

**Федеральное государственное бюджетное учреждение науки  
Институт кристаллографии им. А.В. Шубникова  
Российской академии наук  
(ИК РАН)**

На правах рукописи

Волков Владимир Владимирович

**Спектроскопия и малоугловое рассеяние в решении обратных задач  
исследования многокомпонентных систем**

Специальность 01.04.18 – «кристаллография, физика кристаллов»

Диссертация на соискание ученой степени доктора химических наук

Москва 2013

## ОГЛАВЛЕНИЕ

Введение .....	5	
Часть 1	Компьютерный спектроскопический анализ многокомпонентных систем .....	20
1.1	Литературный обзор .....	20
1.2	Геометрическое представление задачи разложения и единственность решения .....	35
1.3	Методы определения числа компонентов по матрице экспериментальных данных .....	39
1.3.1	Литературный обзор .....	39
1.3.2	Методика приготовления тестовых задач для спектроскопического анализа .....	52
1.3.3	Графический метод анализа сингулярных чисел .....	53
1.3.4	Анализ спектра сингулярных чисел и матрицы остатков ....	56
1.3.5	Анализ сингулярных векторов матрицы данных .....	60
1.3.6	Использование эталонных спектров и библиотечный поиск	74
1.3.7	Определение состава единственного спектра смеси с помощью набора эталонных спектров .....	79
1.4	Методы разделения спектров смесей .....	81
1.4.1	О единственности решения задачи разложения .....	82
1.4.2	Метод итерационного проектирования с отбором признаков .....	85
1.4.3	Метод итерационного моделирования индивидуальных спектров с использованием дополнительной информации и ограничений .....	90
1.4.3.1	Анализ колебательных спектров жидкой воды .....	104
1.4.3.2	Анализ спектров при наличии межмолекулярных взаимодействий: смеси диметилформамида с гексафторэтиленом .....	109

1.4.3.3	Анализ устойчивости решения .....	112
1.5	Использование сведений об относительных концентрациях компонентов в смесях .....	114
1.5.1	Литературный обзор .....	115
1.5.2	Тактика поиска спектров компонентов с анализом устойчивости и надежности решения .....	119
Часть 2	Определение структурных параметров наноразмерных объектов по данным малоуглового рентгеновского и нейтронного рассеяния .....	131
2.1	Литературный обзор .....	132
2.2	Изучение формы наночастиц и биомолекул в растворе по данным малоуглового рассеяния .....	148
2.2.1	Проблема однозначности .....	148
2.2.2	Моделирование формы наночастиц ортогональным рядом	156
2.2.2.1	Численное исследование неоднозначности и стабильности решения .....	159
2.2.2.2	Пример определения форм белковых молекул в растворе по данным малоуглового рентгеновского рассеяния .....	182
2.2.2.3	Построение многокомпонентной оболочечной структуры рибосомы 70S <i>E.coli</i> в растворе по данным нейтронного малоуглового рассеяния .....	185
2.2.3	Определение формы частиц методом шариковых моделей	205
2.2.3.1	Алгоритмические принципы поиска шариковых моделей	206
2.2.3.2	Исследование устойчивости и однозначности определения строения частиц методом шариковых моделей .....	216
2.2.4	Исследование структуры белковой фазы частицы 30S <i>Thermus Thermophilus</i> в растворе методом малоуглового рассеяния с вариацией контраста .....	233

2.2.5	Исследование морфологии макромолекул иммуноглобулинов и ревматоидного фактора человека по данным малоуглового рассеяния от раствора .....	242
2.2.5.1	Исследование устойчивости определения формы молекулы IgM .....	246
2.2.5.2	Результаты исследования растворов молекул иммуноглобулинов и ревматоидного фактора .....	249
2.3	Анализ данных малоуглового рассеяния от смесей полидисперсных частиц .....	277
2.3.1	Применение метода моделирования распределений: анализ микроэмульсий .....	287
Выводы	.....	296
Приложение	Программное обеспечение для решения задач анализа смесей и данных малоуглового рассеяния .....	298
Литература	.....	302
	Список цитируемой литературы .....	302
	Список работ автора .....	323

## ВВЕДЕНИЕ

Разработка эффективных методов определения химического состава и структуры вещества всегда была актуальной научной задачей, и успехи в этой области приобретают с развитием технологий все большее значение, так как изучение связей структура – свойство необходимо для решения всех практических задач по созданию новых материалов, мониторинга окружающей среды, синтеза веществ с заранее заданными функциями и т.д. Постоянно растущая сложность фундаментальных и практических задач делает развитие новых методов анализа неизменно важной частью современной химии. Хорошие перспективы при решении данной проблемы имеют методы компьютерной химии, которая охватывает не только квантово-химические расчеты, с помощью которых можно проводить расчеты электронной структуры и геометрии самых разнообразных химических систем, но и эмпирические и полуэмпирические методы расчета структуры и физико-химических свойств веществ. Компьютерные расчеты стали таким же инструментом исследования, как привычный химический эксперимент.

Определение химического состава позволяет делать выводы о строении и связанных с ним свойствах веществ на уровне атомного разрешения. Многие свойства современных материалов, в частности, включающих биомакромолекулы, определяются надатомной организацией. Поэтому одинаково важны как разработка методов химического компонентного анализа, так и определения строения исследуемых образцов при разрешении от единиц до десятков нанометров.

С другой стороны, современные материалы часто представляют собой системы, функции которых сохраняются в ограниченном диапазоне изменения внешних условий. Это требует применения неразрушающих физических методов исследований, таких, в частности, как спектроскопия в различных диапазонах длин волн излучения и малоугловое рассеяние рентгеновских лучей и нейтронов,

позволяющие исследовать вещество в естественном состоянии. Анализу данных, полученных с помощью таких методов, посвящена данная работа.

В работе **выбраны два неразрушающих метода**, позволяющие характеризовать структуру как со стороны химических связей (спектроскопия), так и пространственного строения (малоугловое рентгеновское и нейтронное рассеяние).

Специфика неразрушающих методов часто заключается в небольшом количестве экспериментально доступной информации об объектах исследований, которые, по существу, представляет собой физически или химически неразделяемые смеси. Компоненты многих таких систем существуют только в комбинации с другими составляющими и/или их невозможно выделить и охарактеризовать в чистом виде, без структурных и химических изменений. Единственный способ получения данных о чистых компонентах в этих случаях - компьютерная обработка результатов измерений с помощью алгоритмов анализа многомерных сигналов и построения физико-химических моделей объектов. Сложность возникающих на практике задач требует разработки алгоритмов и программ, как общего назначения, так и учитывающих особенности частных случаев, что делает эти исследования неизменно актуальными.

Количественные методы спектроскопического анализа многокомпонентных смесей сложного состава являются широко используемым и активно развиваемым инструментом аналитической химии благодаря своей относительной простоте, доступности и информативности. Измеряемые физические характеристики сложных систем (спектры поглощения, испускания, флуоресценции, магнитного и парамагнитного резонансов, рассеяния электромагнитного излучения и элементарных частиц и т.д.) часто представляют собой сумму характеристик отдельных компонентов. Однако соединения, входящие в состав реальных систем, с которыми работают химики-аналитики, иногда невозможно выделить в изолированном и/или неизменном виде. Химический анализ таких смесей, заключающийся в определении спектральных характеристик компонентов сложен, что делает разработку методов разделения и

поиск условий их эффективного применения чрезвычайно актуальной для аналитиков, работающих в различных областях.

Как было отмечено, современное развитие аналитической химии и, в частности спектроскопии, невозможно представить без использования математических методов анализа экспериментальных данных. В настоящее время данное направление получило развитие в создании специализированной дисциплины, получившей название хемометрики (Chemometrics), или хемометрии. Одним из важных направлений хемометрики является анализ многофакторных данных в совокупности, без выделения предварительно одного или нескольких «главных» факторов или компонентов [1]. Хемометрические алгоритмы, разработка которых активно началась в 70-х гг. прошлого века, дали мощный толчок для развития спектроскопических методов анализа. Развитие вычислительной техники применительно к задачам аналитической химии привело к созданию автоматизированного оборудования и в связи с этим очень быстро и плодотворно разрабатываются численные методы, которые предлагают химикам-аналитикам не только новые способы интерпретации данных, но и новые принципы постановки экспериментов. Построенные в результате решения обратных задач многомерные эмпирические модели компонентного состава смесей позволяют обнаружить новые связи или явления, так как учитывают существующие химические и структурные особенности объекта.

Важным предметом исследований в этой области является создание и применение алгоритмов, предназначенных для численного разделения наборов спектров многокомпонентных смесей на спектры индивидуальных соединений. При этом получение данных о составляющих смесей не требует априорных знаний об исследуемой системе. По классификации аналитической спектроскопии такие методы являются безэталонными и "автомодельными" (self-modeling). Другими преимуществами таких методов являются быстрота анализа и относительная дешевизна, применимость не только в различных областях химии, но и в физических исследованиях, например, в исследовании примесей и дефектов в кристаллах, структуры вещества в разном агрегатном состоянии, в мониторинге

объектов окружающей среды. В ряде случаев хемометрические методы анализа спектроскопических данных предпочтительнее дорогостоящих хроматографических и масс-спектрометрических методов разделения, которые хотя и обладают большими возможностями, не всегда пригодны для исследования, например, равновесных систем и в случае дистанционного мониторинга.

Если продолжить рассмотрение возможностей компьютерных методов спектроскопического анализа, то можно прийти к выводу, что невозможно указать области исследований, в которой они не были бы применимы и востребованы в настоящее время. К настоящему времени для разделения аддитивных сигналов разработано и активно используется несколько программ, в некоторых из которых реализованы основные принципы, предложенные автором данной работы за несколько лет до появления этих программ.

Область применения методов разделения аддитивных данных может быть весьма широкой. Под аддитивными данными можно подразумевать данные одновременных измерений различных характеристик объектов с помощью измерительных приборов или многоканальных датчиков, пропорционально регистрирующих физические или какие-либо другие параметры. "Спектр" при этом будет представлять набор таких измерений, представленный в виде зависимости "результат измерения - порядковый номер канала датчика".

В кристаллографии можно указать такие задачи, как исследование спектральных характеристик примесей и различных поглощающих центров в кристаллах, причем варьируемыми параметрами при получении спектральных контуров могут быть концентрации центров, угол поворота поляризатора (анализатора) в спектрометре (если поляризация прошедшего или излученного света различна для разных центров) и т.п.

Параметры, варьирование которых приводит к изменению относительных спектральных вкладов компонентов в смесях, могут быть самыми различными. В химии - это время и условия хроматографирования; рН среды, температура, концентрация исходных веществ и т.д. при исследовании равновесных

химических систем; условия жизнедеятельности биологических объектов; время и другие параметры при исследовании промежуточных продуктов многостадийных реакций; энергия возбуждения в люминесцентном анализе и фотоэлектронном эксперименте; ионная сила раствора и многое другое. В физических экспериментах на современных ярких источниках излучения, таких как синхротроны 4-го поколения и лазеры на свободных электронах получают наборы спектров, или данные рассеяния, снимаемые через определенные интервалы времени. Это позволяет исследовать временное изменение структуры образцов, в частности, путем математического выделения компонентов спектральных контуров. В этих экспериментах существует и серьезная проблема деградации образцов под воздействием высокоинтенсивного излучения, и методы компонентного анализа можно применять в таких случаях в управляющих программах, распознавая момент начала деградации, для которого будет характерен рост числа компонентов по мере поступления новых кадров измерений.

Актуальность разработки новых методов анализа многокомпонентных данных определялась не всегда удовлетворительными характеристиками известных в то время машинных методов: небольшое допустимое число компонентов (как правило, не более 3-х), невозможность учета имеющейся дополнительной информации, отсутствие анализа единственности решения, его устойчивости к ошибкам в исходных данных и т.п. В данной работе сделана попытка упрощения математического аппарата алгоритмов поиска индивидуальных спектров и расширения возможностей метода до анализа смесей, состоящих их 7-10 и более компонентов, сделан акцент на простоту методов спектроскопического анализа и единый математический базис разработанных алгоритмов, основанный на сингулярном анализе матрицы экспериментальных данных.

Исследование многокомпонентных систем включает и определение строения наноразмерных частиц, кластеров и макромолекул, свойства которых зависят не только от их состава, но и от пространственной организации на уровне

разрешения в единицы и десятки нанометров. Для исследования таких объектов как нельзя лучше подходит малоугловое рассеяние (МУР) рентгеновских лучей и нейтронов, так как эти методы не приводят к структурным и другим перестройкам в образцах во время измерений и позволяют изучать разориентированные, неупорядоченные системы. Кроме того, многие материалы представляют собой не полностью упорядоченные системы, что делает неприменимыми такие высокоинформативные методы, как, например, рентгеноструктурный анализ. Наиболее востребованные параметры, которые позволяет определять МУР – это распределения наночастиц по размерам в случае полидисперсных объектов и форма отдельных частиц и их компонентов в монодисперсном случае. Последний вариант особенно важен для изучения конформации биологических макромолекул, внешние условия существования которых весьма ограничены, а структура может претерпевать конформационные изменения в процессе подготовки образцов для электронной и зондовой микроскопии или при кристаллизации. Экспериментальные данные МУР от полидисперсных систем и биологических объектов в большинстве случаев можно рассматривать как спектры неразделяемых смесей. Компонентами, параметры которых необходимо определить, могут быть как сами наночастицы одного размера и формы, так и их составные части, например, домены в белковых комплексах. Некоторые методические подходы, развитые для спектроскопического анализа смесей, применимы в этом случае и для структурного анализа.

Важнейшим фактором развития МУР стали компьютерные методики анализа данных рассеяния и построения структурных моделей. Экспериментальные данные МУР позволяют напрямую определять такие общие характеристические параметры исследуемых систем, как размер частиц, радиус инерции, объем и массу. Построение же трехмерных моделей представляет собой сложнейшую задачу из-за ограниченного количества информации, содержащейся в данных рассеяния. В сравнении с рентгенограммами монокристаллов, где измеряются трехмерные наборы рефлексов, МУР дает картины рассеяния, усредненные по ориентациям частиц, а для полидисперсных систем - и по

размерам рассеивающих объектов. Однозначное решение обратной задачи рассеяния, т.е. восстановление по экспериментальным данным трехмерной структуры объекта, в общем случае невозможно. Тем не менее в ряде работ 1970-80-х гг. (обзор которых дан в [1]) было показано, что при определенных ограничениях трехмерные модели удается определять из данных МУР. Основной прогресс был достигнут в конце 1990-х годов, когда появились новые подходы к интерпретации данных рассеяния изотропными монодисперсными системами. Эти подходы, реализованные в виде компьютерных программ, доступных всему научному сообществу, получили широкое применение сначала в исследованиях биологических систем, а затем и для таких небιологических объектов, как металлические наночастицы.

Основная проблема анализа данных МУР, как и спектроскопического анализа смесей, состоит в плохой обусловленности обратной задачи, то есть сильной зависимости решения от привлекаемых априорных данных, если таковые используются. Ситуация часто усложняется и тем, что целевая функция задачи может иметь несколько локальных минимумов, что приводит к зависимости решений от стартовых значений параметров моделей и от метода поиска минимума. Классические теоретические оценки дисперсии решений на практике малопригодны, так по большей части они представляют собой оценки сверху и не дают ответа на вопрос о надежности полученных параметров моделей.

Среди доступного программного обеспечения можно найти очень небольшое число программ, обеспечивающих получение физически обоснованных решений. В известных программах, как правило, не проводится анализ задач на обусловленность и статистический анализ остатков, необходимый для обоснования результатов. Методы анализа должны обеспечивать оценку числа компонентов, их индивидуальных спектральных или структурных характеристик и позволять учитывать дополнительную информацию о решении. Должна быть обеспечена возможность идентификации компонентов в смесях по априори известным индивидуальным спектрам. При структурном анализе – определении формы компонентов составных наночастиц должна быть обеспечена

проверка на число составляющих по данным измерений малоуглового рассеяния. При анализе размерных параметров наночастиц в полидисперсных смесях должна быть обеспечена возможность учета различных формфакторов и эффективность поиска решения. Понятие эффективности означает, что методы должны обеспечивать получение физически значимого решения за разумное время счета. Для обеспечения наибольшей общности методов используемые статистические оценки качества решения должны быть основаны на критериях, свободных от параметров законов распределения случайных составляющих экспериментальных данных.

Таким образом, важнейшая часть разработки методов анализа состава и структуры вещества заключается в оптимальной с точки зрения обусловленности задачи параметризации модели и обоснованном выборе метода минимизации целевых функций, а также в развитие способов оценки надежности получаемых решений.

**Целью данной работы** Целью настоящей работы была разработка методов химического анализа многокомпонентных систем по спектроскопическим данным и определения структурных параметров наночастиц в моно- и многокомпонентных по данным рентгеновского и нейтронного малоуглового рассеяния. Важной частью была разработка способов оценивания устойчивости решений и рассмотрение вопросов практической реализации предложенных алгоритмов.

Круг исследуемых объектов был **ограничен** системами, в которых можно было пренебречь эффектами взаимодействия компонентов, приводящими к зависимости их характеристик от относительного содержания в объекте исследований. Задача нахождения спектроскопических и структурных характеристик компонентов в таких случаях может быть сформулирована в терминах линейной алгебры. Принципиально нелинейной задача анализа аддитивных смесей становится при попытке учесть дополнительную информацию о решении, которую можно параметризовать в виде штрафных функций. В данной работе не рассматриваются методы анализа смесей сильно взаимодействующих

компонентов, так как взаимодействия и их влияние на экспериментальные данные невозможно формализовать универсальным способом. Для анализа таких систем необходимо разрабатывать индивидуальные алгоритмы.

Объединяющей идеей работы было применение сингулярного анализа матрицы экспериментальных данных как при структурной интерпретации данных от многокомпонентных частиц так и при анализе спектров смесей.

### **Диссертационная работа состоит из двух частей и приложения.**

Первая часть посвящена методам решения обратных задач анализа спектроскопических данных и данных малоуглового рассеяния рентгеновских лучей и нейтронов от многокомпонентных систем. Эта часть работы представляет собой продолжение и завершение исследований и разработок, описанных автором в кандидатской диссертации. Во второй части рассматриваются методы анализа данных малоуглового рассеяния от систем наночастиц. Разработанное программное обеспечение основано на единой библиотеке из более чем 500 процедур, написанных на языке Фортран.

Во второй части работы рассмотрены методы исследования размерных и структурных параметров наночастиц по данным малоуглового рассеяния. Рассмотрены вопросы однозначности определения формы молекул биополимеров по данным малоуглового рассеяния. Данным аспектом часто пренебрегают, тогда как это ключевой момент при интерпретации данных рассеяния. Рассмотрены подходы, связанные с использованием дополнительной информации и требований к решению, улучшающих стабильность решения, но ухудшающих обусловленность задачи (кажущийся парадокс). Обсуждаются результаты численного исследования устойчивости восстановления трехмерной функции формы по данным рассеяния, как в рамках оболочечной ортогональной модели, так и "шариковых" моделей. Рассматривается предложенный автором метод коррекции исходных данных при вариации контраста для многокомпонентных структур на примере структуры рибосомы *E.coli.*, основанный на сингулярном анализе матрицы базисных функций рассеяния, а также коррекция (линеаризация) экспериментальных данных методом обратного восстановления по неполному

базису. Часть заканчивается рассмотрением методов анализа полидисперсных смесей наночастиц по данным малоуглового рассеяния. рассмотрены вопросы расчета вкладов компонентов при известных формфакторах, расчета интенсивности рассеяния от компонентов и моделирование распределений по размерам наночастиц в полидисперсных многокомпонентных системах в виде квазинепрерывного распределения при ограниченном числе формфакторов.

В Приложении приведено перечисление основного программного обеспечения для решения задач анализа смесей и данных малоуглового рассеяния. Рассмотрены особенности реализации базового диалогового пакета программ оптимизации и способы повышения работы алгоритмов минимизации нелинейных целевых функционалов. Приведено описание организации программ для анализа наборов аддитивных спектров смесей, программы для поиска оболочечной модели наночастиц по данным малоуглового рассеяния, модификации программ анализа данных малоуглового рассеяния от многокомпонентных структур и расчета распределений по размерам частиц по данным малоуглового рассеяния.

**Научная новизна.** Разработаны методы спектроскопического компонентного анализа смесей химических соединений различной природы, позволяющие проводить исследование образцов с произвольным числом компонентов, которое ограничивается только уровнем шумов измерений. Новизна методов заключается также в разработанных приемах оценки стабильности и достоверности получаемых результатов. Развитые методы позволяют достаточно просто проводить компонентный анализ как с привлечением дополнительной спектроскопической или другой информации, так и в условиях отсутствия априорных сведений об изучаемой системе. Данные методы применены для выделения индивидуальных спектров в смесях органических и неорганических соединений, а также впервые использованы для коррекции интенсивности в наборах данных измерений малоуглового рассеяния и определения числа рассеивающих компонентов.

Разработанные с непосредственным участием автора методы анализа

данных малоуглового рассеяния развиты в части повышения стабильности результатов и эффективности поиска формы частиц в монодисперсных образцах и распределений по размерам в полидисперсных системах. Методы применены для изучения морфология макромолекул ряда белков и биологических комплексов, которые не поддаются исследованиям другими методами. Определены размерные параметры наночастиц и неоднородностей в системах различной природы. Новизна полученных результатов подтверждается высокими индексами цитирования работ с участием автора.

**Полученные результаты** перечислены ниже.

1. Разработаны алгоритмы и программы решения обратных задач анализа наборов аддитивных неотрицательных спектров смесей, основанные на сингулярном разложении матрицы данных. Разработанные алгоритмы включают оценку числа независимых компонентов в смесях, определение контуров индивидуальных спектров и способы исследования устойчивости решений. Показана возможность анализа систем с числом априори неизвестных компонентов более 3-х, в зависимости от качества экспериментальных данных. Разработанные подходы реализованы в виде компьютерных программ и применены для анализа ИК спектров смесей органических соединений, колебательных спектров жидкой воды, анализа серий измерений интенсивности малоуглового рентгеновского рассеяния.

2. Разработанные способы разложения аддитивных спектров впервые применены, в частности, для анализа и коррекции данных малоуглового нейтронного рассеяния с вариацией контраста, по которым были определены структурные модели селективно дейтерированных частиц рибосомы 70S *E-Coli* и 30S *Thermus Thermophilus*.

3. С помощью модельных расчетов проведен анализ единственности и устойчивости решения задач поиска формы частиц в монодисперсных системах по данным малоуглового рентгеновского рассеяния. Определены параметры анизометрии простых геометрических форм, являющиеся пограничными для устойчивых решений.

5. Программа поиска шариковых моделей частиц по данным малоуглового рассеяния модифицирована способами автоматического определения оптимальных значений параметров и режимов алгоритма поиска, что позволило уменьшить время расчетов в 3-5 раз и уменьшить разброс решений. С помощью разработанного программного обеспечения по данным малоуглового рассеяния от раствора определены формы молекул иммуноглобулина М и ревматоидного фактора человека. Показано, что различие между макромолекулами состоит в меньшей заселенности периферийных участков молекул ревматоидных факторов. Выдвинута гипотеза о вероятной большей гибкости периферийных фрагментов ревматоидного фактора, что независимо подтверждено другими методами исследований.

6. Разработана модифицированная схема поиска размерных распределений частиц по данным малоуглового рассеяния с повышенной устойчивостью решений. Повышение устойчивости достигнуто путем расчета невязки между экспериментальными и модельными интенсивностями рассеяния на неравномерной угловой сетке с определением шага сетки из оценок максимального и минимального радиусов инерции частиц в системе. Проведен анализ ряда систем с наночастицами – металлических катализаторов на органических и углеродных носителях, органико-неорганических композитов, полимерных систем.

**Практическая значимость** работы определяется актуальностью поставленных задач, универсальностью разработанных методов спектроскопического анализа смесей и структурного анализа данных рассеяния от систем наночастиц. Эффективность методов продемонстрирована на примере решения задач из разных областей химического и структурного анализа сложных объектов. Отметим основные применения разработанных методов.

1. Качественный и количественный спектральный анализ неразделяемых смесей любых сложных объектов по набору аддитивных неотрицательных данных измерений физических характеристик – спектров различной природы

(поглощения, испускания, люминесценции, ядерного магнитного резонанса и т.д.) и интенсивности малоуглового рентгеновского и нейтронного рассеяния.

2. Анализ измерений смесей с целью определения числа компонентов и коррекции экспериментальных данных с целью приведения их в соответствие с физической моделью объекта и проверки адекватности самой модели.

3. Структурный анализ смесей наночастиц по данным малоуглового рентгеновского и нейтронного рассеяния.

4. Определение формы наночастиц в разбавленных системах, в частности, анализ строения белковых макромолекул, как в рамках однофазных, так и многофазных моделей.

Предложенные в работе методы анализа смесей и строения наночастиц реализованы в ряде компьютерных программ и программных комплексов, входящих в систему ATSAS [A39], которая использовалась на момент написания работы более чем в 2000 лабораторий и находящийся в свободном доступе (URL: <http://www.embl-hamburg.de/research/unit/svergun/index.html>).

С учетом проведенных разработок подготовлен и выпущен Национальный стандарт Российской Федерации (ГОСТ Р 8.698–2010 Государственная система обеспечения единства измерений. Размерные параметры наночастиц и тонких пленок. Методика выполнения измерений с помощью малоуглового рентгеновского дифрактометра. –М. : Стандартинформ, 2010.– 41 с.). Выпущено методическое пособие "Определение формы частиц по данным малоуглового рентгеновского и нейтронного рассеяния" (URL: [http://nano.msu.ru/files/systems/4\\_2010/practical/41\\_full.pdf](http://nano.msu.ru/files/systems/4_2010/practical/41_full.pdf)).

**Основной вклад** в работу по анализу спектров неразделяемых смесей, выполненные с соавторами, сделан лично автором. Ему принадлежит постановка задачи поиска спектров компонентов при отсутствии дополнительной информации, использование сингулярного анализа матрицы данных и схемы оценки числа компонентов. Автор разработал все алгоритмы и программы, касающиеся анализа спектров смесей, базовые комплексы программ нелинейной оптимизации, разработаны головные модули и ряд библиотечных программ

анализа полидисперсных смесей наночастиц и формы частиц в монодисперсных системах.

Соавторы принимали непосредственное участие в разработке идеологической составляющей схем анализа смешанных спектров и проведении численных расчетов. Базовые схемы поиска формы частиц в монодисперсных системах и распределений частиц по размерам принадлежат соавторам, автору принадлежит численная реализация с использованием высокоэффективных методов оптимизации и применение разработанных им методов оценки числа компонентов в наборах данных рассеяния. Автор принимал активное участие в проведении экспериментальных исследований, разработке части методик измерений и интерпретации данных, в обработке результатов экспериментов.

**Публикация и апробация работы.** Работа проводилась с 1977 года в Институте кристаллографии АН СССР и во Всесоюзном научно-исследовательском институте водоснабжения и гидротехнических сооружений ВНИИ ВОДГЕО. Всего в области анализа спектров смесей и данных малоуглового рентгеновского рассеяния опубликовано 86 работ, в данной диссертации использовано 63, из них 20 изложены более подробно. Результаты неоднократно докладывались на российских и международных научных конференциях [Т1 – Т80].

**Основные положения, выносимые на защиту.**

1. Алгоритмы анализа наборов аддитивных неотрицательных спектров смесей, основанные на сингулярном разложении матрицы данных, линейных преобразованиях спектральных контуров и статистическом анализе результатов, в том числе:

— методы оценки числа компонентов, основанные на статистических критериях автокорреляции элементов сингулярных векторов матрицы спектроскопических данных;

— метод расчета спектров компонентов и их относительных концентраций в отсутствие дополнительной информации о задаче при условиях максимального различия и простоты контуров индивидуальных спектров;

2. Применение разработанных алгоритмов и методов анализа спектров для коррекции наборов данных малоуглового нейтронного рассеяния с вариацией контраста.

3. Выявление степени однозначности и устойчивости решения задач поиска формы частиц в монодисперсных системах по данным малоуглового рентгеновского рассеяния с помощью модифицированных методов сферических гармоник и шариковых моделей.

4. Результаты исследования морфологии по данным малоуглового рассеяния от раствора макромолекул иммуноглобулина М и ревматоидного фактора человека с анализом стабильности решений.

5. Метод поиска размерных распределений частиц в полидисперсных смесях по данным малоуглового рассеяния при заданных формфакторах компонентов, отличающийся улучшенной стабильностью решений в случае широких распределений. Результаты анализа дисперсности ряда систем с наночастицами различной природы.

## Часть 1

# Компьютерный спектроскопический анализ

## МНОГОКОМПОНЕНТНЫХ СИСТЕМ

### 1.1 Литературный обзор

Существующее множество методов численного разделения спектров смесей на спектры компонентов можно разделить на две группы, различающиеся качеством получаемого решения. Одна группа позволяет найти единственный набор спектров компонентов, совпадающих, при отсутствии ошибок в исходных данных, с истинными. С помощью других методов можно найти лишь приближенные контуры индивидуальных спектров, удовлетворяющие тем или иным критериям их предпочтительности перед другими формально допустимыми решениями задачи. Разумеется, применимость тех или иных способов разложения на индивидуальные спектры определяется характером используемой информации, которую, в свою очередь, также можно разделить на две группы. Основной исходной информацией, как уже упоминалось ранее, будем называть матрицу спектроскопических данных. Любые другие сведения о задаче - дополнительной информацией, которая может касаться относительных концентраций компонентов в смесях, формы индивидуальных спектров, наличия тех или иных известных компонентов и т.д.

Задача численного разложения спектральных контуров на спектры компонентов возникает на практике довольно часто. Развитие компьютерных методов анализа данных привело к появлению специальной области химии - хемометрии, или вычислительной химии, которая включает в себя специальный раздел, касающийся анализа сложных смесей по их молекулярным спектрам [1-3]. Будем называть спектры смесей, отличающиеся только относительным вкладом компонентов **родственными смесями, спектрами или контурами.**

Большая группа работ посвящена анализу смесей химических соединений путем численного анализа молекулярных спектров. Это ультрафиолетовые, видимые, инфракрасные, ЯМР, ЭПР, КР, фотоэлектронные и масс-спектры, которые несут информацию о химической структуре компонентов смесей [4]. Имеются работы по разложению контуров неразрешенных хроматограмм, причем разложению подвергаются как наборы контуров хроматограмм родственных смесей, так и наборы молекулярных спектров смесей-фракций, соответствующих одной хроматограмме. Математической обработке подвергают не только "одномерные" (когда спектр представляет собой зависимость интенсивности измеряемой характеристики от одного параметра, например, длины волны проходящего через образец света), но и "многомерные" данные, например, матрицы спектров испускания/возбуждения в многомерном анализе [5], дополняемые также хроматографическими данными [6] ("трехмерные" данные). Вообще говоря, разложению можно подвергать не только спектроскопические и хроматографические контуры, но и любые другие кривые, для которых выполняются условия аддитивности, например, электрохимические данные и кривые титрования [2,4], а также матрицы изображений.

В кристаллографии и физике можно указать такие задачи, как исследование спектральных характеристик примесей и различных поглощающих центров в кристаллах, причем варьируемыми параметрами при получении родственных спектральных контуров могут быть концентрации центров, угол поворота поляризатора (анализатора) в спектрометре (если поляризация прошедшего или излученного света различна для разных центров) и т.п. Под аддитивными спектрами смесей можно также подразумевать данные одновременных измерений различных характеристик объектов с помощью измерительных приборов или многоканальных датчиков, пропорционально регистрирующих физические или какие-либо другие параметры. "Спектр" при этом будет представлять набор таких одновременных измерений, представленный в виде зависимости "результат измерения - порядковый номер канала датчика". В литературе можно найти много десятков работ и целые книги, посвященные

выделению компонентов в задачах анализа фотографических изображений и распознаванию образов на постороннем фоне.

Актуальность диссертационной работы определялась отсутствием многих методов анализа данных малоуглового рассеяния и неудовлетворительными характеристиками известных, на момент начала работы, методов спектроскопического анализа смесей: небольшое допустимое число компонентов (как правило, 2-3), сложность учета дополнительной информации, отсутствие анализа единственности решения, его устойчивости к ошибкам в исходных данных и т.п.

Даже наиболее простой случай разложения одного суммарного спектра на составляющие, для которых известны их число и форма (как это сделано, например, в [4,7,8]) или другие характеристики (например, моменты [9]), обеспечивающие единственность решения, требует применения критериев качества решения, которые не ограничивались бы расчетом значения хи-квадрат, а позволяли бы более детально оценивать надежность получаемых результатов. Большой практический интерес представляет случай, когда спектроскопические характеристики компонентов неизвестны. Единственный спектр смеси можно разложить при этом на множество различных наборов индивидуальных спектров, даже если известно число последних. Новые возможности открываются в том случае, если вместо одного исходного спектра использовать набор спектров смесей различного количественного состава. Основным требованием к исходным спектрам, кроме неотрицательности, является их аддитивность, т.е. форма спектров компонентов не должна зависеть от их относительной концентрации и от каких-либо других параметров, варьируемых при получении спектров смесей. Например, в случае ИК-спектров органических соединений молекулы компонентов не должны образовывать между собой химические связи, которые приводили бы к сдвигам частот спектральных полос и изменению их формы [10,11].

Задача разложения набора смешанных спектров в данной постановке также не имеет единственного решения, поэтому для получения спектров компонентов необходимо привлекать дополнительные требования к решению, которые мы будем называть дополнительной информацией.

В настоящее время техника разделения смешанных аддитивных сигналов развита уже довольно хорошо. Данной проблеме, например, посвящен сайт (URL [http://en.wikipedia.org/wiki/Independent\\_component\\_analysis](http://en.wikipedia.org/wiki/Independent_component_analysis)), написано много десятков книг и пособий, среди которых можно указать [10-14]. Приблизительный подсчет показывает, что число публикаций превышает несколько десятков тысяч. Поэтому подробный обзор даже основных публикаций занял бы слишком много места. Ограничимся основными положениями, на которых основаны методы выделения индивидуальных контуров в случае анализа аддитивных смесей, и рассмотрим основные моменты, которые наиболее близки к данной работе.

1). Аддитивная модель основана на предположении, что форма контуров компонентов (в общем случае, форма сигналов или пространственных моделей) не меняется при изменении их относительного содержания в смесях. Более подробно это положение будет рассмотрено в главе, посвященной единственности решения задачи разложения.

2). Основной успех методам разделения обеспечило применение критериев различия искомых компонентов, что оказалось ключевым моментом в развитии алгоритмов и программ анализа.

3). Многочисленные алгоритмы компонентного (факторного) анализа и родственные ему используют статистические критерии, обеспечивающие поиск решения, как можно более освобожденного от случайных составляющих. Данный класс методов, как видно из публикаций, применим к ограниченному числу задач и менее универсален по сравнению с методами, основанными на максимальном различии.

Большинство методов основано на построении линейных комбинаций или исходных смешанных контуров, или комбинаций собственных векторов матрицы, составленной из экспериментальных данных.

Из наиболее универсальных методов, которые позволяют разделять наборы смешанных сигналов, можно отметить следующие.

**Independent Component Analysis (ICA)** [14]. Основными предположениями, используемыми в данном методе, являются независимость компонентов и отличие их распределения от нормального. Поиск независимых компонентов, как линейных комбинаций наблюдаемых величин, сводится к определению статистически независимых величин, распределение которых максимально далеко от нормального. Реализации метода отличаются способами оценивания степени близости распределения случайной величины к нормальному. В некотором смысле такие критерии родственны критериям максимального различия компонентов, не использующим статистические оценки. Родственным данному подходу является **Blind signal separation (BSS)** [15], метод обработки массивов составных данных, направленный на восстановление малозаметных сигналов. В данном методе также используют только предположения о взаимной независимости сигналов – компонентов.

Естественно, что использование дополнительной информации позволяет значительно улучшить качество и надежность результатов анализа смесей. Один из таких подходов к проблеме объединяет хемометрический подход (soft modeling) с содержательным кинетическим описанием процессов (hard modeling). Он рассмотрен в отдельном пособии "Кинетическое моделирование спектральных данных" (URL <http://rcs.chemometrics.ru/Tutorials/grey.htm>).

Стали уже традиционными методы автомодельного разделения кривых (self-modeling) - **RADICAL**, **SIMPLISMA** [16], **MCR-ALS** [17], и относительно недавно были предложены алгоритмы, реализующие методы независимых компонент - **MILCA**, (Mutual Information Least Dependent Component Analysis [18,19]) и **SNICA** (Stochastic Non-Negative Independent Component Analysis [20]).

Методы независимых компонент во многих случаях превосходят хемометрические методы и, таким образом, представляют арсенал высокоэффективных, производительных и доступных алгоритмов для анализа разнообразных объектов, где другие методы анализа малопродуктивны или в принципе неприменимы. К настоящему времени алгоритмы MILCA и SNICA успешно апробированы для многокомпонентного анализа сложных объектов. В их числе определение жирно- и водорастворимых витаминов (в одном аналитическом цикле) и неорганических соединений в поливитаминных препаратах, определение ароматических (в том числе полициклических) и серосодержащих соединений в различных видах топлив и фракциях бензина, а также различных объектах окружающей среды. Предложенные методы, таким образом, представляют собой мощный инструмент для улучшения производительности и надежности определения различных соединений на производстве и в аналитических лабораториях. Следует отметить, что программное обеспечение, реализующее алгоритмы MILCA и SNICA, доступно для пользователей. По вопросам, касающимся анализа многомерных данных, подготовлены подробные Интернет-ресурсы, в том числе и учебные (URL <http://www.ihes.fr/~zinovyev/#books>, <http://pca.narod.ru/>, <http://rsc.chph.ras.ru/Tutorials/>).

Цель автомоделного разделения кривых заключается в том, чтобы, имея экспериментальную информацию о многокомпонентной системе, выделить реальное поглощение каждого компонента без использования физико-химической модели или априорной информации о системе. Как максимально эффективно и надежно достичь этого — остается задачей исследований в области разделения многомерных сигналов до настоящего времени. К преимуществам методов относится их «слепой» характер, применимость ко многим типам спектроскопического эксперимента, а также экспрессность и достаточная для практики точность.

Хемометрические методы анализа группы Self Modeling представляют для практики особенный интерес. Это связано с тем, что условия, обеспечивающие получение точного решения, строго говоря, невыполнимы в реальном

эксперименте из-за наличия трудно контролируемых вкладов в ошибки измерений. Здесь мы рассмотрим такие методы, которые, будучи применены к невозмущенным данным, позволяют найти точное решение в случае строгого выполнения условий частичной неперекрываемости индивидуальных спектров (более подробно особенности таких случаев будут рассмотрены далее в разделе, посвященном оценкам числа компонентов). Как уже ранее нами отмечалось, при этом для получения единственного решения необходимо привлекать дополнительную информацию. В большинстве методов используют требования неотрицательности спектров компонентов и соответствующих им относительных концентраций.

Для унификации рассмотрения методов разложения введем основные обозначения. Пусть контуры спектров представлены на некоторой сетке из  $N$  значений абсциссы  $i = 1, \dots, N$  (в дальнейшем абсолютные значения и единицы измерений шкалы абсцисс нас интересовать не будут, данные полагаются безразмерными). Обозначим векторы-спектры смесей  $\mathbf{d}_j, j = 1, 2, \dots, M$  ( $M$  – число смесей в исследуемом наборе), спектры компонентов -  $\mathbf{x}_k, k = 1, 2, \dots, K$  ( $K$  – число компонентов). Условия аддитивности означают, что каждый из спектров смесей должен быть линейной комбинацией спектров компонентов (влияние различных шумов мы сейчас рассматривать не будем):

$$d_{ij} = \sum_{k=1}^K x_{ik} \cdot c_{kj}, \quad (01)$$

где  $x_{ik}$  - относительный спектральный вклад  $k$ -го компонента в  $j$ -й спектр смеси с весом (концентрацией)  $c_{kj}$ . Индекс  $i$  обозначает порядковый номер точки в спектре. В матричной форме соотношение линейности записывается в виде

$$D = X C, \quad (02)$$

где  $D$  –  $N \times M$  матрица спектров смесей,  $X$  -  $N \times K$  матрица спектров компонентов,  $C$  –  $K \times M$  матрица относительных концентраций.

Наиболее распространенным является подход Self Modeling, впервые предложенный в [71] для разложения двухкомпонентных смесей. Self Modeling

основан на представлении спектров компонентов  $x_{ik}$  в виде линейной комбинации собственных векторов  $u$  матрицы вторых моментов (ковариационной матрицы)

$$D \cdot D^T = U \cdot \Lambda \cdot U^T, \quad (03)$$

$\Lambda$  – диагональная матрица собственных значений, упорядоченных в порядке убывания::

$$x_{ik} = \sum_{j=1}^M u_{ij} \cdot y_{jk} \quad (04)$$

где  $y_{jk}$  – неизвестные коэффициенты, индекс  $i$  относится к номеру точки в спектре,  $k$  – индекс компонента,  $j$  – индекс собственного вектора.

В методах Self Modeling рассматривают коэффициенты возможных линейных комбинаций  $y_{jk}$  для формально допустимых спектров компонентов при условии неотрицательности элементов матриц  $X$  и  $C$ . В [21] показано, что при выполнении условий неперекрываемости (типа условий Аленцева-Фока [22-24]) наиболее удаленные друг от друга векторы линейных комбинаций соответствуют истинным спектрам компонентов. В наших обозначениях, на Рисунке 1.1 (см. раздел 1.2) зонам допустимых решений в пространстве спектров (а не концентраций) соответствуют области АОВ и СОД. ОА и ОД векторы соответствуют требованиям неотрицательности спектров и соблюдения условий Аленцева-Фока.

В [25,26] техника Self Modeling модифицирована для обработки трехкомпонентных смесей. Зоны допустимых решений (т.е. коэффициентов линейных комбинаций) находят с помощью случайного поиска [27] или довольно сложной логикогеометрической процедуры анализа условий неотрицательности. Последняя процедура состоит из ряда последовательных шагов. Вначале строят симплекс ограничений на величину коэффициентов для условия неотрицательности спектров. Затем строят аналогичный многоугольник для условий неотрицательности концентраций. При этом первый симплекс включает в себя второй. Пространство между ними представляет собой зону допустимых линейных комбинаций собственных векторов  $U$ , дающих спектры компонентов. Однако, в этом пространстве также оказываются запрещенные зоны, связанные с

невыполнением условий аддитивности спектров смесей. На третьем этапе, наиболее сложном, эти зоны отсеиваются с помощью специальных геометрических построений и, в результате, решением являются три непересекающиеся области, в каждой из них может находиться по одному вектору, соответствующему одному из спектров компонентов. Авторы [26] отмечают, что для четырех и более компонентов алгоритм существенно усложняется и ими не разрабатывался.

Аналогичный метод получения спектров компонентов путем линейных комбинаций векторов  $U$  предложен в [28]. Для отбора решений помимо уже рассмотренных требований неотрицательности, которые, как отмечается, лишь сужают зоны неопределенности, предложено использовать дополнительный критерий - максимального различия спектров компонентов. Последнее требование непосредственно относится к форме спектральных контуров, поэтому рассмотрение ведется не в пространстве концентраций, а в пространстве спектров. Первый шаг алгоритма, как обычно, заключается в оценке числа компонентов и получение матрицы собственных векторов (в наших обозначениях  $U$ ). Затем коэффициенты линейных комбинаций нормируют по первому элементу, чтобы спектры возможных компонентов были нормированы так, чтобы концы соответствующих им векторов образовывали  $K-1$  - мерный симплекс в  $N$  - мерном пространстве спектров. С помощью техники линейного программирования подбирают такие коэффициенты линейных комбинаций, которые при соблюдении условий неотрицательности дают симплекс максимального объема, что соответствует максимальному различию индивидуальных спектров. Объем симплекса рассчитывают как определитель матрицы, составленной из полученных векторов-спектров компонентов. Авторы [28] указывают, что, в отличие от техники Self Modeling их метод применим к задачам любой размерности по числу компонентов, однако примеров с  $K > 3$  в литературе приведено не было.

Условия неотрицательности являются основополагающими в методах факторного анализа, широко использующегося при разложении спектров смесей

[1,29] и также базирующегося на построении возможных линейных комбинаций собственных векторов. В [29] сформулированы основные положения техники факторного анализа. Во-первых, спектры смесей центрируют, т.е. преобразуют следующим образом:

$$d'_{ij} = d_{ij} - \hat{d}_i,$$

$$\hat{d}_i = \sum_{l=1}^M \frac{d_{il}}{M}$$

и, при необходимости, взвешивают для выравнивания вкладов экспериментальных ошибок:

$$d''_{ij} = \hat{d}_i \cdot \varpi_i,$$

где  $\varpi_i$  - вектор весов. Центрирование уменьшает число линейно независимых переменных на 1, что легко восстанавливается обратным преобразованием. Далее, спектры смесей нормируют к единичной площади под спектральным контуром ("Unit I" - спектры). Для таких данных авторы [29] сформулировали симплексо-теорему, утверждающую, что концы векторов-спектров компонентов должны лежать внутри симплекса, построенного в  $N$  - мерном пространстве спектров и имеющем размерность  $K-1$ , т.е. позволяет уменьшить размерность задачи еще на 1. Рассмотрение неотрицательности концентраций приводит авторов [29] к выводу, что допустимые решения лежат вне некоторого второго симплекса, построенного внутри первого, т.е. к выводу, аналогичному сделанному в [28]. В нашей трактовке на Рисунке 1.1 первому симплексу соответствует отрезок AD, второму - BC, зоны допустимых решений для двух компонентов - соответственно отрезки AB и CD.

Практически аналогичное применение факторного анализа рассмотрено также в [25,40-41] за исключением того, что вопрос неединственности решения не обсуждается, хотя это имеет, как мы видим, принципиальное значение.

В [31] предложено два метода поиска матрицы  $X$ . Первый из них заключается в итерационном подборе элементов концентрационной матрицы  $C_{mod}$

с помощью программы нелинейной оптимизации, минимизирующей отклонение рассчитанной по уравнению (2) матрицы  $D_{mod}$  от экспериментальной, при условиях неотрицательности  $D$ ,  $X$  и  $C$ . Легко видеть, что этот способ может лишь дать решение, произвольным образом попадающее в допустимые зоны, и соответствующая неоднозначность решений пропорциональна ширине этих зон. Другой метод, названный авторами альтернативным, основан на итерационном нахождении элементов абстрактной концентрационной матрицы, вначале назначаемой исследователем. Один шаг процедуры состоит из вычисления методом неотрицательных наименьших квадратов матрицы  $X$  и, тем же методом, соответствующей матрицы  $C$ . Его повторяют до сходимости, когда элементы  $X$  и  $C$  перестают изменяться от шага к шагу. Результирующую матрицу  $X$  и принимают за решение. Данный метод также не гарантирует сходимость решения к истинному, а численная процедура неоправданно усложнена.

В ряде работ для получения единственного решения привлекают дополнительные требования к форме индивидуальных спектров, что обеспечивает успешную работу современных программных пакетов, например уже упоминавшихся **MILCA** и **SNICA**. Так, авторы [34,35] предлагают дополнить Self Modeling требованием минимума информационной энтропии спектров компонентов (см. также [36,37]), которая вычисляется в предположении, что абсолютное значение первой производной спектрального контура  $p_{ik}$  означает функцию плотности вероятности, и тогда энтропия  $k$ -го спектра равна:

$$H_k = -\sum_{i=1}^N p_{ik} \cdot \log_2 p_{ik} . \quad (05)$$

Подразумевается, что производная берется от спектра, нормированного к единичной площади под кривой. Численная процедура поиска спектра компонента [38] заключается в подборе с помощью программы одномерной минимизации (т.к. разложению в этих работах подвергались только двухкомпонентные смеси) одного из коэффициентов линейной комбинации  $y_{jk}$  собственных векторов в (03) (при фиксированном значении другого) так, чтобы

для индивидуального спектра достигался минимум энтропии  $H_k$ . Критерий минимума информационной энтропии можно рассматривать как критерий минимальной сложности спектров компонентов. Авторы [35,39] используют также и другие критерии сложности. Это или величина

$$\int_{\nu} \left[ 1 + \left( \frac{\partial x(\nu)}{\partial \nu} \right)^2 \right]^{1/2}, \quad (06)$$

где  $x(\nu)$  - спектральный контур [35], или условие минимальной интенсивности для максимального числа точек в спектре компонента [39]. Метод [35] по-прежнему заключается в подборе одного из коэффициентов линейной комбинации (02) для каждого из двух спектров компонентов.

В [35] предложен также другой метод, заключающийся в поиске слабо перекрытых участков в спектрах по подобным участкам в контурах собственных векторов и в построении линейных комбинаций последних с целью получения нулевых или малоинтенсивных участков в результирующих контурах. Фактически этот метод основан на процедуре Аленцева-Фока [22], но примененной к собственным векторам  $U$  ковариационной матрицы (03). Степень близости решения к истинным спектрам компонентов зависит от степени их неперекрываемости. Эта методика успешно применена к разложению 2-х и 3-х компонентных смесей. Подробный обзор других работ можно найти в [1], но с математической точки зрения они аналогичны рассмотренным.

Известные в литературе методы разделения, учитывающие условия частичной неперекрываемости спектров компонентов, оказываются несостоятельными, если зоны неперекрываемости малы или их количество недостаточно по Аленцеву-Фоку. Так, если условия наличия неперекрываемых точек (даже не участков) выполнены для каждой пары спектров компонентов, в спектрах смесей может не существовать областей, в которых ненулевая интенсивность обусловлена только одним компонентом. Однако, из геометрического представления (Рисунок 1.1) легко видеть, что такие жесткие

условия и не требуются для существования единственного решения, хотя процедуры, аналогичные подходу Аленцева-Фока оказываются неприменимыми.

В литературе опубликовано несколько методов поиска неперекрытых точек по матрице спектров смесей. В основном, авторы этих публикаций ссылаются на технику факторного анализа (например, [42,43]) или близкий факторному анализу метод главных компонент [1,21]. И все же, если такие точки не образуют непрерывные зоны, эти методы не дают никакой гарантии, что выделенные ими точки являются действительно неперекрытыми. На самом деле они оказываются наиболее (но не обязательно полностью) неперекрытыми. Такая информация о наиболее неперекрытых точках, по крайней мере, позволяет сузить границы неопределенности решения и найти достаточно близкие к истинным спектры компонентов [21,45-47]. Широко используемым методом локализации неперекрытых спектральных участков является Key Factor Analysis [43,48], заключающийся в поиске наиболее разнесенных в пространстве векторов – спектров компонентов. На самом деле легко видеть, что максимальная разнесенность свидетельствует только о максимальной, но не полной, неперекрытости. В процедуре "Self Modeling" [44-46], которую применяют для 2 - 3 компонентных смесей, учитывают также неотрицательность спектров и их концентраций.

Особенность рассматриваемых случаев, как уже отмечалось, состоит в существовании множества допустимых вариантов решения. Окончательный выбор из них должен делаться путем привлечения той или иной дополнительной информации, что в немалой степени зависит от исследователя. Дополнительные критерии, такие, например, как простота спектров, не внушают полного доверия и не могут использоваться как единственные для выбора окончательного варианта. Поэтому они, в общем случае, не снимают неоднозначности. Это известно в литературе по поиску оптимальных решений как решение задачи в условиях неопределенности [49]. Неопределенность означает, что решение должно удовлетворять нескольким целям, или критериям, как правило противоречивым, для которых характерна различная физическая сущность и невозможность

формализации обоснования их применения. Критерий простоты (сложности) спектра именно такой, т.к. он может быть обоснован в значительной степени только интуитивно, в зависимости от конкретной ситуации.

При решении задач в условиях неопределенности допускается применение различных субъективных критериев отбора решений, но, как правило, на последних этапах, после определения вариантов, рациональных в свете главной цели, при выборе среди них окончательного решения. На этом этапе выводы делает человек (эксперт) на основе своего опыта, с интуитивным учетом других целей и информации, которые имеют отношение к рассматриваемой задаче [49]. Однако, и это нам представляется наиболее важным, даже при таком "волевом" выборе весь предшествующий анализ должен гарантировать, что будет принят один из допустимых вариантов. В нашем случае под допустимостью решения мы подразумеваем неотрицательность спектров и концентраций наряду с выполнением соотношения аддитивности (01).

Схемы решения задач в условиях неопределенности, как отмечается в [49], не предполагают единого вычислительного процесса, а представляют собой фактически "человеко-машинные" системы принятия решений. Участие специалиста предполагается на всех этапах: 1) выбора представительного множества исходных данных, 2) проверки условий единственности решений, 3) формирования дополнительных требований к решению, 4) формирования структуры оптимизируемого критерия и 5) окончательного принятия решения. В следующих разделах мы рассмотрим две разработанные нами процедуры поиска спектров компонентов, позволяющие исследователю активно влиять на ход решения, но в рамках допустимости, о которой речь шла выше.

Несмотря на очевидную перспективность, число работ, посвященных применению алгоритмов декомпозиции для качественного и количественного анализа смесей сложного состава, единично (например, [5,6,50,51]). Только относительно недавно появились достаточно универсальные программы для разложения, упомянутые выше и способные решать задачи разложения при числе

компонентов более 2-3. Таким образом, необходимы исследования по практическому использованию алгоритмов (особенно новых) для разделения различных типов спектральных сигналов, выявлению факторов, влияющих на качество декомпозиции, сравнительному анализу различных подходов и апробации алгоритмов на смесях сложного состава.

## 1.2 Геометрическое представление задачи разложения и единственность решения

Разнообразие математических приемов разложения наборов сложных родственных контуров не так велико, как может показаться на первый взгляд и сводится, как правило, к построению линейных комбинаций (т.е. сложению и вычитанию с некоторыми коэффициентами) исходных спектров смесей. Этот принцип лежит в основе всех рассматриваемых в данной работе методов.

Векторное представление спектроскопической информации с учетом соотношений (0.1 - 0.4) позволяет наглядно продемонстрировать особенности задачи анализа смесей. При числе спектральных точек  $N$ , превышающем число компонентов  $K$ , пучок векторов-спектров компонентов определяет в  $N$ -мерном пространстве  $K$ -мерное подпространство, в котором находятся все возможные спектры смесей (в силу (01)). На Рисунке 1.1 показан случай  $N = 3$ ,  $M = 4$ ,  $K = 2$ , т.е. подпространство спектров смесей представляет собой плоскость в 3-мерном пространстве (по координатным осям отложены интенсивности, а порядковый номер оси соответствует номеру точки в спектре), и спектры компонентов образуют базис этой плоскости (неортогональный в общем случае).

В случае, когда известна только матрица  $D$ , задача нахождения спектров компонентов  $X$  заключается в нахождении разложения (02) относительно неизвестных элементов матриц  $X$  и  $C$ . Поскольку неизвестные входят в (2) в виде парных произведений, то это соотношение представляет собой систему нелинейных уравнений, решение которой неединственно. Действительно, если  $X'$  и  $C'$  - решение (2), то всегда можно найти такую невырожденную матрицу  $A$ , что

$$D = X'' \cdot C'', \quad (1.1) \text{ где}$$

$X'' = X' \cdot A$  и  $C'' = A^{-1} \cdot C'$ . Эта проблема неединственности является основной в задаче анализа многокомпонентных смесей по их спектрам и решается путем привлечения той или иной дополнительной информации и ограничений на вид решения.

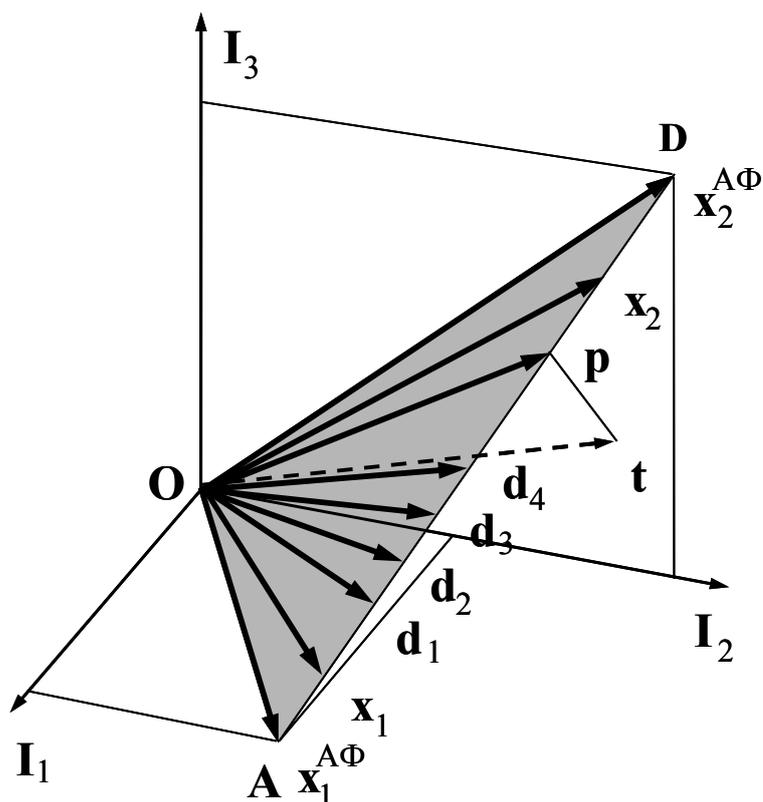


Рисунок 1.1 - Геометрическое представление задачи разложения аддитивных спектров смесей. Здесь  $d_1 - d_4$  спектры смесей,  $x_1$  и  $x_2$  - возможные спектры компонентов,  $x_1^{A\Phi}$  и  $x_2^{A\Phi}$  - частично неперекрываемые спектры компонентов (лежат в плоскостях  $I_1I_2$  и  $I_2I_3$ , соответственно),  $t$  - вектор, не являющийся компонентом смеси,  $p$  - его проекция на плоскость смесей. По осям  $I_1 - I_3$  отложены интенсивности спектров при трех значениях абсциссы (например, трех длинах волн). Все векторы нормированы для наглядности так, что их концы лежат на одной прямой  $AD$ . Области допустимых решений представлены секторами  $AOd_1$  и  $DOd_4$ .

Наиболее общими ограничениями являются требования **неотрицательности** спектров компонентов и их концентраций. Тогда, т.к. спектры смесей есть неотрицательные линейные комбинации спектров компонентов, области существования формально допустимых решений ограничиваются областями подпространства, определяемого векторами-

спектрами смесей (двумерной плоскости AOD на Рисунке 1.1), расположенными в 1-м ортанте и ограниченными изнутри конусом, образованным наиболее разнесенными векторами-спектрами смесей (области AOD<sub>1</sub> и DOD<sub>4</sub> на Рисунке 1.1). Это геометрическое представление наглядно демонстрирует условие единственности решения задачи разложения: необходимо и достаточно, чтобы неотрицательные спектры компонентов имели неперекрываемые точки при разных значениях абсциссы. Таким образом, можно сформулировать следующее **условие единственности разложения (02)**: **в каждой паре спектров компонентов должны существовать по меньшей мере две точки, в одной из которых ненулевую интенсивность имеет только первый компонент, в другой - только второй.** Набор спектров компонентов при этом представляет собой совокупность векторов, представляющих собой линии пересечения гиперплоскости, образованной векторами-спектрами смесей (в дальнейшем будем называть ее **гиперплоскостью смесей**), с координатными плоскостями. **Необходимость** условия доказывается единственностью такого пересечения, **достаточность** – существованием однозначной процедуры поиска пересечения.

В случае невыполнения таких условий неперекрываемости для обеспечения единственности необходима дополнительная информация, чтобы выбрать в областях допустимых решений векторы, которые являлись бы спектрами компонентов. Геометрическое представление на Рисунке 1.1 также наглядно демонстрирует принцип построения областей допустимых решений как гиперплоскостей, ограниченных с одной стороны координатными плоскостями (условие неотрицательности спектров) и с другой стороны выпуклым в сечении конусом, охватывающим набор векторов-спектров смесей (условие неотрицательности концентраций). На Рисунке 1.1 такой конус соответствует набору из двух векторов  $d_1$  и  $d_4$ . Дополнительная информация о задаче нужна для того, чтобы выбрать в областях существования решений векторы, которые являлись бы спектрами компонентов при невыполнении условий неперекрываемости.

Требование аддитивности спектров смесей означает, что форма спектров компонентов не должна зависеть от присутствия других компонентов

(компоненты не должны взаимодействовать, быть **независимыми**) и от условий съемки спектров. Если это условие не соблюдается, число компонентов будет превышать число смесей при любом их количестве и задача будет недоопределенной. Однако и в этом случае разделение при помощи рассмотренных далее процедур может давать полезную информацию о компонентном составе.

## 1.3 Методы определения числа компонентов по матрице экспериментальных данных

### 1.3.1 Литературный обзор

Процедуры поиска независимых компонентов не требуют, в общем случае, знания числа составляющих. Однако, задание этого числа позволяет при наличии значительных шумов стабилизировать решение, с одной стороны, и проверить исследуемую систему на независимость компонентов, с другой. Проверка на независимость заключается в проверке условия, что число смесей строго больше числа компонентов,  $M > K$ . Разумеется, выполнение этого условия не гарантирует независимость всех компонентов, и получаемые решения могут содержать их большее количество, часть индивидуальных спектров будет отражать именно изменения компонентов при взаимодействии, что часто также можно считать полезной информацией.

Предложено довольно много способов оценки числа компонентов  $K$  по матрице спектров родственных смесей  $D$ . Эти способы, как правило, основаны на знании величины экспериментальных ошибок и являются, по существу, статистическими [1,52-55]. Кроме того, предложено несколько нестатистических методов [56-59], которые позволяют получать на практике полезные результаты. При анализе реальной задачи представляется совершенно необходимым использовать несколько оценок, т.к. статистические оценки могут оказаться неверными вследствие использования ошибочных статистических гипотез, касающихся экспериментальных ошибок, а нестатистические оценки не имеют, как правило, достаточного обоснования. Большинство известных методик оценки  $K$  основаны на факторном анализе матрицы исходных данных, который позволяет выделять набор "значимых факторов", число которых и есть число компонентов. Так, в работе [1] делается больший упор на практическое использование факторного анализа, а в [55] - на статистический аспект методов.

Рассмотрение оценки числа факторов - компонентов будет нагляднее, если

его вести с точки зрения сингулярного разложения матрицы данных:

$$D = U \cdot \Lambda \cdot V^T \quad (1.2)$$

где  $U$  - это  $N \times M$  матрица собственных векторов матрицы  $D \cdot D^T$  (в литературе по факторному анализу - ковариационной матрицы),  $V$  - матрица собственных векторов  $D^T \cdot D$  (или корреляционной матрицы),  $\Lambda$  - диагональная матрица сингулярных чисел (положительных квадратных корней из ненулевых собственных значений матрицы  $D \cdot D^T$  или  $D^T \cdot D$ ), расположенных в невозрастающем порядке [60]. Для определенности здесь и далее будем считать, что число спектральных точек  $N$ , количество смесей  $M$  и компонентов  $K$  удовлетворяют неравенству

$$N > M > K, \quad (1.3)$$

хотя достаточными являются неравенства

$$N \geq M \quad \text{и} \quad M \geq K. \quad (1.4)$$

Для облегчения систематизации методов оценки  $K$ , обзор известных и

предлагаемых нами методов будем проводить с использованием разложения (1.2).

Согласно [1], представим матрицу экспериментальных данных  $D$  в виде суммы невозмущенной матрицы  $D^*$  и матрицы экспериментальных ошибок  $F$ :

$$D = D^* + F. \quad (1.5)$$

При условии (1.3) ранг  $D^*$  равен  $K$  (если векторы спектров компонентов линейно независимы и матрица относительных концентраций имеет полный ранг  $K$  - эти условия в дальнейшем также будут предполагаться выполненными), ранг  $D$  в силу некоррелированности экспериментальных ошибок равен числу смесей  $M$ . Следовательно, ранг  $D$  также равен  $M$ .

В свою очередь, матрицу ошибок можно представить [1] в виде

$$F = F^\# + F^0, \quad (1.6)$$

где  $F^\#$  образована частью ошибок, которые дают вклад в первые  $K$  сингулярных

векторов  $U$ , а  $F^0$  образована той их частью, которая порождает ненулевые сингулярные числа  $\lambda_i$ ,  $i = M+1, \dots, M$  и соответствующие им вектор-столбцы  $u_i$ . Из (1.5) и (1.6) получаем:

$$\begin{aligned} D^\# &= D^* + F^\# \\ D &= D^\# + F^0 \end{aligned} \quad (1.7)$$

Матрица  $F^0$  обычно называется остаточной ошибкой [1], которую можно отбросить, восстанавливая матрицу данных  $D^\#$  по первым сингулярным векторам исходной матрицы  $D$  согласно соотношению (1.2). Следует заметить, что исходные ошибки распределены по матрицам  $D^\#$  и  $F^0$ , т. е. последняя образована только некоторой их частью.

Первая группа методов нахождения числа компонентов, по сути, сводится к нахождению расщепления (1.7) (анализ остатков). На первом этапе вычисляют разложение (1.2) и затем разделяют сингулярные векторы на две группы, называемыми в факторном анализе первичными и вторичными осями. Число первичных осей при этом равно оценке числа компонентов и они являются первыми  $K$  сингулярными векторами, которые образуют базис пространства столбцов (в случае  $u$ -векторов) или строк (в случае  $v$ -векторов) матрицы  $D^\#$ . Запишем соотношение (1.7) в терминах сингулярного разложения: если  $Q = U \cdot S$ , то

$$D = Q \cdot V^T = Q^\# \cdot V^{\#T} + Q^0 \cdot V^{0T}, \quad (1.8)$$

где  $Q^\#$  - матрица первичных осей размера  $N \times M$ , у которой столбцы  $K+1, \dots, M$  нулевые, а  $Q^0$  - матрица  $N \times M$  вторичных осей с нулевыми столбцами  $1, 2, \dots, K$ . Тогда, согласно [1], первый метод оценки числа компонентов (Residual Standard Deviation) будет заключаться в вычислении выражений

$$RSD = \sum_{i=1}^N \sum_{k=K+1}^M (q_{ik}^0)^2, \quad (1.9)$$

где  $q_{ik}^0$  - элементы матрицы  $Q^0$ ;

$$RSD = \left( \frac{\sum_{k=K+1}^M \lambda_k^2}{N \cdot (M - K)} \right)^{1/2}, \quad (1.10)$$

( $\lambda_k$  - сингулярные числа), что эквивалентно (1.9). Вычисления проводят, изменяя гипотезу о числе компонентов  $K$  начиная с  $K = 1$  и сравнивая соответствующую величину  $RSD$  с величиной оценки экспериментальной ошибки  $\sigma_{\text{э}}$ . То значение  $K$ , которое приводит к величине  $RSD$ , приблизительно равной оценке ошибки  $\sigma_{\text{э}}$ , и считают числом компонентов в смесях.

Наша практика показала, что лучше иметь завышенную оценку числа компонентов, поэтому вычисления надо продолжать до тех пор, пока  $RSD$  не станет строго меньше экспериментальной ошибки).

В работе [52] предложен сходный с рассмотренным метод, названный Residual Standard Deviation Method, основанный на вычислении

$$\sigma_K = \left( \frac{\text{trace}(D^T \cdot D) - \sum_{k=1}^K \lambda_k^2}{N - K} \right)^{1/2}, \quad (1.11)$$

где  $K$  также принимает увеличивающиеся значения, а величина  $\sigma_K$  сравнивается с  $\sigma_{\text{э}}$ . Наибольшую сложность, конечно, вызывает корректное оценивание экспериментальной ошибки  $\sigma_{\text{э}}$ , которая складывается не только из случайных шумов с известной дисперсией, но и содержит на практике неконтролируемую долю систематических погрешностей. Более того, рассмотренные оценки разработаны для случая нормально распределенных шумов измерений. Как показано в [23,24,64], распределения измерений для различных приборов существенно отклоняются от нормальных. Например, для ИК-спектров распределение спектроскопических ошибок в зависимости от поглощения можно представить в виде [23]:  $A_i = -\text{Log}(G \cdot W_i \cdot e^{-A_o} + H) + \text{Log}(W_i)$ , где  $A_i$  - измерение

поглощения при  $i$ -ой длине волны,  $W_i$  - принимаемая детектором мощность,  $A_0$  - общее поглощение,  $G$  - модуляционный оптический шум (считают, что он распределен лог-нормально со средним значением 1,  $H$  - шум детектора, который обычно имеет распределение, близкое к нормальному и задается в виде эквивалентной мощности шумов для данного детектора. Из этих работ видно, что по вопросу характера распределения спектроскопических шумов нет единого мнения. По-видимому, перед использованием статистических методов оценки числа компонентов в каждом конкретном случае следует проводить специальное исследование природы и параметров шумов. Из трудно учитываемых вкладов в ошибки можно указать искажения спектров вследствие межмолекулярных взаимодействий, рост шумов при двухлучевой компенсации полос поглощения растворителя, шумы источника излучения, искажения спектров из-за дрейфа спектрального прибора по длинам волн и дрейфа нулевой линии. Некоторые из этих ошибок определить очень трудно, что ограничивает ценность статистических методов определения числа компонентов и вынуждает нас использовать различные полуэмпирические методики.

Можно показать, что  $RSD$  оценка связана с общей квадратичной невязкой между исходной  $D$  и восстановленной  $D^\#$  по (1.7) матрицами данных:

$$RSD = \left( \frac{\sum_{i=1}^N \sum_{j=1}^M (d_{ij} - d_{ij}^\#)^2}{N \cdot (M - K)} \right)^{1/2}. \quad (1.12)$$

Величина  $RSD$  используется в часто применяемой Factor Indicator Function ( $IND$ ) [1,57,58]:

$$IND = \frac{RSD}{(M - K)^2}, \quad (1.13)$$

которую вычисляют, изменяя оценку  $K$  от 1 до  $M$ . Кривая зависимости величины  $IND$  от  $M$  должна достигать минимального значения при верной оценке  $K$ .

Кажется наиболее естественным метод оценивания  $K$ , основанный на

сравнении дисперсии элементов матрицы остатков  $F^0$  с дисперсией экспериментальных ошибок. Этот метод реализован в некоторых пакетах программ, но на практике он часто приводит к неверным оценкам, так как критерий равенства дисперсий двух совокупностей случайных величин, использующий F - распределение Фишера-Снедекора чрезвычайно чувствителен к нарушениям нормальности распределения [65]. В свою очередь, как отмечалось выше, нет никаких оснований считать распределение ошибок в матрице данных нормальным. Для применения F -критерия сначала необходимо доказать гипотезу о нормальности закона распределения с помощью критерия согласия (Пирсона, Колмогорова, Смирнова и т.п. [66]). Поэтому особую ценность приобретают такие статистические методы оценки числа компонентов, которые являются свободными от знания закона распределения элементов матрицы ошибок или устойчивыми к отклонениям от нормальности распределения.

Естественным требованием для матрицы остатков является отсутствие корреляции вдоль последовательности ее элементов. В данной работе в качестве поверочной статистики остатков мы выбрали критерий Дарбина-Ватсона [67], который вычисляется по формуле :

$$DW = \frac{\sum_{i=2}^{N \cdot M} (F_i^0 - F_{i-1}^0)^2}{\sum_{i=1}^{N \cdot M} (F_i^0)^2}, \quad (1.14)$$

где  $F_i^0$  представляет собой последовательный элемент матрицы, превращенной в одномерный массив размера  $M \cdot N$  путем последовательного соединения векторов-столбцов:  $i = (1,1); (1,2); \dots (1,M); (2,1); (2,2); \dots (2,M); \dots (N,M)$ . Величина критерия лежит в области  $\{0 - 4\}$ , принимая значения около 2.0 в случае отсутствия корреляции. Так, при уровне значимости гипотезы о наличии корреляции между последовательными элементами массива 0.05, ее отвергают, если величина критерия лежит в диапазоне  $\{1.7 - 2.2\}$ . На практике приходится допускать наличие слабой остаточной корреляции из-за нарушения аддитивности

спектров при их взаимодействии в смесях или неучтенных слабых систематических ошибок измерений и в большинстве случаев границу необходимо немного расширить, например, до  $\{1.5 - 2.5\}$ , обосновывая это расширение эмпирически на основании опыта решения аналогичных задач из данной предметной области. Такое расширение, конечно, не является строгим, а относится к экспертным оценкам и рассматривается также в (Бородич, С. А. Вводный курс эконометрики: Учебное пособие / С. А. Бородич. – Мн.: БГУ, 2000. – 354 с.).

Другая группа методов оценки основывается на анализе относительных величин собственных значений корреляционной  $D^T \cdot D$  или ковариационной  $D \cdot D^T$  матриц или, в терминах разложения (1.2), сингулярных чисел матрицы данных  $D$ . Авторы [8,68] за число компонентов принимают число наибольших собственных значений, которые описывают 99% дисперсии спектров смесей (метод главных компонент). Разумеется, пороговое значение этого критерия должно зависеть от уровня экспериментальных шумов.

Еще один критерий основан на вычислении дисперсии собственных значений матриц  $D \cdot D^T$  или  $D^T \cdot D$ . По определению сингулярного разложения (1.2), эти собственные значения есть квадраты соответствующих сингулярных чисел.

Стандартная ошибка в  $m$ -ом собственном значении вычисляется в виде [1,52]:

$$\sigma_m = \left( \sum_{j=1}^M \sum_{k=1}^M v_{mj}^2 \cdot v_{mk}^2 \cdot \sigma(Z)_{jk}^2 \right)^{1/2}, \quad (1.15)$$

где  $v_{mj}^2$  и  $v_{mk}^2$  -  $j$ -й и  $k$ -й элементы  $m$ -го сингулярного вектора, а

$$\sigma(Z)_{jk}^2 = \begin{pmatrix} \sum_{i=1}^N (d_{ij}^2 \cdot \sigma_{ik}^2 + d_{ik}^2 \cdot \sigma_{ij}^2), & j \neq k \\ \sum_{i=1}^N (4d_{ij}^2 \cdot \sigma_{ij}^2), & j = k \end{pmatrix}^{1/2}, \quad (1.16)$$

$\sigma_{ij}$  обозначает ошибку в экспериментальной точке  $d_{ij}$  и определяется в эксперименте. Числом компонентов считают число собственных значений, превосходящих по величине оценку (1.15) при заданном уровне значимости). Этот метод позволяет, как видно из (1.16), учитывать оценку ошибки в каждой из спектральных точек матрицы данных  $D$  индивидуально.

В тех случаях, когда среднее стандартное отклонение меняется от точки к точке в спектре и не является константой для данной матрицы данных, предложено применять хи-квадрат критерий Бартлета [1,52]. Недостаток этого метода (как и предыдущего) в том, что необходимо иметь достаточно справедливую оценку ошибки в каждой экспериментальной точке спектра. Если такие оценки  $\sigma_{ik}^2$  получены, то  $\chi_N^2$  определяют как

$$\chi_N^2 = \sum_{i=1}^N \sum_{j=1}^M \frac{(d_{ij} - d_{ij}^{\#})^2}{\sigma_{ij}^2} \quad (1.17)$$

где  $d_{ij}^{\#}$  - по-прежнему элементы матрицы данных, восстановленной по первым  $K$  сингулярным векторам (1.2). Число компонентов  $K$  определяют, изменяя его, начиная с 1, вычисляя для каждого пробного значения  $K$  величину  $\chi_N^2$  и сравнивая ее с ожидаемым значением  $\chi_{\alpha}^2$ ,

$$\chi_{\alpha}^2 = (N - K)(M - K), \quad (1.18)$$

где  $(N - K)(M - K)$  есть число степеней свободы,  $\alpha$  - уровень значимости.

Выбирают такое значение  $K$ , при котором  $\chi_N^2$  ближе всего к  $\chi_{\alpha}^2$ .

С практической точки зрения интересны методы, основанные на анализе спектра сингулярных чисел  $\lambda_j$  матрицы  $D$  (или собственных значений

ковариационной матрицы [56-58]). В работах [25,56,57,69] показано, что при нормальном распределении ошибки и постоянной дисперсии, сингулярные числа матрицы ошибок  $F$  из (1.5) лежат приблизительно на прямой, если расположить их в убывающем порядке на графике в координатах величина числа – порядковый номер. Среднее квадратичное отклонение от этой прямой предложено оценивать по формуле [25]:

$$\xi_0 = 0.3 \cdot \sigma_D^2 \cdot \sqrt{N \cdot (M - 2)}, \quad (1.19)$$

где  $\sigma_D^2$  - интегральная оценка дисперсии ошибок в исходных данных. К сожалению, во многих случаях дисперсия есть функция, значение которой зависит от интенсивности в спектрах и данный критерий оказывается неприменим. Отклонение собственных значений от прямой на графике становится тем более значительным, чем больше зависимость  $\sigma$  от  $d_{ij}$ . Число первых "выпавших" из линейной зависимости больших сингулярных чисел принимают за число компонентов.

Среди методов оценки числа компонентов, не требующих знания величины экспериментальной ошибки, следует отметить часто используемую Imbedded Error Function ( $IE$ ) [1,57]:

$$IE = \left( \frac{K \cdot \sum_{j=K+1}^M \lambda_j^0}{N \cdot M \cdot (M - K)} \right)^{1/2}, \quad (1.20)$$

которую вычисляют, изменяя оценку  $K$  от 1 до  $M$ . Кривая зависимости величины  $IE$  от  $K$  имеет вид, аналогичный зависимости  $\text{Log}(\lambda_j)$  от номера  $j$ , но число компонентов соответствует точке излома на кривой (точка резкого изменения угла наклона кривой относительно оси абсцисс).

Интересен метод определения числа компонентов по результатам частотного анализа левых сингулярных векторов  $D$ , представленных в виде спектральных контуров. В работах [41,59] отмечается, что сингулярные векторы, соответствующие "значимым" сингулярным числам, содержат преимущественно

низшие гармоники (в спектре Фурье), а "шумовые" векторы представляют высшие гармоники, соответствующие шумам. Это, безусловно, верно только в том случае, если шумовые составляющие более высокочастотны по своей природе, чем спектроскопическая информация. Поэтому при оценке числа компонентов этим методом важно, чтобы оцифровка спектров проводилась с достаточно малым шагом по длинам волн. Тогда шумы дискретизации будут превосходить по своей верхней частоте спектра Фурье высшую из гармоник спектров компонентов.

Для целей определения числа компонентов, конечно, лучше всего иметь оценку такой высшей Фурье-частоты  $y_h$  для, например, косинусного преобразования:

$$f(y) = \int_0^{y_{\max}} u(\nu) \cdot \cos(y \cdot \nu) \cdot d\nu, \quad (1.21)$$

где  $f(y)$  - спектр Фурье,  $y_{\max}$  - высшая частота спектра,  $y_{\max} > y_h$ . Такая оценка позволяет эмпирически установить предел (для конкретной задачи или класса задач)  $y_{\lim} = y_h$  для вычисления отношения:

$$R = \frac{\int_0^{y_h} |f(y)| dy}{\int_0^{y_{\max}} |f(y)| dy}, \quad (1.22)$$

в котором интегрирование можно заменить простым суммированием абсолютных значений амплитуд Фурье [59]) для каждого из собственных векторов  $u_j(\nu)$ . Если шумовые компоненты более высокочастотны, чем спектроскопические, то, начиная с некоторого  $u_j(\nu)$ , величина  $R$  скачком уменьшает свое значение. Номер последнего сингулярного вектора  $u_k$ , имеющего большое значение  $R$  и принимают за число компонентов [59]. При таком анализе сингулярные векторы располагают соответственно убыванию сингулярных чисел. При использовании этого метода нужно иметь физические основания для утверждений о "разночастотности" спектров компонентов и спектров помех.

Таким образом, существует большое разнообразие методов оценивания числа компонентов, одновременное применение которых позволяет повысить надежность ответа путем сравнения результатов, полученных разными способами. Сам факт несовпадения оценок может нести полезную информацию. Так, заниженное число компонентов, полученное нестатистическими методами по сравнению со статистическими, может служить свидетельством того, что преобладающими являются ошибки, обусловленные, например, временной нестабильностью прибора (дрейф по длинам волн и т.п.); межмолекулярными взаимодействиями в смесях, приводящими к изменению спектров компонентов и т.д. В нашей работе мы применяли как все рассмотренные способы оценивания  $K$ , так и разработанные методы, описываемые в следующей главе.

В дополнение к литературному обзору по методам оценки числа компонентов следует указать еще два других известных критерия проверки известных спектров на присутствие в спектрах смесей [70-72], основанные на теории ошибок в факторном анализе: Reliability Function (RELI) и Spoil Function (SPOIL). Данные критерии, разработанные для целей библиотечного поиска компонентов, показывают удовлетворительные результаты и при оценивании числа компонентов методом проектирования известного спектра компонента на пространство столбцов матрицы данных  $D^\#$ , восстановленной по  $K$  первым сингулярным векторам разложения (1.2). Эти критерии используют величину "кажущейся" (apparent) ошибки в векторе пробного спектра компонента  $t$ :

$$AET = \left( \frac{\sum_{i=1}^N (p_i - t_i)^2}{N} \right)^{1/2},$$

где  $p$  – вектор проекции  $t$  на  $D$  (или  $D^\#$ ) – см. Рисунок 1.1 – мы предлагаем вычислять с помощью SVD-разложения (1.2):

$$p = U \cdot U^T t, \quad \text{или} \quad p = U^\# \cdot U^{\#T} t, \quad (1.23)$$

где знак (#) означает матрицу, составленную из первых  $K$  столбцов  $U$ , а  $K$  – предполагаемое (проверяемое) число компонентов.

Величину АЕТ сравнивают с "действительной" (real) ошибкой в векторе проекции  $\mathbf{p}$  ( $\lambda_j$  - сингулярные числа из (1.2)):

$$REP = \left( \frac{\sum_{j=1}^M \lambda_j^2}{\max(N, M) \cdot (M - K)} \right)^{1/2}, \quad (1.24)$$

$$RET = (AET^2 - REP^2)^{1/2}.$$

Критерии для принятия решения о приемлемости пробного вектора в качестве компонента (а в случае оценки числа компонентов – о приемлемости значения  $K$  при условии, что пробный спектр  $\mathbf{t}$  заведомо содержится в спектрах смесей) вычисляют по формулам:

$$SPOIL = \frac{RET}{REP}, \quad RELI = \left( \frac{RET^2 - RET_{est}^2}{AET^2} \right)^{1/2}, \quad (1.25)$$

где  $RET_{est}^2$  - оценка ошибки для данных, которые использовали для построения тестового вектора. В [70-72] предложены следующие пороговые значения:  $RELI > 0.5$  означает приемлемость  $\mathbf{t}$  в качестве спектра компонента, так же как и  $SPOIL < 3.0$ .  $SPOIL = 3.0 \dots 6.0$  указывает на возможность того, что  $\mathbf{t}$  есть спектр компонента,  $SPOIL > 6.0$  означает отрицательный ответ. Несмотря на кажущуюся строгость теории ошибок в [70-72], основанной на "соотношении треугольника":

$$AET = RET^2 - REP^2,$$

это равенство не подкрепляется никакими статистическими выводами и вводится искусственно. По существу, это соотношение означает, что ошибки  $RET$  и  $REP$  мы считаем независимыми. Только в этом случае их дисперсии можно суммировать. Тем не менее, практика показывает (см., например, [73]), что рассмотренные критерии позволяют получать неплохие результаты.

В большинстве практических случаев предложенные статистические оценки оказываются несостоятельными и смещенными из-за отклонений реальных ошибок измерений от закона распределения, для которого эти оценки построены. Ситуация усложняется еще и тем, что в данных, как правило, присутствуют систематические ошибки, связанные как с дрейфами аппаратуры, так и с нарушениями аддитивности спектров (например, вследствие химического взаимодействия компонентов). В этом случае оценка числа компонентов должна представлять, по существу, экспертную оценку, для построения которой исследователю необходимо располагать дополнительной информацией. Кроме того, некоторые из предложенных критериев являются оценками сверху, то есть, как правило, значительно завышенными.

В ряде работ [52-55] число компонентов предложено определять по графику отсортированных в порядке убывания собственных значений (или квадратных корней из них) ковариационной матрицы спектроскопических данных  $DD^T$ . В терминах сингулярного разложения (1.2) это соответствует анализу графика сингулярных чисел. В [52] утверждается, что согласно оценке (1.11) часть графика, состоящая из "шумовых" собственных значений в случае нормального распределения шумов в данных имеет вид линейно спадающей зависимости.

### 1.3.2 Методика приготовления тестовых задач для спектроскопического анализа

Для демонстрации численных методов оценки числа компонентов и их спектров по матрице спектров родственных смесей были приготовлены несколько модельных смесей органических соединений. Для получения родственных спектров использовали такие методы разделения, как вакуумная перегонка и жидкостная хроматография. Хроматограф представлял собой стеклянную колонку размером 250x3,5 мм, заполненную сорбентом Силасорб 600 (LC) 30 мкм, оборудованную системой впуска пробы шприцем и создания давления (азотный баллон с редуктором на избыточное давление 0-0,4 мПа). Конструктивная схема рассматриваемого прибора предельно проста и ее рисунок мы здесь не приводим, так как она не имеет принципиального значения. В качестве детектора использовали ИК спектрометр Perkin Elmer 580В, сопряженный с микро-ЭВМ Interdata 6/16.

Спектры смесей-фракций снимали в проточной кювете из КВг толщиной 0,15 мм, останавливая поток растворителя игольчатым вентилем примерно через равные промежутки времени. Уровень спектроскопических шумов составлял 0,4%Т. Полученные спектры сглаживали с помощью стандартных программ и компенсировали в них спектр растворителя (четырёххлористого углерода, который фактически представлял собой известный компонент) с помощью кюветы переменной толщины, помещенной в канал сравнения спектрометра, или программой вычитания спектральных контуров. Исходные данные (матрицу спектров смесей) формировали, кодируя наиболее информативные участки в спектрах смесей с переменным шагом. Матрицу относительных концентраций компонентов определяли по площадям соответствующих хроматографических пиков, полученных на газовом хроматографе ЛХМ 8МД (3%Е-30 на панхроме 3, длина колонки 3м, температура 120°C) с погрешностью 3,5% отн. [А1,А4]).

Работу численных методов, рассматриваемых в данной работе, мы преимущественно проверяли на модельных смесях гексана, толуола, ацетона,

циклогексанона, диэтилового эфира и амилацетата. Хроматограммы записывали при длине волны  $1450 \text{ см}^{-1}$ , где поглощение невелико, но примерно одинаково для всех компонентов (пример типичной хроматограммы показан на Рисунке 1.2). Давление растворителя при хроматографировании было  $0,4 \text{ мПа}$ , расход  $0,75 \text{ мл/мин}$ . Спектры смесей-фракций записывали в диапазоне  $1750\text{-}1110 \text{ см}^{-1}$  с разрешением не хуже  $5 \text{ см}^{-1}$ . После этого при подготовке матрицы спектроскопических данных  $D$  были отброшены малоинтенсивные участки и оставлены области  $1750\text{-}1690$ ,  $1610\text{-}1595$ ,  $1510\text{-}1300$ ,  $1160\text{-}1110 \text{ см}^{-1}$ .

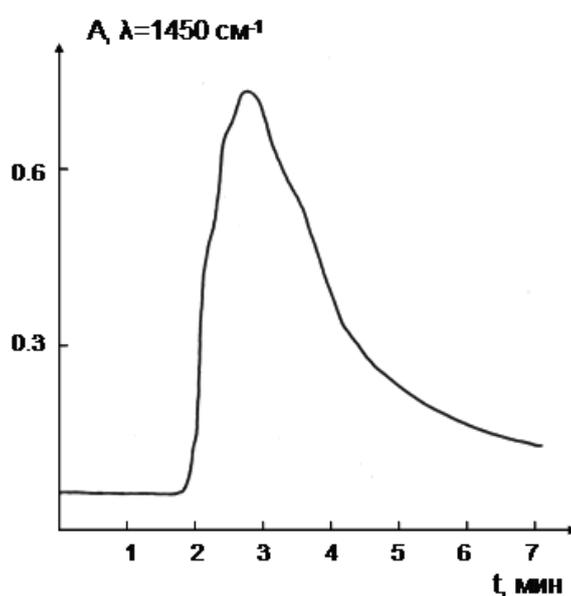


Рисунок 1.2 – Типичная хроматограмма 5-компонентной смеси гексана, толуола, ацетона, циклогексанона и диэтилового эфира при длине волны поглощения  $1450 \text{ см}^{-1}$ . Гладкость пика говорит о плохом разделении компонентов во времени.

Оцифровку спектров проводили с фиксированным шагом  $5 \text{ см}^{-1}$ . На Рисунке 1.3 показаны типичные спектры исходной смеси №1 после кодирования. От каждой смеси получали 10-12 спектров фракций. Так как для проверки работоспособности методов разложения параметры задачи желательно знать как можно точнее, то концентрационную матрицу получали не с помощью газовой хроматографии, а рассчитывали по уравнению (1.1), используя снятые на этом же спектрометре спектры чистых компонентов в качестве матрицы коэффициентов.

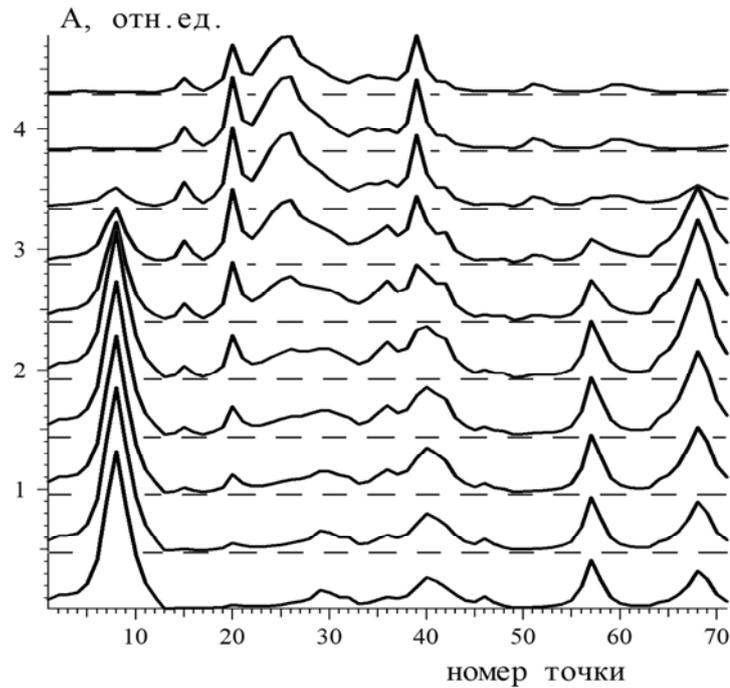


Рисунок 1.3 – ИК спектры 5-компонентной смеси №1 после оцифровки и выбрасывания. Диапазон спектра  $1750-1110 \text{ см}^{-1}$ . По оси абсцисс отложены порядковые номера спектральных точек (т.к. величина длины волны при разложении не учитывается, спектры оцифрованы с переменным шагом с целью исключения неинформативных участков).

Спектры смещены по вертикали для лучшей визуализации, пунктиром обозначены уровни нулевого поглощения.

Вторая (**№2**) модельная задача представляла собой смесь шести компонентов: изооктана, толуола, бензола, диэтилового эфира, ацетона и амилацетата в четыреххлористом углероде как неполярном растворителе. В этом наборе относительно сильно перекрытыми оказываются две пары компонентов: ацетон - амилацетат и бензол - толуол (см. Рисунок 1.13). Кроме того, эти вещества плохо разделяются при данных условиях хроматографирования. Набор спектров поглощения показан на Рисунке 1.4.

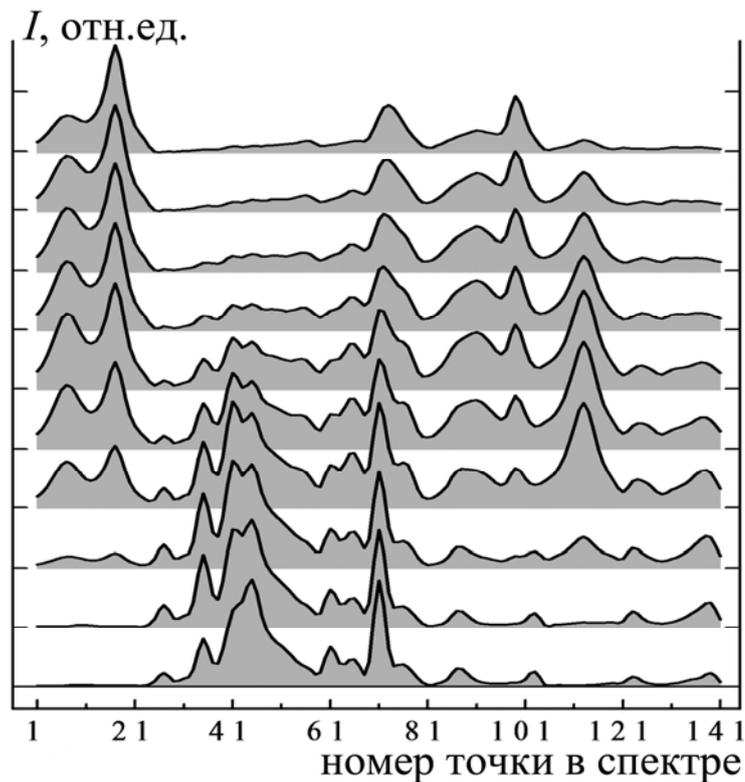


Рисунок 1.4 – ИК спектры фракций 6-компонентной смеси (№2) изооктана, толуола, бензола, диэтилового эфира, ацетона и амилацетата, полученные на выходе препаративного жидкостного хроматографа конструкции автора. По оси абсцисс отложены порядковые номера спектральных точек (т.к. величина длины волны при разложении не учитывается, спектры оцифрованы с переменным шагом с целью исключения неинформативных участков). Спектры смещены по вертикали для лучшей визуализации.

Третья модель (№3) была составлена из гексана, изооктана, бензола, диоксана, диэтилового эфира и ацетона. Смеси-фракции (10 фракций) получали путем перегонки с дефлегматором. Определение числа компонентов по матрице (Рисунок 1.5б) дало оценку 4 вместо 6.

Четвертая модель (№4) представляла собой смесь ацетона, диэтилового эфира, циклогексанона, и толуола. Фракционирование также проводили на жидкостном хроматографе, но с худшим разрешением из-за большего диаметра колонки (5 мм).

Спектры смесей №3 и №4 мы здесь не приводим, так как они не содержат какой-либо новой полезной информации для дальнейшего изложения.

Смеси подвергали хроматографированию и съемке несколько раз, меняя скорость элюации от 0.1 мл/мин до 1 мл/мин.

### 1.3.4 Анализ спектра сингулярных чисел и матрицы остатков

В наших работах [A1-A5] было показано, что при исследовании смесей по ИК и видимым спектрам, на практике оценку числа компонентов по спектру сингулярных чисел лучше проводить не в абсолютных значениях, а в логарифмическом масштабе. Это лишает метод строгого статистического обоснования, рассмотренного в [41,52,59], делает его эмпирическим, но дающим на практике приемлемые результаты. Причина использования логарифмической шкалы заключается в том, что шумы данных реального эксперимента часто распределены не по нормальному или какому-либо другому статистическому закону, а содержат неконтролируемые ошибки измерений (например, дрейф и скачки параметров оборудования, взаимодействия компонентов в смесях, вносящие систематическую составляющую в данные). Эта систематическая составляющая нарушает идеальную аддитивность данных и выражается в виде малоинтенсивных компонентов, интенсивность которых уже не обязана быть неотрицательной. Наличие автокорреляций между точками в контурах интенсивности таких "компонентов", естественно, приводит к увеличению значений малых сингулярных чисел матрицы данных и к оценке числа компонентов, равным всегда числу смесей. В этом случае хорошие результаты дает анализ контуров левых сингулярных векторов матриц смесей, рассматриваемый в разделе 1.3.5.

Практика автора показала, что часто систематическая составляющая ошибок относительно невелика, но закон распределения случайной составляющей оказывается формально близок к лог-нормальному. При лог-нормальном распределении шумов в  $D$  линейный характер зависимости сингулярных чисел  $\lambda$  приблизительно выполняется для координат  $\log(\lambda_k) - k$ , где  $k$  – порядковый номер числа, хотя даже для этого случая известного закона распределения получить оценку вида (1.18) представляется затруднительным. Тем не менее, число компонентов можно получить, подсчитав количество сингулярных чисел, расположенных выше прямой, проведенной через набор малых  $\log(\lambda_k)$ . По

характеру излома на графике при переходе от "шумовых"  $\log(\lambda_k)$  к "значимым" можно судить о характере исходной спектроскопической информации: чем увереннее можно проводить расщепление на значимые и шумовые компоненты, тем лучше обусловлена задача разложения. На Рисунке 1.5 приведены примеры спектров сингулярных чисел для модельных и реальных наборов данных.

Второй предложенный нами ранее способ оценки числа компонентов [A3,A5] основан на статистических критериях, работающих не с самими элементами выборок, а с их рангами, что делает оценки независимыми от закона распределения [75-78]. Ранговые критерии инвариантны по отношению к любому монотонному преобразованию шкалы измерения и пригодны к анализу разномасштабных данных. Один из наиболее широко используемых ранговых критериев связывают с именами Манна-Уитни [79] и Вилкоксона [80]. У-критерий Манна-Уитни (Mann-Whitney U test) — непараметрический статистический критерий, используемый для оценки различий между двумя выборками. С его помощью, например, проверяют гипотезу об идентичности законов распределения двух совокупностей независимых случайных величин. Этот критерий можно использовать непосредственно как критерий числа компонентов, проверяя гипотезу для выборки элементов матрицы остатков  $F^0(K)$ , которую получают, вычисляя  $D^\#$  при разных значениях  $K$  (см. уравнения (1.6, 1.9) выборки матрицы экспериментальных ошибок. Последнюю можно получить, например, путем вычитания спектров, записанных с некоторым временным интервалом. За оценку  $K$  принимают то минимальное значение, для которого гипотезу об идентичности законов распределения уже можно считать верной при выбранном уровне значимости. Другими словами,  $K$  полагают равным минимальному количеству сингулярных чисел, которому соответствует матрица остатков  $F^0$  с элементами, имеющими тот же закон распределения, что и ошибки в  $D$ .

Вначале все  $n_1$  элементов  $F^0(K)$  и  $n_2$  элементов массива экспериментальных ошибок объединяют в один вектор  $x$  в порядке убывания. Затем формируют вектор  $y$  из  $n_1$  чисел, представляющих собой порядковые

номера элементов  $F^0(K)$  в векторе  $x$  и вычисляют сумму

$$S = \sum_{i=1}^{n_1} y_i \quad (1.26)$$

По таблицам [75] для данных  $n_1$  и  $n_2$  и выбранного уровня значимости находят критическую область ( $S_{\min}$  и  $S_{\max}$ ). Попадание величины  $S$  в найденный интервал свидетельствует о том, что законы распределения одинаковы. Для ИК спектров 6-компонентных смесей, показанных на Рисунке 1.4, величина  $S$  для гипотезы о 4-х компонентах равна 194, что выходит за пределы интервала (348, 472) при 5%-ном уровне значимости (из всей совокупности данных были отобраны случайным образом по 20 последовательных точек для вектора остатков и вектора экспериментальных ошибок). Для гипотез о 5, 6 и 7 компонентах  $S$  было равно соответственно 429, 394 и 414. Это может свидетельствовать о наличии ошибок, связанных с дрейфом прибора и приводящих к тому, что экспериментально оцененный вектор шумов содержит долю информации о спектрах веществ. Отсюда и приемлемость заниженной оценки, равной 5.

Небольшая модификация процедуры Вилкоксона позволяет создать критерий равенства дисперсий при условии равенства характеристик положения для двух совокупностей [75]. Поскольку средние значения  $F^0(K)$  и экспериментальных ошибок мы полагаем равными нулю, этот критерий применим к нашему случаю.

Модификация заключается в изменении способа приписывания рангов элементам вектора  $x$ : вместо последовательности 1,2,3,..., $n_1+n_2$  строят такую: 1,4,5,8,9,...,7,6,3,2 и из этой цепочки формируют вектор  $y$  из  $n_1$  рангов (порядковых номеров в  $x$ ) для выборки  $F^0(K)$ . Затем вычисляют  $S$  и оценивают критерий по таблицам, как и раньше. В рассматриваемом примере величины  $S$  равны 292, 362 и 440 для гипотез о 5, 6 и 7 компонентах. Видно, что приемлемой оказывается оценка  $K = 6$ , т.е. критерий равенства дисперсий в данном случае оказался более эффективным. Следует заметить, что рассмотренные критерии лучше применять в совокупности: за  $K$  принимают минимальную оценку, одновременно справедливую для обоих подходов. Если оценка

экспериментальных ошибок неизвестна, то число компонентов можно найти, проверяя гипотезу о равенстве нулю среднего значения элементов  $F^0(K)$ , если для этого есть достаточные основания.

Практика показала, что анализ матрицы остатков на наличие автокорреляции дает менее надежные результаты из-за влияния систематических погрешностей измерений. Графики зависимости величин критериев автокорреляции от предполагаемого числа компонентов имели в случае рассматриваемых тестовых систем существенно менее выраженный излом, чем графики логарифмов сингулярных чисел. Поэтому здесь мы эти данные не приводим. Но это не означает, что данный подход не следует применять в других случаях.

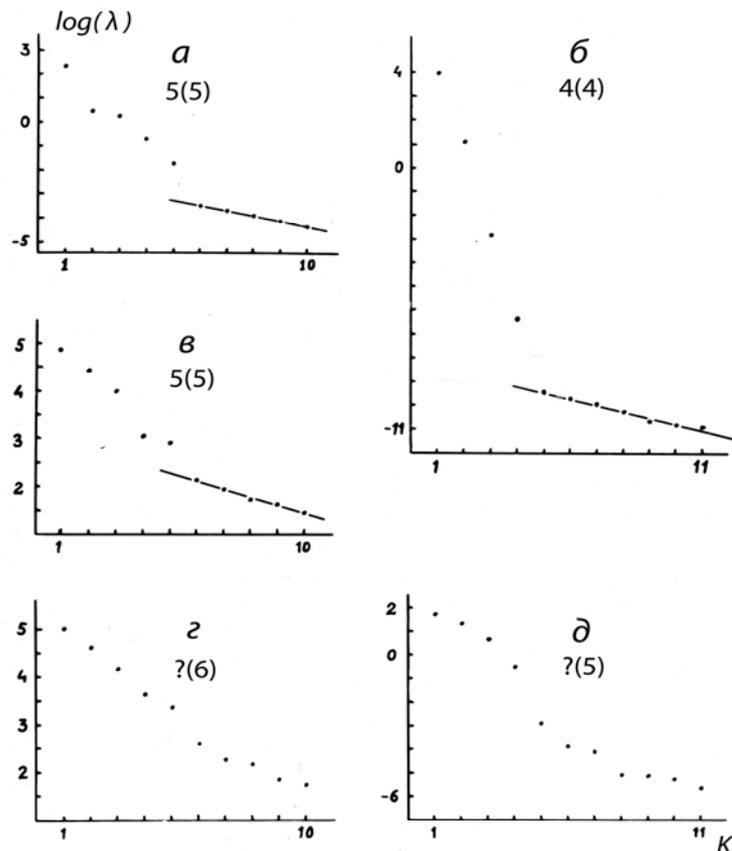


Рисунок 1.5 - Спектры логарифмов сингулярных чисел матриц ИК спектров тестовых смесей: а) модельная матрица данных при уровне спектральных шумов 0,2%Т, б) – смесь № 4, в) – №2, г) – №2, д) – №3 . На каждом графике указано найденное число поглощающих компонентов и априори известное (в скобках). По абсцисс отложены порядковые номера сингулярных чисел ( $K$ ). Знаки вопроса означают, что линейная зависимость отсутствует при  $R^2 > 0.7$ .

### 1.3.5 Анализ сингулярных векторов матрицы данных

Вклады спектров компонентов и спектроскопических шумов в сингулярные векторы  $U$  из (1.2) распределяются по ним неоднородно. Пусть  $u_j$  -  $j$ -й столбец матрицы  $U$ .  $u_1$  есть некоторая положительная линейная комбинация столбцов матрицы  $D$  (все сингулярные числа неотрицательны, как и элементы первого столбца матрицы  $V$  из (1.2)) и представляет собой некоторый "усредненный" спектр смеси, исчерпывающий максимум дисперсии данных в исходной матрице. По определению,

$$\mathbf{u}_j = \frac{1}{\lambda_j} D \cdot \mathbf{v}_j \quad (1.27)$$

$$\mathbf{v}_j = \frac{1}{\lambda_j} D^T \cdot \mathbf{u}_j$$

Кроме того, матрицы  $DD^T$  и  $D^TD$  симметричны и неотрицательны (т.к.  $d_{ij} \geq 0$  – мы рассматриваем случай неотрицательных спектров). Если в спектре смесей нет участков, в которых интенсивность равна нулю одновременно для всех спектров, то матрицы  $DD^T$  и  $D^TD$  являются неприводимыми. Согласно теореме Перрона Фробениуса [81], неотрицательные и неприводимые матрицы имеют положительное собственное значение  $\lambda_1$ , равное спектральному радиусу:

$$\eta_1 = \lambda_1^2 = \max_{\mathbf{x} \neq 0} \frac{|D \cdot D^T \cdot \mathbf{x}|}{|\mathbf{x}|} = \max \frac{|D^T \cdot D \cdot \mathbf{y}|}{|\mathbf{y}|}, \quad (1.28)$$

которое имеет кратность 1, и ему соответствует положительный собственный вектор  $\mathbf{u}_1$  (или  $\mathbf{v}_1$ , соответственно). Учитывая (1.27) можно сказать, что  $\mathbf{u}_1$  есть положительная линейная комбинация спектров смесей (столбцов  $D$ ), а  $\mathbf{v}_1$  - положительная линейная комбинация строк  $D$ . Далее, существует теорема, согласно которой при аппроксимации матрицы  $A$  матрицей  $A_k$  меньшего ранга  $k$  с использованием  $k$  первых сингулярных векторов:

$$A_k = U \cdot \Lambda \cdot V^T, \quad (1.29)$$

остаточная матрица  $A - A_k$  будет иметь минимально возможную согласованную

норму для всех  $A - B_k$ , где  $B_k$  - матрица ранга  $k$ , т.е.

$$\|A - B_k\|_E = \min_{\text{при } B_k = A_k} \quad (1.30)$$

Следовательно, при  $k = 1$ , остаточная матрица после удаления 1-го сингулярного вектора будет иметь минимально возможную дисперсию элементов. Это означает, что в  $\mathbf{u}_1$  или  $\mathbf{v}_1$  отношение сигнал/шум будет максимальным, т.к. ошибки некоррелированы между собой. Аналогично  $\mathbf{u}_2$  и  $\mathbf{v}_2$  максимально исчерпывают корреляции в матрице остатков  $D_2 = D - \mathbf{u}_1 \cdot \lambda_1 \cdot \mathbf{v}_1^T$  и т.д. до тех пор, пока векторы  $\mathbf{u}_{k+1, \dots, M}$  и  $\mathbf{v}_{k+1, \dots, M}$  не окажутся состоящими из элементов, ненулевые значения которых обусловлены практически только некоррелирующей составляющей в исходных данных, на практике - шумами.

На Рисунке 1.6 показаны левые сингулярные векторы, рассчитанные по матрице из 10 спектров смесей (смесь №2 из 1.3.2), показанных на Рисунке 1.4. Из Рисунка видно, что относительный вклад высокочастотных составляющих возрастает с увеличением номера вектора.

Вопрос о том, расширяется ли спектр Фурье для  $\mathbf{u}_j$  с ростом  $j$ , необходимо решать для каждого конкретного случая индивидуально, т.к. легко построить примеры, в которых наблюдается сужение частотного диапазона Фурье для больших  $j$ , путем простого перемасштабирования модельной задачи по сингулярным числам.

Нами предложен метод оценивания числа компонентов, который показал свою практическую эффективность, основанный на статистическом анализе последовательности элементов в левых сингулярных векторах матрицы данных. Как и ранее, условием применимости данного метода является достаточно частый шаг оцифровки спектров, обеспечивающий коррелированность (систематичность изменения) последовательных точек в незашумленных спектрах. В этом случае случайные составляющие, представленные гармониками Фурье исходного спектра, превышающими частоту дискретизации, должны быть некоррелированы.

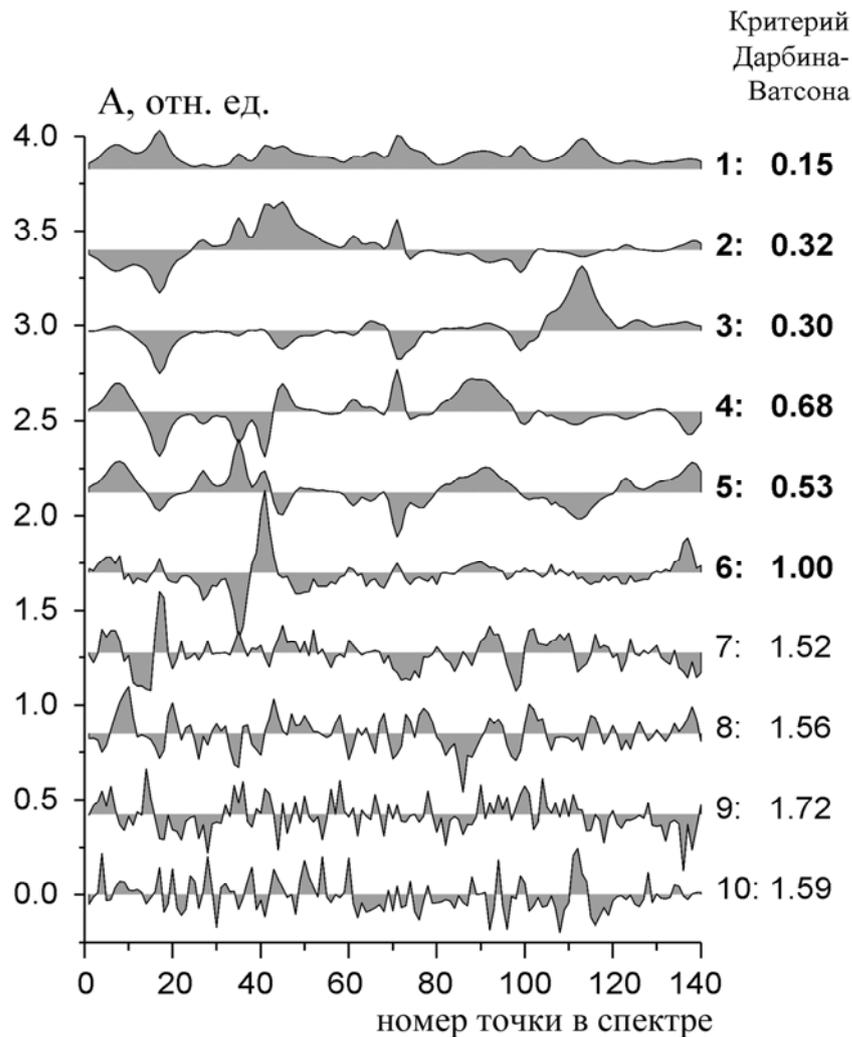


Рисунок 1.6 - Левые сингулярные векторы  $u_j$  матрицы данных (Рисунок 1.4), представленные в виде спектральных контуров. Контуров смещены по вертикали, нулевая линия соответствует горизонтальной оси в серой закрашке их площади.

Здесь надо сделать одно важное замечание. Довольно широко используемый статистический анализ на отсутствие корреляций в элементах матрицы остатков  $F^0$ , которая соответствует разности между исходной матрицей данных и матрицей, восстановленной по последним сингулярным векторам с индексами  $j=K+1, \dots, M$ :

$$F^0 = D - D^0, \quad D^0 = U_{j=K+1, \dots, M} \cdot \Lambda \cdot V_{j=K+1, \dots, M}^T, \quad (1.31)$$

и которая теоретически должна содержать только шумовую составляющую, на практике может давать заниженное число компонентов, так как малоинтенсивные компоненты маскируются шумами и "захватываются" в  $F^0$ .

Поэтому этот метод был нами далее рассматриваться не будет.

Рассмотрим применение данных статистик на примере набора 6-компонентных смесей (модель №2). Набор ИК спектров смесей органических соединений, снятых в разные моменты времени на выходе жидкостного хроматографа с помощью проточной кюветы, показан на Рисунке 1.4. На Рисунке 1.7 показан график зависимости величин логарифмов сингулярных чисел матрицы данных. Видно, что в данном случае выбор может быть сделан между 6 и 7 компонентами, что соответствует действительности: было взято 6 веществ, 7-й компонент – растворитель, спектр которого компенсировали двухлучевой схемой съемки. Рисунок 1.6 визуально более информативен: сингулярные векторы 1 - 6 заведомо описывают спектральные данные, векторы 7 - 10 имеют шумовой характер.

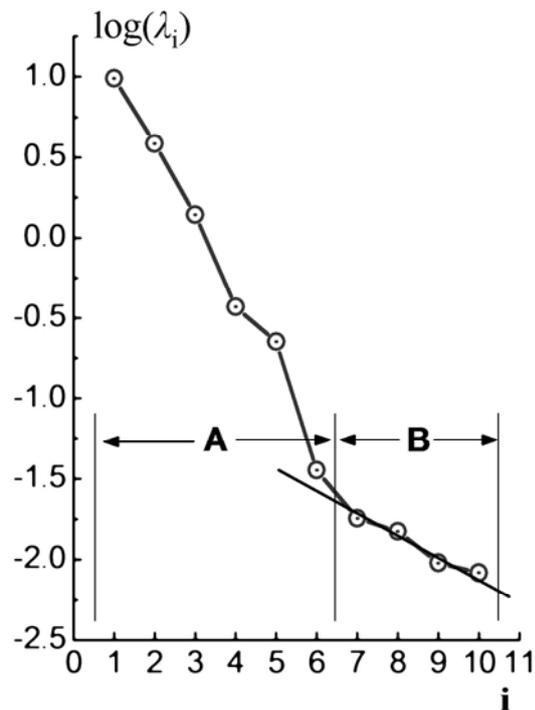


Рисунок 1.7 – График логарифмов сингулярных чисел матрицы 6-компонентных смесей (модель №2 из 1.3.2).  $R^2$  критерий для линейной аппроксимации правого участка графика равен  $0.89 > 0.7$ . Область А соответствует значимым числам, В - шумовым.

В Таблице 1.1 представлены результаты статистического анализа сингулярных векторов с помощью модифицированного критерия Вилкоксона и Дарбина-Ватсона. Результаты показывают, что при наиболее часто употребляемом уровне

значимости 0.02 за оценку числа следует принять 7 вместо 6. В дальнейшем исследование показало, что 7-й компонент представляет собой неполностью вычтенный их данных спектр растворителя – четыреххлористого углерода.

Таблица 1.1. Уровни значимости для критерия согласия Вилкоксона, рассчитанные для набора спектроскопических данных №2.

Уровень значимости Q	0.05	0.02	0.01	0.002
Нижнее критическое значение	28.00	27.00	26.00	23.00
Верхнее критическое значение	43.00	45.00	46.00	48.00

Критерий Дарбина – Ватсона  $DW$  (1.14) показал в данном случае завышенную оценку  $K = 8$  из-за наличия слабых корреляций вдоль последовательности точек в сингулярных векторах вследствие межмолекулярных взаимодействий компонентов. Это критерий оказался более чувствителен к наличию таких нарушений аддитивности. Как было указано ранее, гипотезу о наличии корреляций отвергают, если величина  $DW$  находится за пределами доверительного интервала  $\{1.7-2.2\}$  для часто используемого уровня значимости  $Q = 0.05$ . Однако, если принять в качестве доверительного интервала рекомендуемый диапазон  $\{1.5-2.4\}$  (см. раздел 1.3.1), то можно принять за оценку числа компонентов число 6.

Общим выводом для приведенного примера является оценка числа компонентов 6 или 7. Однако, число 8 также необходимо рассматривать при поиске индивидуальных спектров, что будет рассмотрено далее.

Таблица 1.2. Значения критерия согласия Вилкоксона и Дарбина-Ватсона (1.14), рассчитанные для сингулярных векторов набора данных №2.

Номер сингулярного вектора	Величина критерия Вилкоксона	Вывод о числе компонент (при уровне значимости)	Величина критерия Дарбина-Ватсона (1.14)	Вывод о числе компонент (при уровне значимости 0.05)
1	1.0	-	0.15	-
2	2.0	-	0.32	-
3	8.0	-	0.30	-
4	14.0	-	0.68	-
5	11.0	-	0.53	-
6	17.0	-	1.00	-
7	25.0	6 (0.002)	1.52	(6)
8	28.0	7 (0.02)	1.56	-
9	28.0	8 (0.02)	1.72	8
10	30.0	9 (0.05)	1.59	-

Оценка  $K = 8$  по критерию Дарбина-Ватсона была в данном эксперименте отвергнута, так как вкладом 8-го сингулярного вектора в матрицу данных можно пренебречь из-за относительной малости соответствующего сингулярного числа (Рисунок 1.7).

Помимо критерия Дарбина-Ватсона, в качестве критерия случайности элементов как матрицы остатков так и самих сингулярных векторов  $u_j$  нами предложено использовать непараметрический критерий серий Вальда-Вольфовица [78,82], позволяющий получать несмещенные и состоятельные оценки при малых объемах выборок (параметрические критерии, как правило, требуют объема не менее 1000) и заранее неизвестных законах распределения [83]. Статистика Вальда-Вольфовица проверяет случайность чередования



временами накопления). Цель мониторинга заключается в определении момента начала радиационного повреждения образца, или динамики появления компонентов в процессе протекающей в образце реакции. Совпадающие фреймы в дальнейшем объединяют в одно измерение. В процессе повреждения или другого изменения образца с его объеме появляются новые компоненты, что приводит к увеличению числа компонентов в данных измерений и изменению, как следствие, формы кривых рассеяния. Эти изменения можно контролировать не только с помощью рассмотренных в следующих разделах критериев различия формы одномерных контуров, но и с помощью оценивания ранга матрицы данных (составленной из отдельных фреймов) методом сингулярного анализа. Так как фреймы в наборе данных должны совпадать, то ранг матрицы измерений должен быть равен 1. Сингулярный анализ позволяет не только зафиксировать момент увеличения ранга, но и в ряде случаев предоставить более детальную информацию о происходящих в образце изменениях, еще до предварительной обработки данных и вычитания рассеяния буферным раствором для его компенсации. В качестве примера можно привести результаты on-line обработки последовательности измерений водного раствора белкового комплекса на синхротронной станции малоуглового рассеяния X33 (EMBL c/o DESY, Гамбург, Германия). Время измерения одного фрейма составляло 15 сек.

Из Рисунка 1.8 видно, что изменения в форме кривых рассеяния начинаются практически сразу. Форма сингулярных векторов (Рисунок 1.9) свидетельствует о наличии двух основных компонентов в наборе данных, о чем говорят и критерии проверки гипотез об отсутствии автокорреляции в последовательности элементов левых сингулярных векторов (Таблица 1.3).

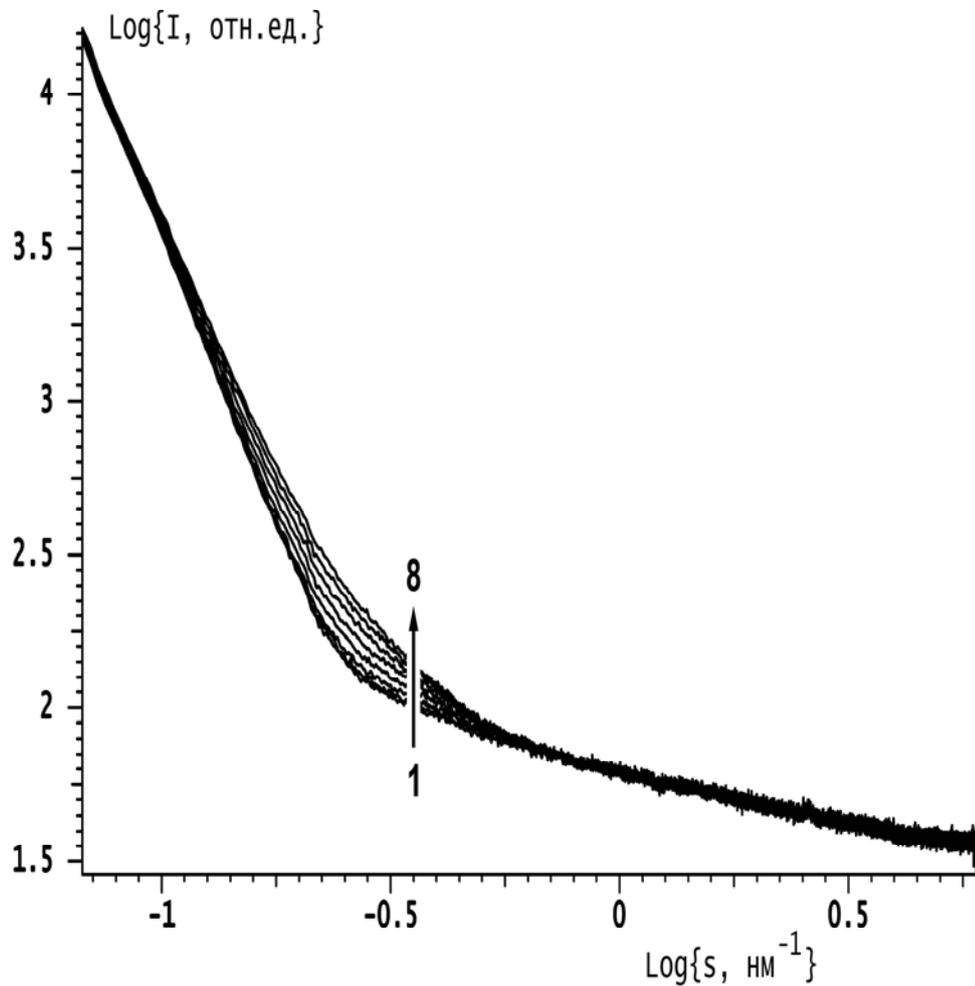


Рисунок 1.8 - Данные 8 последовательных измерений интенсивности малоуглового рассеяния от образца раствора белкового комплекса в трис-буфере на синхротронной станции X33 (EMBL c/o DESY, Гамбург, Германия). Видно, что интенсивность рассеяния в центральной области углов растет со временем - нумерация фреймов идет снизу вверх. По оси абсцисс отложены значения логарифма модуля вектора рассеяния  $|s| = \frac{4\pi \sin \theta}{\lambda}$ ,  $\lambda$  - длина волны,  $2\theta$  - угол рассеяния.

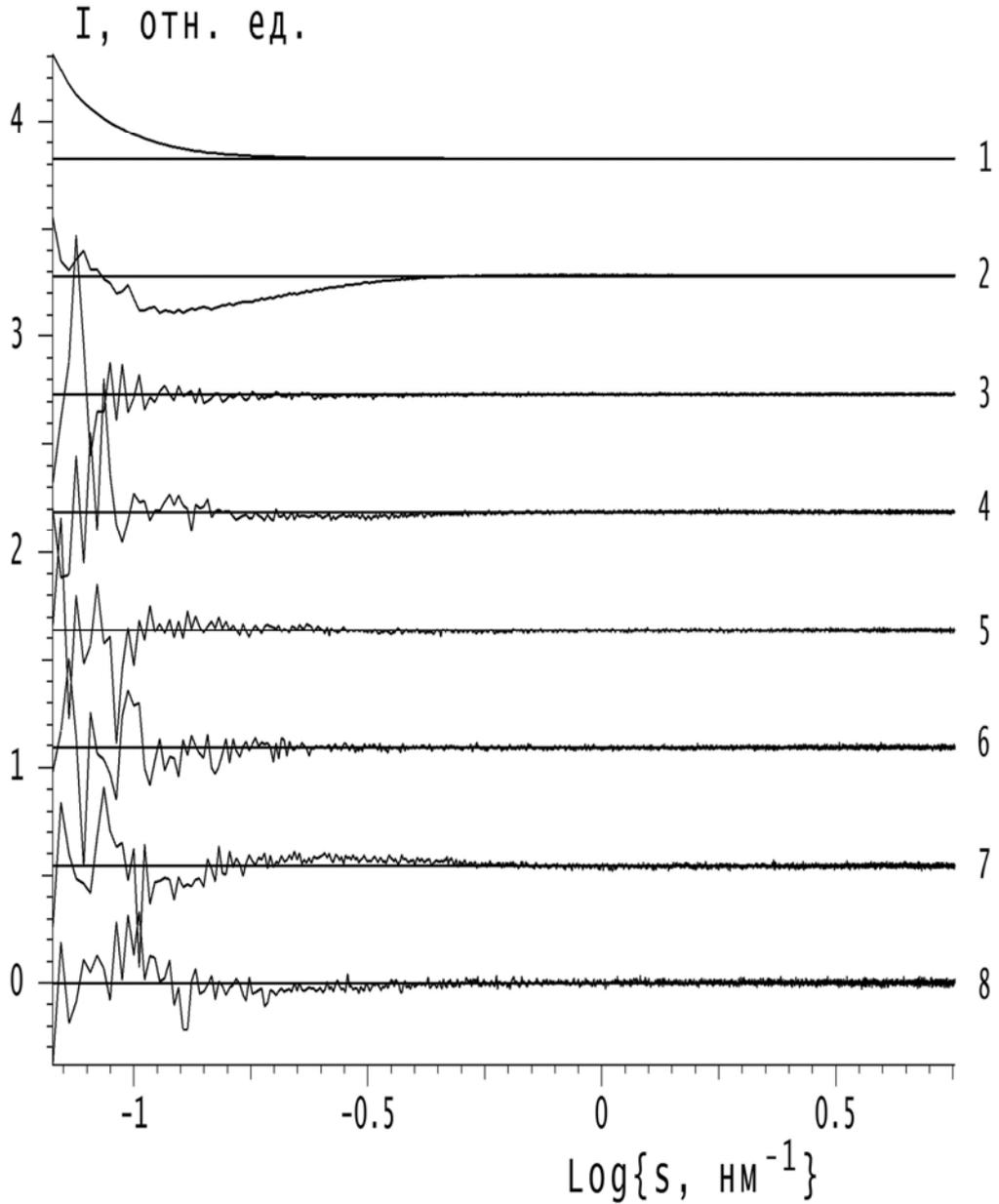


Рисунок 1.9 - Левые сингулярные векторы  $u_j$  матрицы данных (Рисунок 1.8, представленные в виде спектральных контуров. Контуров смещены по вертикали, нулевые линии показаны отдельно для каждого вектора. Единицы на осях аналогичны Рисунок 1.8.

Таблица 1.3. Значения критерия согласия Вальда-Вольфовица (1.32) и Дарбина-Ватсона (1.14), рассчитанные для сингулярных векторов набора данных малоуглового рентгеновского рассеяния (Рисунок 1.8).

Номер сингулярного вектора	Величина критерия Вальда-Вольфовица (1.32)	Нижний и верхний уровни значимости для (1.32) при $P=0.95$	Вывод о числе компонентов	Величина критерия Дарбина-Ватсона (1.14)	Вывод о числе компонентов (при уровне значимости 0.05)
1	1.0	0	-	$1.6 \cdot 10^{-2}$	-
2	0.88	$9.4 \cdot 10^{-3}$ $4.7 \cdot 10^{-2}$	> 1	$7.1 \cdot 10^{-2}$	> 1
3	-0.24	-1.98 1.98	<b>2</b>	1.4	> 2
4	-1.79	-1.98 1.98	< 4	2.1	<b>3</b>
5	-1.56	-1.98 1.98	< 5	2.4	< 5
6	0.71	-1.98 1.98	< 6	1.9	< 6
7	-2.9	-1.98 1.98	<b>7?</b>	1.7	< 7
8	-2.2	-1.98 1.98	<b>8?</b>	1.7	< 8

Здесь необходимо дать пояснения. При оценивании числа компонентов согласно приведенной таблице, необходимо сделать выбор между 2, 3, 7 и 8 компонентами. Общим правилом может служить выбор минимальной из полученных оценок, так как оценки 7 и 8 соответствуют слишком малым величинам сингулярных чисел (Рисунок 1.10) и неслучайный характер элементов 7-го и 8-го сингулярных векторов по критерию Вальда-Вольфовица чаще всего обусловлен наличием малых систематических дрейфов оборудования и условий

измерения и их вклад в экспериментальные контуры данных очень мал. Таким образом, можно сделать вывод, что число компонентов в наборе данных равно 2 и они, вероятно, соответствуют димерным и мономерным макромолекулам в исследуемом растворе.

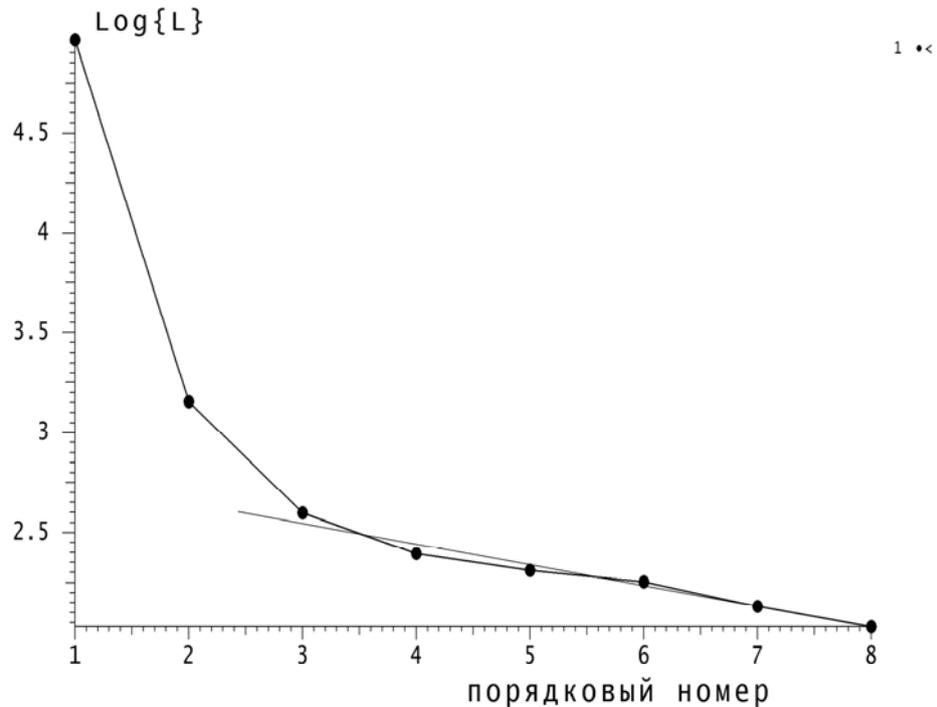


Рисунок 1.10 – График логарифмов сингулярных чисел матрицы данных Рисунка 1.8. R2 критерий для линейной аппроксимации правого участка графика равен 0.69 ~ 0.7. Выше прямой линии располагается 2 сингулярных числа, которые соответствуют числу значимых компонентов.

Еще один предлагаемый способ оценки числа компонентов и качества решения, который был успешно применен в программах анализа данных малоуглового рассеяния, рассматриваемых во 2-й части работы, основан на оконном расчете критерия автокорреляции элементов сингулярных векторов и векторов остатков (разностей между экспериментальными данными и контурами, восстановленными в результате той или иной аппроксимации). Критерий рассчитывают внутри окна, охватывающего заданный пользователем участок данных (мы рекомендуем 10% от всего диапазона, но не менее 20 точек в окне) и приписывают полученное значение центральной точке окна. Концевые участки рассчитывают по половине функции окна. В качестве сканирующей функции

предложено применять трапецеидальный контур с отношением оснований 1:2 (возможны и другие варианты функций окна, если для их выбора имеются достаточные основания). Графики зависимости величины критерия, построенные в шкале абсцисс исходных спектров наглядно демонстрируют области, в которых присутствуют систематические отклонения (значение  $DW$  (1.14) менее 1.7) и которые соответствуют случайным шумам ( $1.7 < DW < 2.2$ ). Граничные величины критериев значимости указаны здесь для типичного случая числа точек в тестируемой последовательности внутри окна 30-100 и уровне значимости гипотезы о наличии корреляций 0.05.

На Рисунке 1.11 показаны графики критерия  $DW$  для приведенного выше примера анализа данных малоуглового рассеяния. Критерий  $DW$  рассчитан для матрицы остатков (уравнение 29)  $D - D_k$ , где  $D$  - исходные экспериментальные данные,  $D_k = U_k \cdot \Lambda_k \cdot V_k^T$  - матрица, восстановленная по первым  $k$  сингулярным векторам,  $k$  представляет собой оценку числа компонентов. Ширина окна сканирования в данном примере составляла 10% (210 точек) от всего диапазона данных. Из Рисунка видно, что график статистики для 3-го сингулярного вектора близок к диапазону 1.7 - 2.2 и данный вектор можно считать состоящим из "независимых случайных" величин, что подтверждает гипотезу о наличии 2 основных, или как их называют, значимых, компонентов. Значительные отклонения в графиках с номерами 3 - 8 говорят о наличии малоинтенсивных компонентов, которые можно отнести к приборным искажениям, так как ширина пиков выбросов слишком мала для того, чтобы соответствовать гладкому угловому характеру зависимости малоуглового рассеяния от разупорядоченных систем наночастиц. Расчеты были сделаны с помощью разработанных автором программ SVDPLOT и многоцелевой программы графического анализа данных SASPLOT [A34].

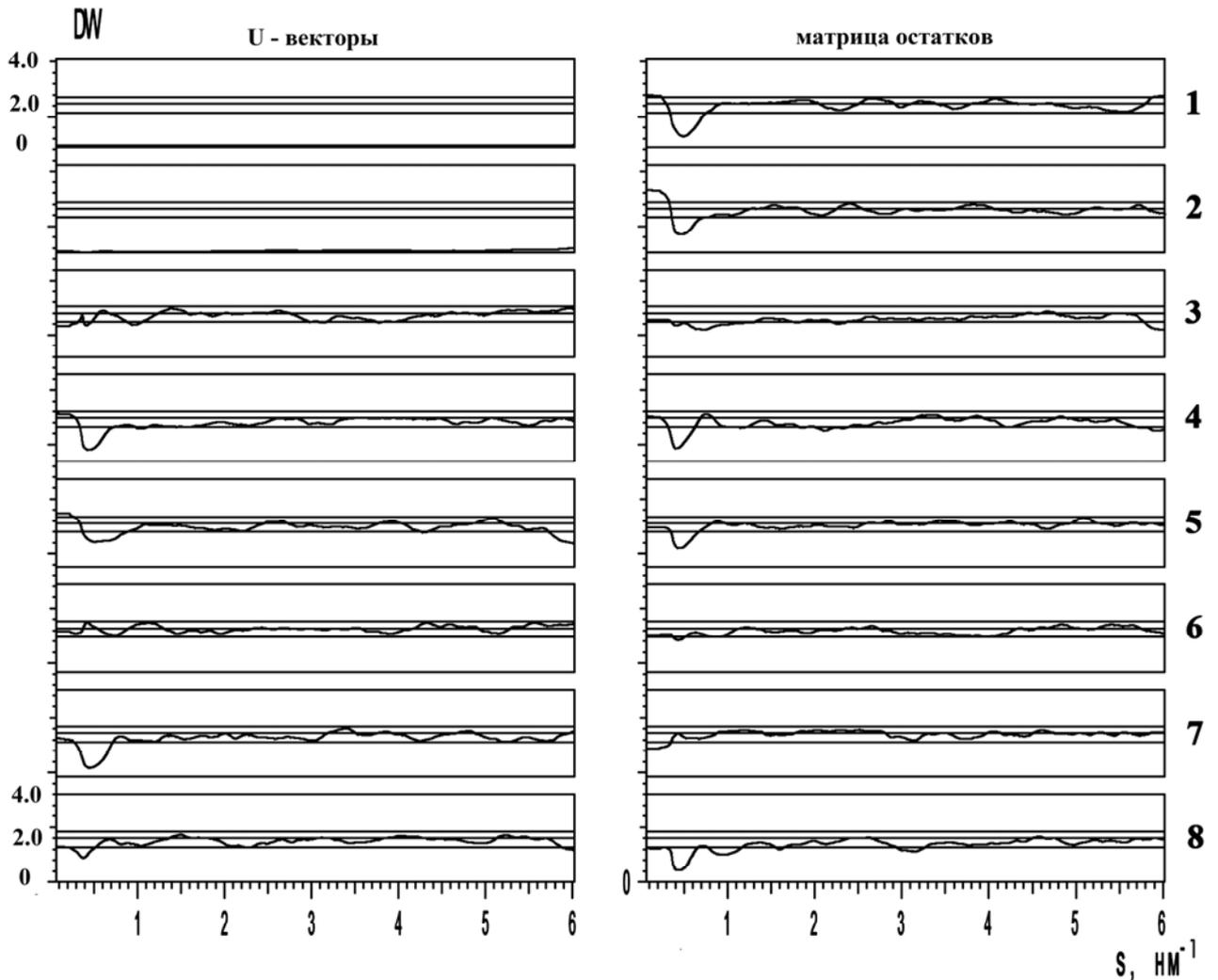


Рисунок 1.11 - Графики оконной статистики Дарбина-Ватсона для элементов левых сингулярных векторов матрицы данных Рисунка 1.8 (левая панель) и элементов матрицы остатков для гипотезы о числе компонент, равному 2 (правая панель). Номера графиков 1 - 8 соответствуют номерам сингулярных векторов или номерам экспериментальных кривых рассеяния. Каждый график критерия показан со смещением по вертикали в рамке, соответствующей значениям  $0 < DW < 4.0$ , уровни критических значений 1.7, 2.0 и 2.2 (см. текст) показаны внутри рамок тремя горизонтальными линиями. Значения критерия внутри диапазона 1.7 - 2.2 соответствуют отклонению гипотезы о наличии автокорреляций в окне сканирования. Длина окна 10% от углового диапазона, показанного на оси абсцисс.

### 1.3.6 Использование эталонных спектров и библиотечный поиск

Другой предложенный нами метод, достаточно удобный для работы эксперта, основан на преобразовании проектирования эталонного спектра компонента, заведомо присутствующего в смеси, на подпространства, образуемые возрастающим числом левых сингулярных векторов. Эта операция известна в литературе по факторному анализу под названием Target Transformation [1]. В нашей работе [A2] мы ранее предложили вычислять проекцию по уравнению (1.23). Впоследствии этот метод был дополнен нами статистическим анализом вектора разности между исходным вектором и его проекцией с использованием критерия Дарбина-Ватсона. Пусть  $t$  - вектор-спектр чистого вещества из библиотеки,  $p$  - его проекция на пространство столбцов матрицы данных  $D$ . Использование наборов спектров родственных смесей вместо одного позволяет упростить численные процедуры проверки присутствия компонента. Методы контроля в этом случае основываются на уже упоминавшемся факте, что при отсутствии экспериментальных ошибок вектор-спектр истинного компонента должен лежать в гиперплоскости  $AOD$ , образуемой спектрами смесей (см. Рисунок 1.1). Следовательно, для проверки решений необходимо организовывать процесс численной оценки степени близости эталонного спектра к подпространству смесей. При совпадении числа базисных векторов с истинным числом компонентов спектр проекции совпадает с эталонным с точностью до случайных ошибок и увеличение размерности подпространства не приводит к существенному увеличению критериев сходства, в качестве которого в простейшем случае можно использовать коэффициент корреляции, обратную величину угла между векторами, косинус угла и т.п. В нашей работе [A2] в качестве критериев присутствия пробных спектров мы ранее использовали величину (косинуса) угла между  $t$  и  $p$ ,

$$\cos(\mathbf{t}, \mathbf{p}) = \frac{\sum_{i=1}^N t_i \cdot p_i}{\left[ \sum_{i=1}^N t_i^2 \cdot \sum_{i=1}^N p_i^2 \right]^{1/2}} \quad (1.33)$$

а также коэффициента корреляции, вычисляемого по формуле

$$r(\mathbf{t}, \mathbf{p}) = \frac{N \sum_{i=1}^N t_i \cdot p_i - \sum_{i=1}^N t_i \cdot \sum_{i=1}^N p_i}{\left\{ \left[ N \sum_{i=1}^N t_i^2 - \left( \sum_{i=1}^N t_i \right)^2 \right] \cdot \left[ N \sum_{i=1}^N p_i^2 - \left( \sum_{i=1}^N p_i \right)^2 \right] \right\}^{1/2}} \quad (1.34)$$

Наиболее сложным оказывается выбор пороговой величины критерия присутствия компонента в смеси. Практика решения модельных задач показала, что для ИК-спектров органических соединений за пороговые значения критериев можно принимать величину угла = 8 – 10°,  $r = 0,92 - 0,95$ . Визуальное сравнение спектров  $\mathbf{t}$  и  $\mathbf{p}$  мы не рекомендуем использовать, т.к. внешне похожие контуры могут давать совершенно неудовлетворительные значения критериев. Аналогично предыдущему методу, критерием совпадения проекции с эталоном могут быть непараметрические критерии случайности выборки, примененные к вектору разности между эталоном и проекцией. В Таблице 1.4 приведены результаты определения числа компонентов для набора спектров №2 (Рисунок 1.4) при проектировании известных спектров для каждого из компонентов. В данном случае применен критерий величины угла между вектором эталонного спектра  $\mathbf{t}$  (Рисунок 1.1) и его проекцией на пространство столбцов матрицы  $D^\#$ , восстановленную по уравнению (1.2) при разном числе сингулярных векторов, начиная с 3-х (соответствующие сингулярные числа упорядочены в убывающем порядке). Скачкообразное уменьшение величины угла соответствует числу векторов, равному числу компонентов.

При использовании этого метода в наборы смесей можно искусственно добавлять известный компонент и использовать его в качестве эталонного.

Таблица 1.4. Проектирование известных спектров компонентов на подпространство столбцов матрицы данных 6-компонентной смеси №2. Приведены величины углов (в градусах) между вектором известного компонента и его проекцией на матрицу  $D^\#$ , восстановленную с использованием гипотез  $K = 1, 2, \dots, 10$ . Скачок величины угла (уменьшение) соответствует числу компонентов, уменьшенному на 1. Оценка, равная 6, совпадает с истинным числом основных компонентов.

Гипотеза о числе компонентов	10	8	6	5	4	3
Компонент						
изооктан	1.52	1.53	1.95	5.7	14.5	20.0
бензол	6.1	6.5	8.5	28.7	41.1	44.8
толуол	8.1	9.0	10.0	41.0	45.3	53.8
диэтиловый эфир	1.93	2.54	3.0	8.83	20.0	21.8
ацетон	2.83	3.62	5.1	10.4	31.3	48.0
амилацетат	0.81	0.82	1.60	2.70	9.73	17.3

Помимо упомянутых критериев сходства в [84] выделено три группы сходных процедур сравнения, дающих на практике удовлетворительные результаты. Первая заключается в явном вычислении расстояния между векторами  $t$  и  $p$ , например, методом ортогонализации Грама-Шмидта. Это расстояние представляет, по сути, величину невязки для несовместной системы линейных уравнений  $Dx = p \approx t$ . Второй метод, Target Transformation Analysis (TFA), на самом деле полностью аналогичен первому за исключением того, что матрица  $D$  перед проектированием аппроксимируется матрицей  $D^\#$  меньшего ранга, равного предполагаемому числу компонентов. Третий метод (Bessel's

inequality test) заключается в вычислении длины проекции  $\mathbf{p}$  при условии, что  $\mathbf{t}$  нормирован к 1. Коэффициент сходства,  $b$ , вычисляют в виде суммы скалярных произведений  $\mathbf{t}$  с  $K$  первыми сингулярными векторами  $D$ :

$$b = \sum_{k=1}^K (\mathbf{t}^T \cdot \mathbf{u}_k)^2. \quad (1.35)$$

Он принимает значения от 0 до 1 и равен, по сути, квадрату длины проекции  $\mathbf{p}$ . Близкое к единице значение  $b$  является свидетельством хорошего соответствия данных.

В нашей работе мы предложили для проверки приемлемости эталонного спектра использовать аппарат теории квадратичных форм, применяемый к соотношению

$$Q = (\mathbf{s}^T \cdot \mathbf{s}), \quad \mathbf{s} = \mathbf{t} - U^\# \cdot U^{\#T} \cdot \mathbf{t}, \quad (1.36)$$

где  $U^\#$  - левая сингулярная матрица, в которой оставлено  $K$  первых столбцов.

Утверждается [73], что  $Q$  распределена как  $\chi^2$  с  $N - K$  степенями свободы.

Несмещенная оценка дисперсии  $Q$  равна

$$\sigma_Q^2 = \frac{Q}{N - K}. \quad (1.37)$$

Здесь нужно заметить, что  $\mathbf{s}$  представляет собой вектор расстояния между эталонным спектром  $\mathbf{t}$  и его проекцией на  $D^\#$ . Величину ошибки в матрице данных вычисляют аналогично, используя в качестве пробных спектров сами спектры смесей  $\mathbf{d}$  и проектируя их на  $D^\#$ :

$$Q_D = (\mathbf{d} - U^\# \cdot U^{\#T} \cdot \mathbf{d})^T \cdot (\mathbf{d} - U^\# \cdot U^{\#T} \cdot \mathbf{d}), \quad (1.38)$$

$Q_D$  имеет  $\chi^2$  распределение с  $N - K$  степенями свободы. Поэтому оценка дисперсии равна

$$\sigma_D^2 = \frac{Q_D}{N - K}. \quad (1.39)$$

Для нормированной по столбцам матрицы  $D$  соответствующая оценка равна

$$\sigma_D^2 = \frac{1}{(M-1)(M-K)} \sum_{j=1}^M \mathbf{d}_j^T \cdot \frac{U \cdot U^T - U^\# \cdot U^{\#T}}{\mathbf{d}_j^T \cdot \mathbf{d}_j} \cdot \mathbf{d}_j \quad (1.40)$$

Используя величину  $p^2 = \frac{Q}{\mathbf{t}^T \cdot \mathbf{t}}$ , проверяют Фишеровское отношение дисперсий

$$F = \frac{\sigma_Q^2}{\sigma_D^2} \quad (1.41)$$

с  $N-K-(M-1)(M-K)$  степенями свободы. Задаваясь определенным уровнем значимости  $p$ , по величине  $F$  принимают или отвергают гипотезу о приемлемости пробного спектра в качестве компонента и, следовательно, гипотезу о правильности оценки числа компонентов  $K$ . В работе [73] отмечается, что для рассмотренного метода имеются два ограничения: пробный вектор  $\mathbf{t}$  должен отстоять от подпространства  $D^\#$  далее, чем любой из векторов-спектров смесей  $\mathbf{d}_j$  (иначе будут получаться отрицательные величины дисперсии) и должно соблюдаться неравенство

$$(N - K) > M(M - K),$$

чтобы  $\sigma_Q^2$  был определен. Легко видеть, что данный метод неприменим и в случае, когда  $N-K$  близко к  $(M-1)(M-K)$ :  $\sigma_Q^2$  тогда может принимать необоснованно большие значения.

Используемый нами в работе критерий автокорреляций Дарбина-Ватсона (1.14), был применен к элементам вектора  $\mathbf{s}$  (1.36). Однако, из-за возможного наличия систематических ошибок в  $D$  его непосредственное использование для принятия решения о числе компонентов также невозможно, но решение можно принимать по наличию скачка на графике этого критерия относительно числа используемых в (1.36) сингулярных векторов.

Рассмотренные оценки справедливости гипотезы о присутствии индивидуального спектра в наборе спектров смесей могут использоваться для анализа данных методом библиотечного поиска. Надежность решения в данном случае зависит от полноты базы спектроскопических данных.

### 1.3.7 Определение состава единственного спектра смеси с помощью набора эталонных спектров

Если исследователь располагает информацией о возможном составе смеси и соответствующими спектрами компонентов, то проверить их наличие можно простым способом, решая систему линейных уравнений относительно неизвестных вкладов компонентов  $\mathbf{C}$  при восстановлении единственного доступного спектра смеси. При этом можно рассмотреть два подхода к получению ответа на вопрос о ее составе.

Первый подход заключается в анализе качества восстановления исходного контура путем решения методом наименьших квадратов (если число отсчетов  $N$  превышает число компонентов  $K$ ) системы линейных уравнений

$$\mathbf{E} \cdot \mathbf{c} \approx \mathbf{J}, \quad (1.42)$$

где  $\mathbf{E}$  -  $N \times K$  матрица системы, столбцы которой представляют собой априори заданные спектры пробных компонентов,  $\mathbf{J}$  - спектр исследуемой смеси,  $\mathbf{c}$  - вектор искомых спектральных вкладов (концентраций). Критериями качества решения могут служить, помимо суммарной квадратичной невязки или величины хи-квадрат, критерии случайного характера остатков, рассмотренные в разделе 1.3.1. Суммарный критерий качества, разумеется, должен включать критерий неотрицательности концентраций. Отрицательное относительное значение концентрации означает неверность гипотезы о составе, то есть отсутствие соответствующего компонента в смеси и / или неверность набора компонентов в целом.

Второй подход основан на решении системы (1.42) методом неотрицательных наименьших квадратов, то есть поиске минимума суммарной квадратичной невязки  $\|\mathbf{E} \cdot \mathbf{c} - \mathbf{J}\|^2$  при условии  $\forall c_i \geq 0, \quad i = 1, \dots, K$ . В этом случае основную роль при оценке приемлемости решения играют критерии невязки и статистики остатков  $\mathbf{E} \cdot \mathbf{c} - \mathbf{J}$ .

Оба подхода были реализованы при участии автора в программе анализа данных малоуглового рентгеновского рассеяния от многокомпонентных

разбавленных растворов наночастиц OLIGOMER, которая входит в программный пакет PRIMUS [A34]. Естественно, данная программа носит общий характер и ее применение не ограничивается анализом данных интенсивности малоуглового рассеяния.

На практике, определение состава смеси по данным малоуглового рассеяния носит характер вспомогательной задачи, которая решается в процессе предварительного анализа данных и выделения данных рассеяния от одного из компонентов, например, как это сделано в [85].

## 1.4 Методы разделения спектров смесей

Расчет спектров индивидуальных компонентов и их концентраций только по экспериментально полученным смешанным спектрам представляет собой одно из важных направлений хемометрии. Эта группа методов, рассмотренная в разделе 1.1, объединена названием "Self Modeling", или "автомоделирование". Во многих случаях этот подход оказывается единственно доступным способом определения контуров сигналов компонентов в смесях неизвестного состава.

Цель автоматического разделения смешанных контуров сигналов заключается в выделении контуров компонентов без использования физико-химической модели или априорной информации о системе. К преимуществу автоматического разделения относится его применимость ко многим типам спектроскопических данных, быстрота получения ответа и возможность относительно простого учета любой дополнительной информации.

Число работ, посвященных применению алгоритмов разложения для качественного и количественного анализа смесей сложного состава, не так велико, как можно было бы ожидать. В какой-то мере это обусловлено практической сложностью статистического обоснования приемов факторного анализа и формализации физически обоснованных требований к решению – фильтров неоднозначности.

В данной работе предложен сравнительно простой для использования метод разделения, использующий в качестве фильтров решений такие очевидные требования как неотрицательность компонентов и соответствующих им концентраций в совокупности с требованиями простоты формы и максимального различия индивидуальных контуров. Для данного набора условий можно сформулировать условия существования единственного решения.

### 1.4.1 О единственности решения задачи разложения

Основным вопросом в анализе смешанных сигналов является то, какие свойства формы контуров компонентов могут обеспечивать их нахождение единственным образом. Существует, по крайней мере, одно такое свойство: наличие в спектрах компонентов областей взаимной неперекрываемости, расположенных таким образом, что в любой паре спектров смесей должны найтись, по крайней мере, два участка, ненулевая интенсивность которых обусловлена только одним компонентом, своим для каждого участка (условия Аленцева-Фока [22 - 24]). Это условие на практике оказывается очень жестким и в дальнейшем мы покажем, что его можно значительно ослабить.

Для метода, предложенного в [22], важно наличие именно участков неперекрываемости, а не отдельных точек, т.к. он основан на поиске таких областей путем анализа кривых отношений двух контуров спектров смесей. Аналогичный метод расчета был позднее предложен в [64], а также применен к анализу "многомерных" данных (спектров испускания-возбуждения в люминесцентном анализе) [86]. Основная трудность при реализации схемы разделения Аленцева-Фока заключается в нахождении неперекрываемых областей в присутствии спектроскопических шумов. Таким областям должны соответствовать горизонтальные участки в кривых отношения двух исходных смешанных контуров. Даже сравнительно небольшие спектральные шумы и другие нарушения аддитивности спектров смесей приводят к существенным искажениям горизонтальности и максимальное число компонентов ограничивается на практике 2-3. Сам способ получения спектров компонентов в работе [22] может рассматриваться как доказательство **достаточности** условий неперекрываемости Аленцева-Фока для единственности решения. Легко также получить доказательство этого аналитически, считая известными относительные концентрации компонентов в смесях. Эти концентрации при соблюдении условий Аленцева-Фока можно однозначно определить из анализа спектров смесей, например, при помощи оригинальной процедуры [22], основанной на вычислении

коэффициентов отношения спектров смесей по областям, в которых индивидуальные спектры неперекрываются. Если же известны относительные спектральные вклады (матрица  $C$ ), то задача определения спектров сводится к решению невырожденной (при линейной независимости кривых концентрационного распределения компонентов по смесям) системы линейных уравнений (02)  $D = X \cdot C$ .

Можно показать, что условия Аленцева-Фока являются и **необходимыми**. В самом деле, если предположить, что некоторое решение единственно, хотя спектры компонентов перекрываются полностью и от них не требуют наличия взаимно неперекрываются участков, то из этого решения путем каких-либо линейных комбинаций индивидуальных спектров (а они теперь допустимы, т.к. не выводят результирующие спектры из неотрицательного подпространства возможных решений  $AOD$ , см. Рисунок 1.1) можно получить новое решение в виде разложения (1.1) с другими матрицами  $X$  и  $C$ . Полученное противоречие с исходным предположением о единственности и доказывает необходимость условий Аленцева-Фока для однозначности разложения (1.1). Следовательно, решение задачи о нахождении спектров компонентов по спектрам родственных смесей при отсутствии другой априорной информации имеет единственное решение (совпадающее с истинными индивидуальными спектрами) тогда и только тогда, когда выполнены условия неперекрываемости Аленцева-Фока.

Доказав этот факт, мы разработали численные методы, использующие информацию о наличии (но уже не обязательно расположении) нулевых участков или отдельных точек в спектрах компонентов (получаемую, например, из анализа горизонтальных участков в кривых отношений двух спектров смесей, как в [22], или каким-либо другим образом) [A5]. Однако на практике учет неперекрываются точек лучше вводить в виде штрафного члена в целевую функцию задачи, как это рассмотрено в разделе 1.4.3.

Некоторые авторы [30,31] в виде необходимых и достаточных условий единственности разложения (2) рассматривают "условия разрешимости" вида

$$K \leq \frac{N \cdot M}{N + M} \quad (1.43), \text{ или}$$

$$K \geq \frac{N \cdot M}{N - M} \quad (1.44)$$

которые ограничивают минимальное число условных уравнений в (2) при неизвестных  $X$  и  $C$ . Однако, соотношение (2) при этом вовсе не является системой линейных алгебраических уравнений, так как содержит произведение неизвестных  $X$  и  $C$ . Поэтому подсчитывать число условных уравнений для выведения "условий разрешимости", как это сделано в [87], некорректно, даже рассматривая их в совокупности с разнообразными требованиями нормировки спектров и соответствующих относительных концентраций. Впрочем, условиями единственности в [33] считают совокупность условий неперекрываемости спектров компонентов, так и кривых их концентрационного распределения по смесям, что есть, по сути, есть "двойные" условия Аленцева-Фока, которые можно использовать и независимо, но на практике они оказываются слишком жесткими и редко выполняемыми.

Метод Аленцева-Фока неприменим, когда в спектрах компонентов участки неперекрываемости малы или даже представляют собой единичные точки (иногда это можно преодолеть более подробной оцифровкой спектров). Многие известные методы разложения спектров сложных смесей позволяют учесть эти свойства индивидуальных спектров, не проводя специального поиска точек неперекрываемости, как это реализовано в методах, рассматриваемых в данной работе, а также в относительно недавно разработанных пакетах программ **RADICAL**, **SIMPLISMA**, **MCR-ALS**, **MILCA**, **SNICA** (раздел 1.1). Однако, для обеспечения надежности решения, доказательство выполнимости условий Аленцева-Фока необходимо получать каким-либо другим способом, из физических соображений, или привлекая дополнительную информацию о задаче.

### 1.4.2 Метод итерационного проектирования с отбором признаков

Располагая набором спектров смесей, спектр одного компонента можно найти с помощью предлагаемой нами процедуры, задавая одну или несколько ненулевых точек в пробном спектре  $t_1$  (Рисунок 1.1), отвечающих положению наиболее интенсивного пика в спектрах смесей и проектируя  $t_1$  на подпространство столбцов матрицы с помощью преобразования

$$p_1 = U \cdot U^T \cdot t_1, \quad (1.45)$$

где  $p_1$  – вектор проекции. Матрица  $U$  здесь – матрица левых сингулярных векторов из разложения (1.2). Операция проектирования обеспечивает одно из условий допустимости решений - спектров проекций: они представляют собой линейные комбинации спектров смесей. Это условие и наличие спектроструктурных корреляций не позволяют сконструировать спектры компонентов из имеющихся в спектрах смесей полос произвольным образом.

Как уже сказано, на первом этапе за начальное приближение  $t_1$  берут искусственно приготовленный спектр, состоящий из одной или нескольких интенсивных полос, заведомо содержащихся в искомом индивидуальном контуре. Их относительные интенсивности, форму и положение можно не стремиться указывать совершенно точно, желательно лишь не включать в их число "чужие" интенсивные полосы, т.к. это иногда может несколько осложнить дальнейшую процедуру. В нашей практике обработки ИК-спектров органических соединений в качестве стартового мы брали спектр, состоящий всего из одной полосы. Опыт показывает, что лучше начинать с максимальной по интенсивности полосы в спектрах смесей, хотя всегда возможны другие варианты.

Полученную оценку  $p$  уточняют в итерационном процессе, который заключается в последовательном конструировании очередного пробного вектора  $t_2$  из  $p_1$ , полученного на предыдущем шаге, путем зануления всех отрицательных участков и тех полос, которые резко уменьшились после проектирования (из-за чего можно предположить, что они оказались "чужими"). Во-вторых, в спектре проекции можно усилить вновь появившиеся полосы, т.к. они могут оказаться

присущими компоненту, но не учтенными на предыдущем шаге. Нами рекомендуется занулять все участки, кроме второго по интенсивности пика. После этого полученный контур опять проектируют на и отмечают степень изменений. Последнее можно делать визуально, или с помощью какого-нибудь численного критерия сравнения двух спектров (коэффициента корреляции (1.34), косинуса угла (1.33) и т.п.).

Процедуру проектирования и модификации спектра проекции повторяют до тех пор, пока векторы  $t_n$  и  $p_n$  перестают меняться в основных деталях (в случае отсутствия шумов они должны совпасть с истинными спектрами компонентов с точностью до масштаба) [A2, 88]. Когда фиксация группы последовательно найденных полос не вызывает появления новых при проектировании, необходимо поочередно исключать как основные, так и малоинтенсивные полосы из полученного контура с помощью аналогичной последовательности операций. Если эта последовательность покажет, что чужих полос в найденном спектре нет (это видно из того, что исключаемые полосы восстанавливаются после проектирования), то финальный спектр проекции принимают за спектр чистого компонента.

Когда все полосы в спектрах смесей исчерпаны и найдены спектры компонентов, проводят расчет соответствующей концентрационной матрицы  $C$  по уравнению, аналогичному (1.1). Отрицательность элементов этой матрицы свидетельствует о неправильной форме индивидуальных спектров и их следует исправить путем зануления отрицательных значений в  $C$  и расчета соответствующей матрицы компонентов  $X$  с последующей повторной коррекцией результирующих спектров с помощью проектирования. На этом этапе можно привлекать дополнительные сведения о концентрационном распределении компонентов по смесям и вводить соответствующие поправки в  $X$ . В любом случае, эти сведения можно проверить по виду соответствующих спектров компонентов.

Являясь проекциями на пространство столбцов  $D$ , полученные индивидуальные спектры точно (в пределах ошибок исходных данных, если при

проектировании брали число базисных векторов в  $U$  меньше числа смесей) воспроизводят весь набор спектров смесей и, следовательно, удовлетворяют разложению (1.1). С другой стороны, следует отдавать себе отчет в том, что найденное решение лишь приближенно совпадает с истинными спектрами компонентов, если при его получении мы не требовали выполнения условий неперекрываемости Аленцева-Фока, т.е. не требовали нулевой интенсивности в заранее известных участках или точках спектров. На степень совпадения оказывают большое влияние такие факторы, как уровень ошибок в  $D$ , степень близости спектров компонентов (чем они ближе друг к другу, тем хуже обусловлена  $D$  и больше вероятность искажения решений), степень разделения компонентов в смесях (чем она хуже, тем больше число обусловленности  $C$  – отношение максимального к минимальному сингулярных чисел матрицы, и тем сильнее искажения в  $X$ ). Все эти моменты, как правило, выясняются только в ходе самого анализа.

Существенное ограничение этого метода состоит в необходимости априорного знания расположения малоинтенсивных участков в спектрах компонентов. Если эту информацию извлекать при помощи процедуры Аленцева-Фока [22], то нет необходимости использовать метод итераций, т.к. спектры компонентов будут вычисляться в процессе поиска неперекрываемых зон. Однако, как показала практика, при числе компонентов более 2-х и наличии спектроскопических шумов, процедура Аленцева-Фока, являясь разностным методом, не может обеспечить нахождение некоторых из индивидуальных спектров. Метод итерационного проектирования оказывается в этом случае значительно устойчивее и поэтому предпочтительнее.

При поиске остальных компонентов можно использовать требование максимального отличия получаемых контуров. Критерии различия будут рассмотрены далее.

На Рисунке 1.12 показаны результаты итерационного проектирования для трех компонентов 5-компонентного набора ИК-спектров смесей №2 (Рисунок 1.4).

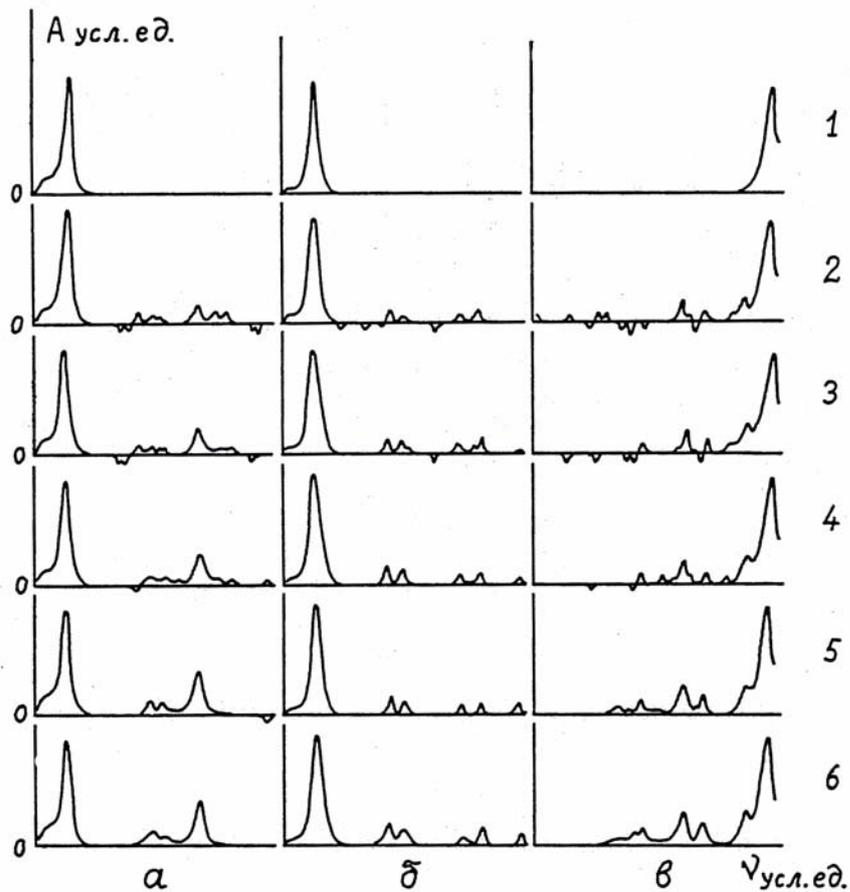


Рисунок 1.12 - Шаги восстановления с помощью преобразования проектирования спектров компонентов из спектров 5-компонентных смесей. Спектр ацетона (а), циклогексанона (б) и диэтилового эфира (в). 1 - исходный спектр  $t_1$ , 2 - 5 - спектры проекций  $p_{1-4}$  для последовательных шагов проектирования, 6 - спектры компонентов, измеренные для чистых веществ.

При рассмотрении спектров смесей было видно, что первая слева интенсивная полоса поглощения (при  $1717\text{ см}^{-1}$ ) изменяла положение максимума при варьировании их состава от  $1718$  до  $1716\text{ см}^{-1}$  и при этом менялась форма ее крыльев. Этот факт позволил предположить, что данная полоса есть суперпозиция двух полос, принадлежащих разным компонентам. И действительно, из Рисунка 12(а,б) видно, что, если в качестве исходного спектра задавать контур, состоящий только из полосы  $1718\text{ см}^{-1}$  или только  $1716\text{ см}^{-1}$ , можно найти два спектра, различающиеся в других областях - спектры ацетона и циклогексанона. Следует отметить, что выделение спектра циклогексанона не представляло особых трудностей, а для ацетона сходимость метода оказалась значительно хуже, пока

не была достигнута правильная форма полосы при  $1718 \text{ см}^{-1}$ . Сложность нахождения этих спектров заключалась именно в том, что наиболее характерные спектральные признаки у них существенно перекрыты, а остальные участки малоинтенсивны и перекрыты с другими компонентами. Выделение остальных спектров не представило трудностей, т.к. они были достаточно характеристичны.

Эксплуатация программы, реализующей описанный способ, показала, что наиболее эффективен именно диалоговый режим работы, при котором исправления в спектры вносит оператор, наблюдая их на экране дисплея. Запрограммировать автоматический режим работы для всех случаев достаточно трудно, так как при этом необходимо создавать эффективные и универсальные алгоритмы распознавания и анализа спектральных образов, формализующие процессы исправления. С другой стороны, диалоговый режим позволяет эксперту легко и наглядно использовать всю доступную информацию о решаемой задаче.

### **1.4.3 Метод итерационного моделирования индивидуальных спектров с использованием дополнительной информации и ограничений**

Существует группа задач, в которых даже точная исходная информация не обеспечивает совпадения решений с истинными индивидуальными спектрами. Главная особенность рассматриваемого случая состоит в том, что имеется множество допустимых вариантов решения, которые формально удовлетворяют основным условиям. Выделение из этого множества решений одного требует привлечения самой разнообразной дополнительной, часто трудно формализуемой в рамках используемого метода разложения, информации. Это, в свою очередь, может приводить к субъективности оценок индивидуальных спектров. Необходимо подчеркнуть, что субъективность состоит не в полной произвольности выбора решения (оно всегда должно удовлетворять условиям аддитивности и неотрицательности спектров и концентраций), а в субъективности оценки важности дополнительной информации, которая (оценка) в дальнейшем примет вид весового коэффициента. В этом разделе мы рассмотрим разработанный нами метод поиска, относящийся к группе Self Modeling, основанный на сингулярном анализе матрицы данных и позволяющий учитывать разнообразную дополнительную информацию о задаче и требования к решению. Метод поиска является универсальным и мы применили его к решению задач спектроскопического анализа, а методы коррекции данных к малоугловому нейтронному рассеянию [A1–A5,A8,A10–A13,A17,A26].

Под дополнительной информацией мы будем понимать наличие известных спектров компонентов, под требованиями – условия неотрицательности спектров и концентраций компонентов (сигналов), максимальное различие контуров индивидуальных сигналов и их простота.

Большинство методов решения задач спектрального разложения используют различные численные процедуры, как правило, адаптированные к конкретному типу используемой дополнительной информации. При всех различиях подходов их объединяет то, что поиск векторов-спектров компонентов

ведут в областях допустимых решений линейной задачи (АОВ и COD на Рисунке 1.1). Так или иначе, поиск спектров можно свести к построению линейных комбинаций векторов, образующих некоторый базис подпространства спектров смесей. Таким базисом можно считать как саму матрицу векторов-спектров смесей, так и ортонормированный набор левых сингулярных векторов матрицы  $U$ .

Предлагаемый нами подход основан на построении последовательности пробных спектров компонентов

$$x_{ik} = \sum_{j=1}^L u_{ij} \cdot y_{jk}, \quad K \leq L \leq M, \quad (1.46)$$

с коэффициентами  $y_{jk}$ , которые есть искомые параметры и аргументы некоторой целевой функции  $W(\mathbf{y})$ , точка минимума которой должна удовлетворять используемым ограничениям и условиям, отражающими разнообразную дополнительную информацию. Здесь  $K$  - число компонентов,  $M$  - число смесей,  $L$  - число сингулярных векторов в сумме. Вектор аргументов  $\mathbf{y}$  здесь означает искомые параметры, по которым будут восстановлены (смоделированы) спектры компонентов.

Особенность предложенного метода заключается в том, что при наличии неперекрытых точек в спектрах компонентов нет необходимости включать в целевую функцию невязку  $RES$  между модельными спектрами смесей, рассчитываемыми по соотношению

$$D^{\#} = X_n C_n, \quad (1.47)$$

где  $n$  – номер шага, или итерации поиска, и экспериментальными данными:

$$RES = \frac{1}{N \cdot M - K} \|D_{эксн} - D_n\|^2 \quad (1.48)$$

Эта невязка определяется матрицей остатков  $F^0$  из (1.7):  $D = D^{\#} + F^0$ , которая, в свою очередь равна произведению сингулярных матриц:  $F^0 = U_{M-L} \Lambda_{M-L} V_{M-L}^T$ . Индекс  $M - L$  обозначает, что в матрицах оставлены последние  $M - L$  столбцов, а  $L$  равно числу первых сингулярных векторов, взятых при построении модельной матрицы компонентов (1.46) (сингулярные векторы всегда упорядочены в

соответствии с убыванием сингулярных чисел в зависимости от их порядкового номера). Таким образом, какие бы пробные спектры компонентов ни вычислялись в процессе поиска по (1.46), они будут лежать строго в подпространстве  $D^\#$ , и формально удовлетворять условиям решения. Однако, накладываемые на вид контуров компонентов ограничения увеличивают минимально возможную невязку, которую можно оценить по матрице остатков, вычисляемую с использованием сингулярных векторов с индексами  $L+1, \dots, M$ , не вошедших в число векторов, из которых конструируются спектры компонентов:

$$RES_{\lim} = \frac{1}{N \cdot M - L - 1} \|R\|^2, \quad R_{ik} = \sum_{j=L+1}^M u_{ij} \cdot y_{jk}. \quad (1.49)$$

Включение в состав целевой функции невязки (1.48), таким образом, не является необходимым в случае выполнения условий неперекрываемости, но, как показала практика, ускоряет поиск минимума и достижение глобального экстремума в задачах большой размерности, при числе компонентов более 4-5. Большое число штрафных членов в составе целевой функции часто делает ее мультимодальной, то есть может существовать множество локальных минимумов, соответствующих разным наборам спектров компонентов, отвечающих совокупности наложенных на них условий. В этом случае включение члена с невязкой также улучшает сходимость метода.

Предложенная нами целевая функция  $W$  [A8] в расширенном варианте имеет следующий вид:

$$\begin{aligned} W(\mathbf{y}) = & \alpha \cdot (RES - RES_{\lim}) + \beta \cdot [\text{cond}(X) - 1]^2 + \gamma \cdot \frac{1}{N \cdot K} \cdot \sum_{\forall(x_{ik} < 0)}^{N, K} x_{ik}^2 + \\ & + \delta \cdot \frac{1}{M \cdot K} \cdot \sum_{\forall(c_{k,m} < 0)}^{K, M} c_{ik}^2 + \varepsilon \cdot \frac{1}{N \cdot K} \cdot \sum_{k=1}^K \text{len}(\mathbf{x}_k) + \\ & + \eta \cdot \frac{1}{N \cdot K} \cdot \sum_{k=1}^K \text{cur}(\mathbf{x}_k) + \xi \cdot \frac{2}{K \cdot (K - 1)} \cdot \sum_{\forall(l \neq m)}^K [\cos(\mathbf{x}_l, \mathbf{x}_m)] + \\ & + \tau \cdot \frac{2}{K \cdot (K - 1)} \cdot \sum_{\forall(l \neq m)}^K r(\mathbf{x}_l, \mathbf{x}_m) + \dots \end{aligned} \quad (1.50)$$

где  $X$  - матрица пробных спектров компонентов  $\mathbf{x}_l$ ,  $\mathbf{x}_m$  -  $l$ -й и  $m$ -й вектор-спектры компонентов, составляющих матрицу  $X$ ,  $i = 1, \dots, N$  - индекс точки в спектре,  $k = 1, \dots, K$  - индекс компонента,  $r$  - коэффициент корреляции между двумя спектрами.

Несколько подробнее надо остановиться на втором члене  $[cond(X) - 1]^2$ . Здесь  $cond(X)$  означает число обусловленности матрицы нормированных пробных спектров компонентов  $X$ , вычисляемое как отношение максимального к минимальному сингулярным числам  $\lambda_x$  - диагональных элементов матрицы  $\Lambda_x$  из разложения

$$X = U_x \Lambda_x V_x^T, \quad cond(X) = \frac{\lambda_{\max}}{\lambda_{\min}}. \quad (1.51)$$

Число обусловленности в данном случае отражает суммарное сходство вектор-спектров компонентов, достигая 1 в случае их максимального формального различия (но в этом случае, будучи ортогональными, векторы компонентов должны содержать отрицательные элементы). Вычитание единицы из критерия сходимости обусловлено именно этим фактом. С геометрической точки зрения, меньшие значения  $cond(X)$  соответствуют большим значениям телесного угла, образованного векторами компонентов  $\mathbf{x}_j$ .

Третий и четвертый члены в (1.50) есть штрафы за наличие отрицательных элементов в модельных спектрах и за отрицательность элементов концентрационной матрицы, определяемой соотношением (1.47).

Пятое слагаемое, содержащее выражение

$$len(\mathbf{x}_k) = \sum_{i=1}^{N-1} \left[ h^2 + (\mathbf{x}_{i+1,k} - \mathbf{x}_{ik})^2 \right]^{1/2} \quad (1.52)$$

есть длина нормированного по максимальному элементу спектрального контура, аппроксимированного ломаной кривой с шагом  $h$ . Длину шага по абсциссе  $h$  можно выбирать достаточно произвольно, например, так, чтобы  $h^2$  и  $\mathbf{x}_{i+1,k} - \mathbf{x}_{ik}$  в среднем были соизмеримы.  $h = 0$  также оказывается неплохим выбором в большинстве случаев, что и реализовано в программе REMIX. Выражение для длины выступает в качестве штрафа за сложность спектрального контура. Как

уже указывалось, контур спектра нормируют перед вычислением длины, деля его на значение максимальной интенсивности. Кажется более разумным использовать нормировки к единичной евклидовой норме, или к норме 1 (сумме модулей элементов), однако численная проверка показала, что чувствительность целевой функции к взаимопроникновению контуров спектров компонентов при таких масштабированиях значительно ухудшается.

Остальные слагаемые в (1.50) являются избыточными, или альтернативными соответствующим слагаемым, рассмотренным выше. Так, шестое слагаемое содержит выражение

$$cur(\mathbf{x}_k) = \sum_{i=2}^{N-1} \left| (\mathbf{x}''_{ik}) \right|, \quad (1.53)$$

которое пропорционально кривизне спектрального контура, вычисляемой как сумма по вторым производным от нормированного по максимальному элементу вектор-спектра. Этот критерий также рассматривается как критерий сложности спектра. Практика решения модельных и реальных задач показала, что включение этого члена в (1.50) в редких случаях помогает достичь более гладких контуров спектров, но, как правило, его использование ухудшает обусловленность задачи и препятствует получению приемлемого решения.

7-й и 8-й члены в (1.50) (суммы косинусов углов и коэффициентов корреляций между векторами компонентов как критерии попарного сходства спектров) также являются по своему смыслу избыточными и соответствуют второму члену в сумме (1.50), так как их минимизация также ведет к увеличению углов между векторами компонентов. Член  $[cond(X) - 1]^2$  отражает интегральную по всем компонентам характеристику и при большом сходстве некоторых компонентов между собой сходимость алгоритма ухудшается из-за ухудшения обусловленности задачи. В этом случае 7-й и/или 8-й члены, непосредственно выражающие попарное различие спектров, могут несколько улучшить скорость сходимости поиска минимума (1.50). Так 7-й и 8-й члены выражают одно и то же требование к решению, их одновременное использование вряд ли целесообразно.

Введение нескольких аналогичных по своей сути штрафных членов в целевую функцию полезно на практике, так как такие слагаемые обладают разной чувствительностью к деталям спектральных контуров, при этом исследователь имеет возможность подобрать соответствующие весовые коэффициенты для конкретного класса решаемых задач и добиться максимально эффективной работы алгоритма.

Еще одним критерием сходства (или различия) спектров, иногда используемым другими авторами в своих работах, является объем многогранника, образованного концами особым образом нормированных векторов-спектров компонентов [28]. Применение этого критерия резко усложняется с ростом числа компонентов, т. к. требует большого объема вычислений по построению многогранника, имеющего максимально возможный объем (для этого в [28] с достаточно малым шагом генерируют всевозможные пробные неотрицательные спектры компонентов и отбирают наилучший вариант). В литературе по статистике, помимо коэффициента корреляции, известно несколько критериев сходства объектов. Это Евклидова метрика, определяемая квадратом нормы

разности двух векторов, манхеттенская метрика  $\sum_{i=1}^N |a_i - b_i|$ , метрика Брея –

Кертиса  $\frac{\sum_{i=1}^N |a_i - b_i|}{\sum_{i=1}^N a_i - \sum_{i=1}^N b_i}$ , Канберровская метрика  $\sum_{i=1}^N \frac{|a_i - b_i|}{a_i - b_i}$  и некоторые другие, в

том числе использующие величину информационной энтропии [36]. Проверка соответствующих выражений показала, что они существенно ухудшают свойства целевой функции в окрестностях решения из-за невысокой чувствительности к малым изменениям формы спектральных контуров.

Вообще говоря, можно предлагать самые разнообразные критерии сложности или простоты спектров, но трудно представить достаточно веские

основания для их использования, так как связь таких критериев с физическим смыслом и особенностями конкретных задач весьма опосредована и поэтому трудно формализуема. Наша практика решения модельных и реальных задач разложения показала, что величина информационной энтропии (05) в качестве критерия сложности для УФ, видимых, ИК и масс-спектров оказывается весьма непредставительной и не может обеспечить приемлемой скорости сходимости процедуры минимизации (по крайней мере, при числе компонентов более двух, чем, может быть, и объясняется тот факт, что в литературе в качестве примеров разложения приводятся не более чем 2-3 компонентные задачи), поэтому от использования критерия (05) мы отказались.

Часть слагаемых в (1.50) имеют квадратичную форму. Последнее, как отмечается в литературе по методам нелинейной минимизации [89-92], должно обеспечивать приемлемую скорость сходимости процедуры поиска решения при использовании квазиньютоновских алгоритмов. Но, начиная с пятого слагаемого, возведение их в квадрат ухудшает обусловленность задачи в окрестности минимума и увеличивает разброс решений при малых вариациях весовых коэффициентов при штрафных членах.

**Таким образом, минимизация функционала (1.50) означает поиск заданного количества  $K$  спектров, лежащих в подпространстве спектров смесей и, значит, формально могущих быть спектрами компонентов, при условиях их максимального различия и простоты формы контуров.**

Итак, выражение (1.50) представляет собой штрафную функцию. Известно [89-92], что метод штрафных функций далеко не оптимален с точки зрения устойчивости решения задачи – число обусловленности матрицы вторых производных целевой функции часто оказывается настолько большим, что численная процедура поиска минимума иногда не в состоянии достичь окрестности решения. Тем не менее мы выбрали именно этот подход, так как он позволяет расширять целевую функцию простым добавлением новых штрафных членов, учитывающих любую дополнительную информацию, если ее можно формализовать в виде квадратичного слагаемого.

Рассмотренная структура целевой функции соответствует учету дополнительных **требований** к решению и **ограничений** на него. Использование **дополнительной информации** можно также свести к построению штрафных слагаемых в (1.50). Но этого можно избежать в случае, когда в качестве дополнительных сведений выступают известные спектры некоторых компонентов. В таком случае их можно просто включить в матрицу  $X_n$  и зафиксировать, запретив их изменение (по крайней мере в начале поиска). Тем самым мы избегаем усложнения целевой функции и уменьшаем число неизвестных.

В этом месте необходимо сделать следующее замечание. Включаемые в матрицу компонентов априори известные спектры должны быть сняты с высокой точностью, так как даже небольшие систематические ошибки в интенсивности при плохой обусловленности задачи могут привести к большим искажениям искомым спектров, вплоть до невозможности их идентификации. Автор рекомендует не использовать известные спектры при решении задачи разложения, если нет уверенности в их точности, но контролировать с их помощью получаемые решения, проводя сопоставление известных спектров с найденными. Критерии сопоставления были рассмотрены ранее.

Нами были реализованы и проверены практически все основные известные методы минимизации нулевого и первого порядков (т.е. не использующие и использующие первые производные целевой функции), представленные в литературе. Было найдено, что наиболее устойчивыми (т.е. достигающими окрестности решения без дополнительных рестартов и других искусственных приемов) оказались метод деформируемого многогранника Нелдера-Мида (метод нулевого порядка), метод переменной метрики в варианте Бройдена-Флетчера-Голдфарба-Шанно (класс квазиньютоновских алгоритмов первого порядка) и метод, известный в литературе под названием NL2SOL [93,94] (комбинация алгоритмов Левенберга-Маркардта решения нелинейных задач наименьших квадратов и переменной метрики Бройдена-Флетчера-Голдфарба-Шанно). Предлагаемый метод разделения реализован нами в виде диалоговой

программы REMIX, применение которой было показано в [A8] на примере разделения спектров 6-компонентных смесей.

Величины весовых коэффициентов нами предложено выбирать следующими (при условии нормировки матрицы спектров смесей к максимальной среди всех спектров интенсивности, и пробных спектров компонентов - как это было описано выше при рассмотрении слагаемых целевой функции):  $\alpha = 2 \cdot 10^4$ ,  $\beta = 1.0$ ,  $\gamma = 10^5$ ,  $\delta = 10^3$ ,  $\varepsilon = 10^2$ ,  $\eta = 5 \cdot 10^2$ ,  $\xi = 2 \cdot 10^2$ ,  $\tau = 10^2$ . Величины коэффициентов были выбраны из соображения примерного равенства слагаемых целевой функции в точке минимума, кроме первого члена, вклад которого достаточен на уровне  $\sim 0.01$  от вкладов остальных членов. Вклад слагаемых (1.52) и (1.53) желательно поддерживать на уровне 10-30 от вклада  $\beta \cdot [\text{cond}(X) - 1]^2$ . Практика показывает, что скорость сходимости к решению при таком выборе достаточно высока в большинстве случаев и решение отвечает выбранным требованиям.

Работу алгоритма REMIX мы продемонстрируем на примере анализа 6-компонентной смеси органических веществ (набор №2 из раздела 1.4.2) [A8]. Набор исходных спектров показан на Рисунке 1.4, оценка числа компонентов для этой задачи рассмотрена в разделе 1.4.3. На Рисунке 1.13 показаны решения при гипотезах  $K = 6, 7, 8$  и числе используемых при моделировании компонентов сингулярных векторов в (47)  $L = 6, 7, 8$ , соответственно.

Данный эксперимент демонстрирует, что увеличение числа модельных компонентов не приводит к существенным искажениям спектральных контуров 1 – 6. Вместе с тем введение дополнительных компонентов позволяет очистить некоторые из модельных компонентов от полос, которые обусловлены нарушениями аддитивности из-за взаимодействия компонентов между собой.

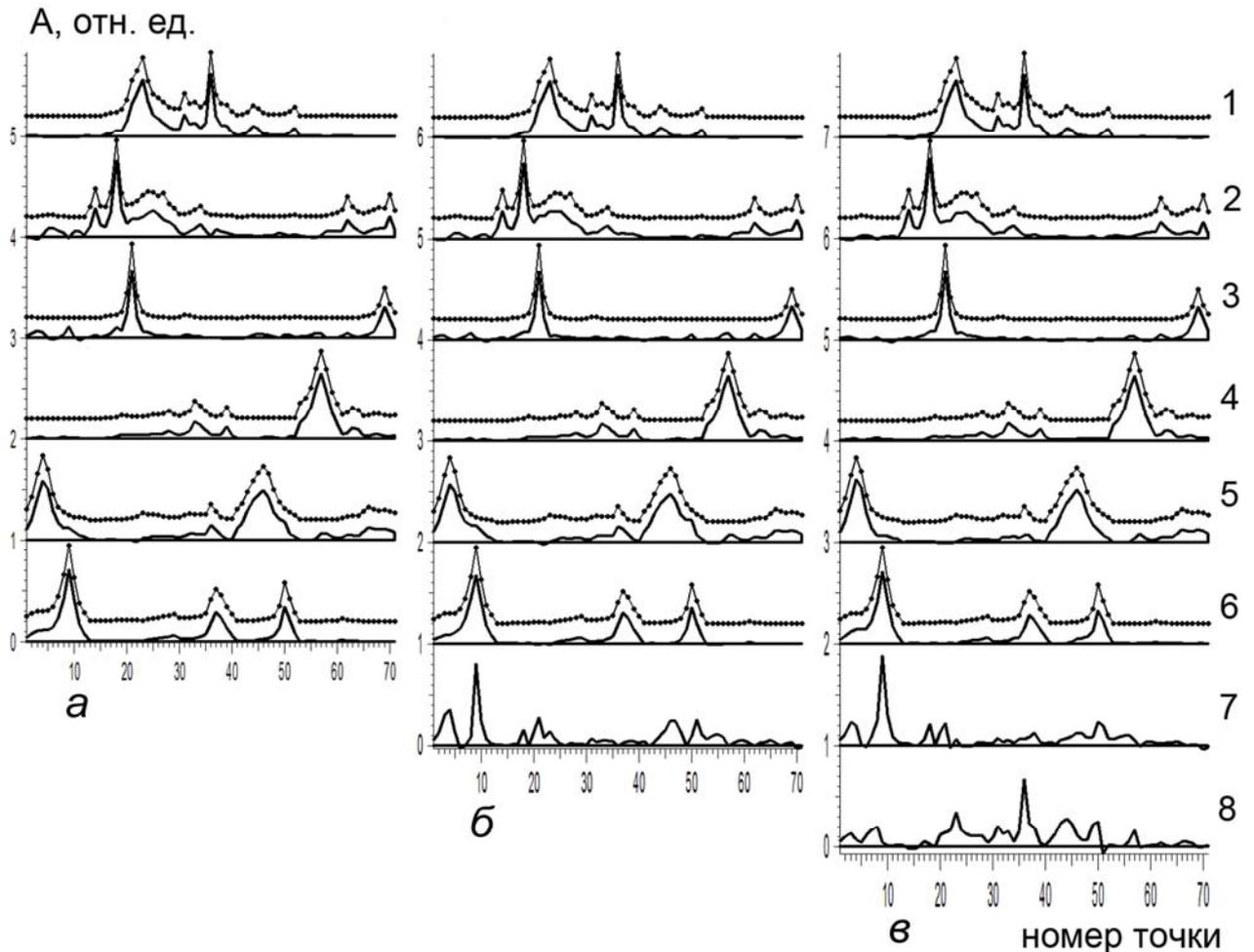


Рисунок 1.13 - Результаты моделирования спектров компонентов по набору из 10 спектров смесей (Рисунок 1.4). Жирными линиями показаны модельные спектры, линиями с точками – экспериментально полученные спектры чистых веществ: 1 – изооктана, 2 – толуола, 3 – бензола, 4 – диэтилового эфира, 5 – амилацетата, 6 – ацетона. **а) – в)** – решения для гипотез о 6, 7 и 8 компонентах, соответственно.

Линии нулевой интенсивности показаны для модельных спектров, спектры чистых веществ смещены по вертикали для лучшей визуализации.

В частности, это видно на примере компонентов 2 и 3: при  $K = 6$  (решение **а)** интенсивность пиков в области точек 1-15 уменьшается, а сами пики проявляются в компоненте 7, который содержит спектральную информацию о не полностью скомпенсированном в спектрах смесей растворителе – четыреххлористом углероде и сдвигах полос в спектрах из-за межмолекулярных взаимодействий. Следовательно, оценка числа компонентов 7 (см. раздел 1.3.4) оказывается ближе к истине. При  $K = 8$  спектр 8 не удается идентифицировать из-за большого числа полос, основная из которых коррелирует с полосой в спектре 1. По-видимому, данный компонент является суммарным спектром примесей в

изооктане, использованном для приготовления модельной смеси. Кроме того, при завышенном числе компонентов программа поиска вынуждена конструировать неотрицательный контур с соответствующим неотрицательным спектральным вкладом в смеси. Этот контур, вследствие ограничений на неотрицательность, может представлять собой сумму контуров остальных компонентов плюс вклад от "шумового" сингулярного вектора №8, участвующего в построении модельной матрицы  $X$  и его интерпретация должна проводится с осторожностью.

Распределение концентраций основных компонентов для решения  $K = 7$  показано на Рисунке 1.14. Из рисунка видно, что вклад 7-го компонента незначителен.

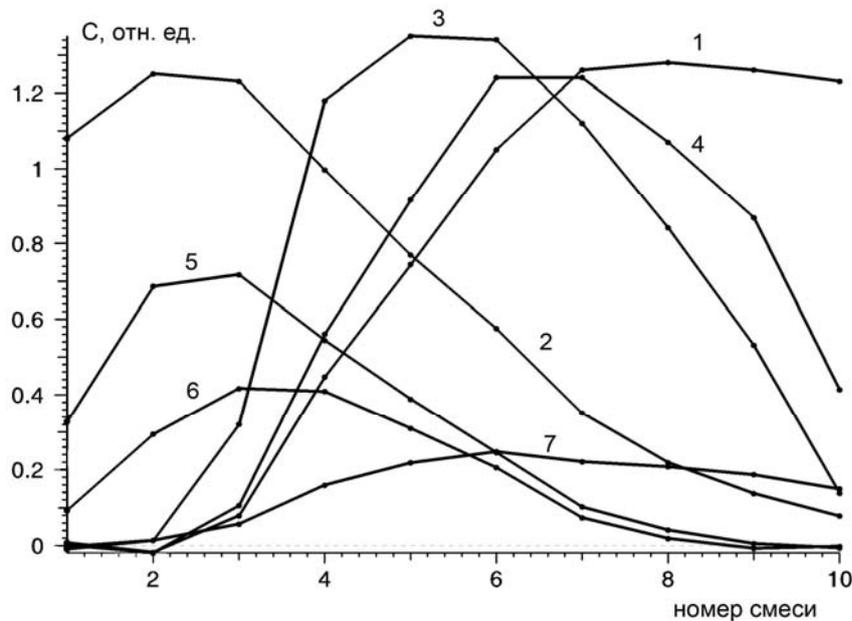


Рисунок 1.14 – Концентрационные вклады компонентов для решения б (Рисунок 1.13), рассчитанные по уравнению (02). Номера компонентов соответствуют Рисунку 1.13.

Матрицы остатков  $F_K^\# = D - X_K \cdot C_K$  при  $K = 5, 6, 7$  и  $8$ , представленные в виде контуров, составленных из столбцов  $F_K^\#$ , показаны на Рисунке 1.15. Видно, как с ростом числа компонентов в модели уменьшается амплитуда элементов матрицы остатков. Критерий Вальда-Вольфовица (1.32) для гипотезы о случайности последовательности амплитуд элементов матрицы остатков не показывает в данном случае (и, как отмечено в разделе 1.3.2, и во многих других

случаях) отсутствия корреляций и не может обеспечить определения числа компонентов. Критерий Дарбина-Ватсона  $DW$  (1.14) оказывается менее чувствителен к быстро меняющимся систематическим отклонениям, причина которых кроется в межмолекулярных взаимодействиях компонентов, как видно из Таблицы 4 и, при доверительном интервале  $\{1.7-2.2\}$  показывает 8 компонентов. Данный пример демонстрирует типичную сложность, возникающую при обработке реальных систем, связанную с тем, что дисперсия шумов измерений может значительно варьироваться от смеси к смеси и делать статистический анализ величины дисперсии матрицы остатков (предполагающий одинаковую дисперсию для всех спектров) неравномерным. Предлагаемое нами использование тестов на отсутствие корреляций в последовательности остатков в таких случаях в значительной степени свободно от таких ограничений.

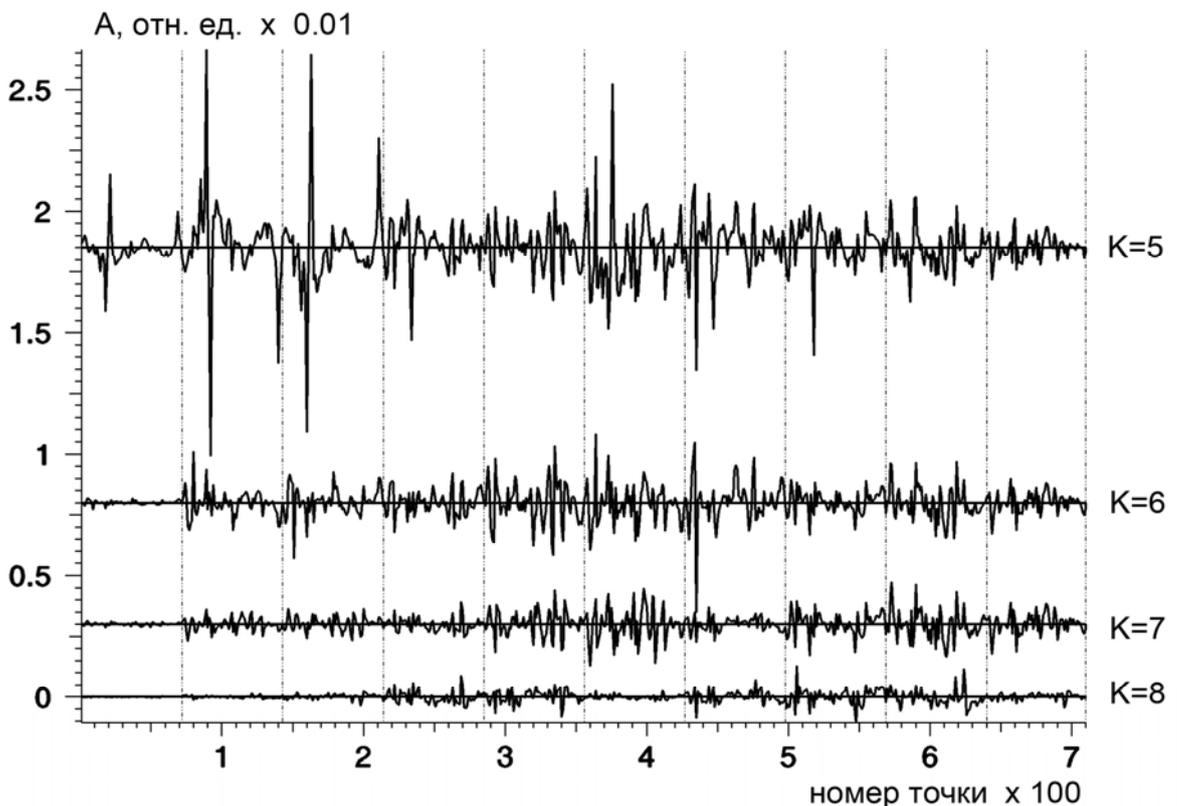


Рисунок 1.15 – Представление матриц остатков при разном числе компонентов  $K$  в виде одномерных контуров. Штриховые вертикальные линии обозначают границы спектров смесей в матрице данных (71 точка в спектре).

Таблица 1.5. Величины критериев автокорреляции и случайности элементов матриц остатков для разного числа компонентов в модели.

Число компонентов $K$	Критерий Вальда-Вольфовица (1.32)		Критерий Дарбина-Ватсона (1.14)	
	WW	Доверитель- ный интервал при $Q=0.05$		Доверитель- ный интервал при $Q=0.05$
5	-12.2	-3.13, 3.13	1.16	1.65, 2.18
6	-6.38		1.56	
7	-6.23		1.59	
8	-7.35		<b>1.68</b>	

Для демонстрации особенностей разложения в случае невыполнения условий неперекрываемости спектров компонентов был подготовлен модельный набор спектров компонентов и смесей, показанный на Рисунке 1.16а,в. Шумовая составляющая в спектры смесей не вводилась для лучшей демонстрации особенностей форм найденных спектров компонентов. Дополнительной сложностью был значительный вклад каждого компонента в смеси (Рисунок 1.16д).

Решение 1.16,б получено при использовании первых пяти слагаемых в формуле (48) и указанных выше весовых коэффициентах. Видно, что спектр третьего компонента достигает нулевого значения в районе 5-й точки. Увеличение штрафа за относительную длину контура спектров в 3 раза приводит к правильной форме 3-го компонента, но спектр 2-го компонента оказывается слегка искажен добавкой 3-го контура. Таким образом, подбирая относительные величины весов штрафных членов, можно влиять на качество решения и исследовать взаимовлияние находимых индивидуальных контуров. Конечно, в сложных случаях при необходимости можно увеличивать веса штрафов для каждого компонента индивидуально в диалоговой итерационной процедуре.

С помощью разработанного программного обеспечения был проведен анализ состава комплексов, образующихся в системе  $NiCl_2-CO(NH_2)_2-H_2O$  методом электронной спектроскопии [Аб]. Были определены приближенные спектры комплексов в их смесях и сделаны выводы об их составе. На конкретных

результатах этой работы мы здесь останавливаться не будем, а рассмотрим в следующем разделе более показательный пример попытки разложения сильно перекрытых малоинформативных спектральных контуров, для которых разработанные методы, строго говоря, не могут считаться применимыми. Тем не менее, с помощью предложенных подходов можно получать полезную информацию.

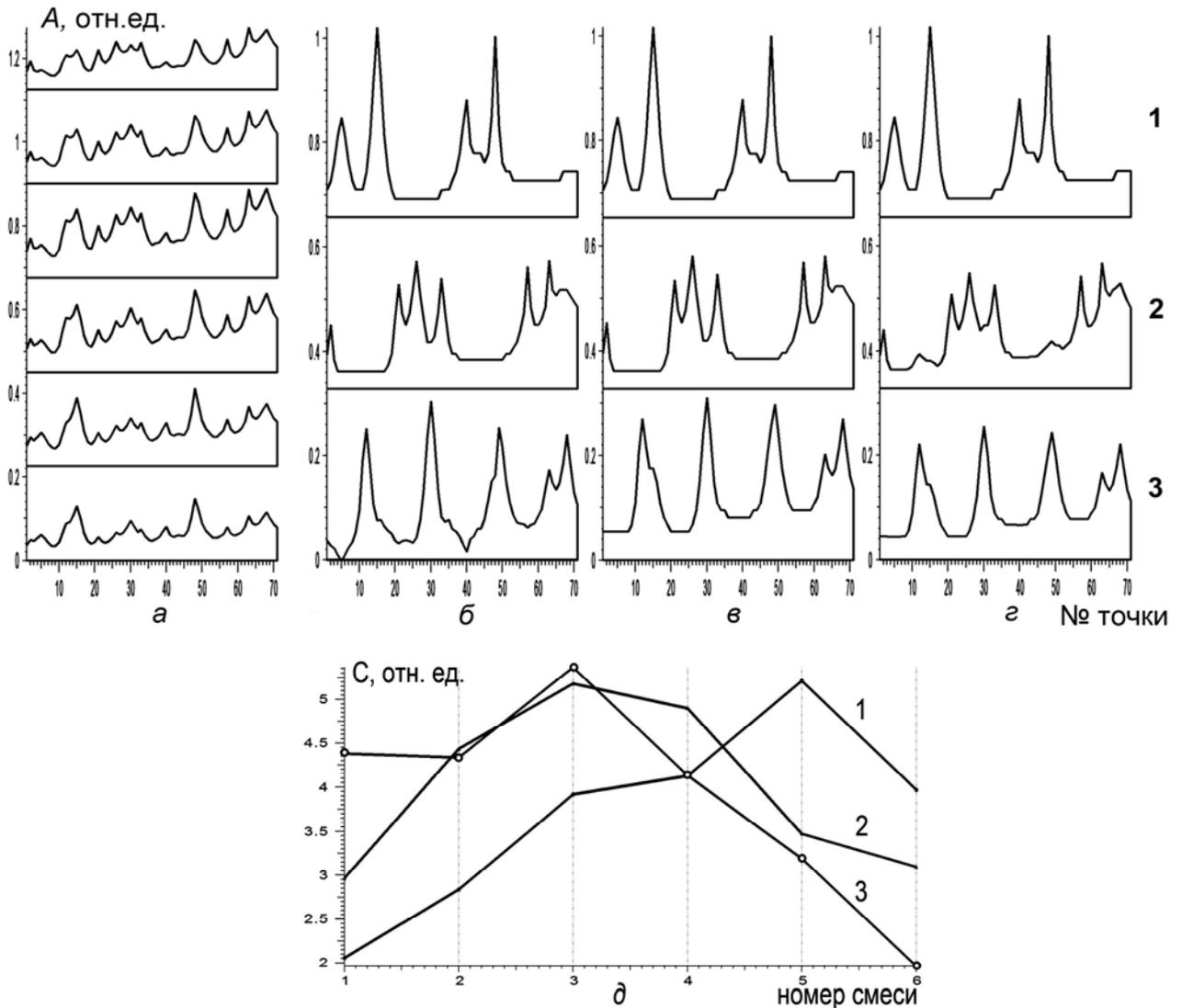


Рисунок 1.16 – Модельная 3-х компонентная задача с полностью перекрытыми спектрами компонентов. **а)** – спектры смесей, **б)** – решение при весах 10 – 10 - 10, **в)** – исходные спектры компонентов, **г)** – решение при весах 10 – 10 - 50, **д)** – относительные концентрации. Спектры разнесены по вертикали для лучшей визуализации, нулевой уровень интенсивности показан горизонтальными линиями для каждого спектра. Цифрами 1, 2, 3 обозначены номера компонентов.

### 1.4.3.1 Анализ колебательных спектров жидкой воды

С помощью разработанной процедуры был исследован набор колебательных спектров жидкой воды при разных температурах - 1, 16, 39 и 50°C. [A8]. Данный пример иллюстрирует сложность интерпретации спектров смесей при большой ширине спектральных линий и недостаточном количестве исходных спектров. Первые измерения спектров колебаний жидкой воды показали, что полуширина линии при длине волны 3 мкм группа очень большая, около 400 см<sup>-1</sup> [95]. Первые объяснения этого факта базировались на предположениях, что группы ОН могут образовывать водородные связи с необычными параметрами. И преддиссоциативная [96] и модульная [97] теории объясняли большую ширину полосы  $\nu_{\text{OH}}$  дисперсионными особенностями водородной связи. Одновременно с этими теориями, совершенно разные объяснения аномальной ширины полосы были предложены в [98], основанные на значительных отклонениях геометрических параметров структуры и результирующей дисперсии  $\nu_{\text{OH}}$  частот ОН-подобных молекулярных ансамблей в жидкой воде. Это явление было зарегистрировано для разных объектов и названо позже неоднородностью уширения.

Через шесть лет после первых исследований полосы  $\nu_{\text{OH}}$  было установлено, что она имеет сложную структуру [99]. Спустя два десятилетия было показано, что она состоит как минимум из трех компонентов [100,101]. С этого времени основной упор делается не на форму полосы, а на количество и параметры ее составляющих. Однако эту проблема так и не удалось решить однозначно. Хотя результаты, полученные разными авторами, привели почти к одинаковым формам контура  $\nu_{\text{OH}}$ , интенсивность и положение компонентов значительно отличались. Разница была намного больше, чем погрешности эксперимента. Новые методы измерения спектров поглощения (в частности, нарушенного полного внутреннего отражения) позволили уменьшить ошибки эксперимента. Наконец, спустя почти пятьдесят лет с момента первого эксперимента, согласованные значения оптических констант были получены при

четырёх различных температурах в одном эксперименте [102]. До сих пор эти данные остаются непревзойденными [103].

В соответствии с теорией колебаний, два из компонентов  $\nu_{\text{OH}}$  считаются валентными колебаниями и третий обертоном - деформационными колебаниями молекул воды. Анализ положения максимумов компонент показал, что эти спектральные составляющие вызваны колебаниями несимметричных молекул воды, которые имеют два неэквивалентных положения ОН-групп. А именно, сила водородной связи образованной с одной из ОН-групп в среднем в 1,5 раза больше, чем прочность второй [104]. Кроме того, было очевидно, что предположений о наличии дополнительных слабых компонентов полосы ОН-контура недостаточно, чтобы обеспечить разумное решение проблемы структуры. Дело в том, что спектральные полосы, соответствующие двум валентным колебаниям молекулы воды, находящейся в окружении четырех ближайших соседей, могут быть размещены в произвольно широкой области  $\nu_{\text{OH}}$  контура (3700-3000  $\text{cm}^{-1}$ ). Поэтому для анализа набора  $\nu_{\text{OH}}$  спектров был применен разработанный алгоритм разделения, не требующий никаких априорных знаний о форме составляющих и было интересно оценить не только число компонент, но и их форму. Было очевидно, что для компонент в данном случае не выполняются условия взаимной неперекрываемости, но приближенное решение часто содержит полезную информацию, поскольку неопределенность расчетных контуров только увеличивает различия между отдельными спектрами. В результате расчетные спектры не могут точно соответствовать структуре фрагментов, но могут помочь выяснить основные черты структуры, а также оценить общее количество фрагментов и их относительный вклад в измеренные спектры.

К сожалению, с приемлемой достоверностью до настоящего времени удалось измерить только четыре спектральных контура: при 1, 16, 39 и 50°C (Рисунок 1.17а). Сингулярные векторы матрицы данных показаны представлены в виде "спектров" на Рисунке 1.17б.

Анализ сингулярных векторов не позволяет в данном случае оценить число компонент, которое, скорее всего, превышает число участвующих в

анализе смесей - 4. Поэтому был проведен расчет 4 компонентов, контуры спектров которых показаны на Рисунке 1.18а. По результатам разложения был сделан вывод, что спектры воды состоят из линейных комбинаций не менее четырех компонентов и сеть водородных связей состоит не менее чем из 4 основных фрагментов структуры. Вклад двух основных компонентов составляют около 94%. Другие фрагменты структуры присутствуют в системе в незначительных количествах. Этот вывод согласуется с экспериментальными результатами рентгеноструктурного исследования воды при комнатной температуре [105].

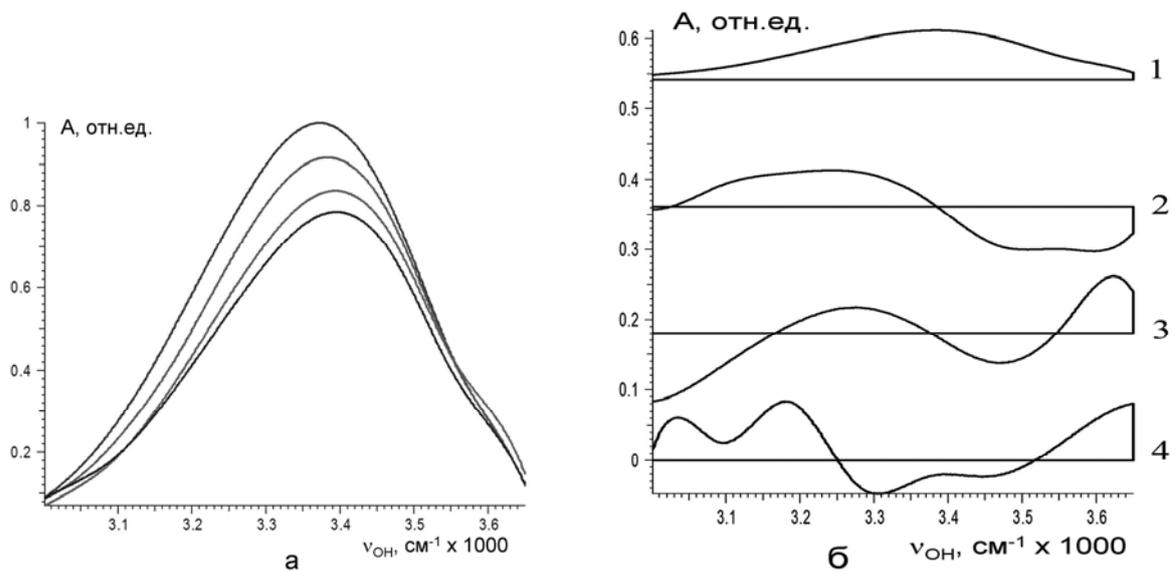


Рисунок 1.17. а): колебательные спектры воды при 1, 16, 39 и 50°C (снизу вверх); б): сингулярные векторы (смещены по вертикали).

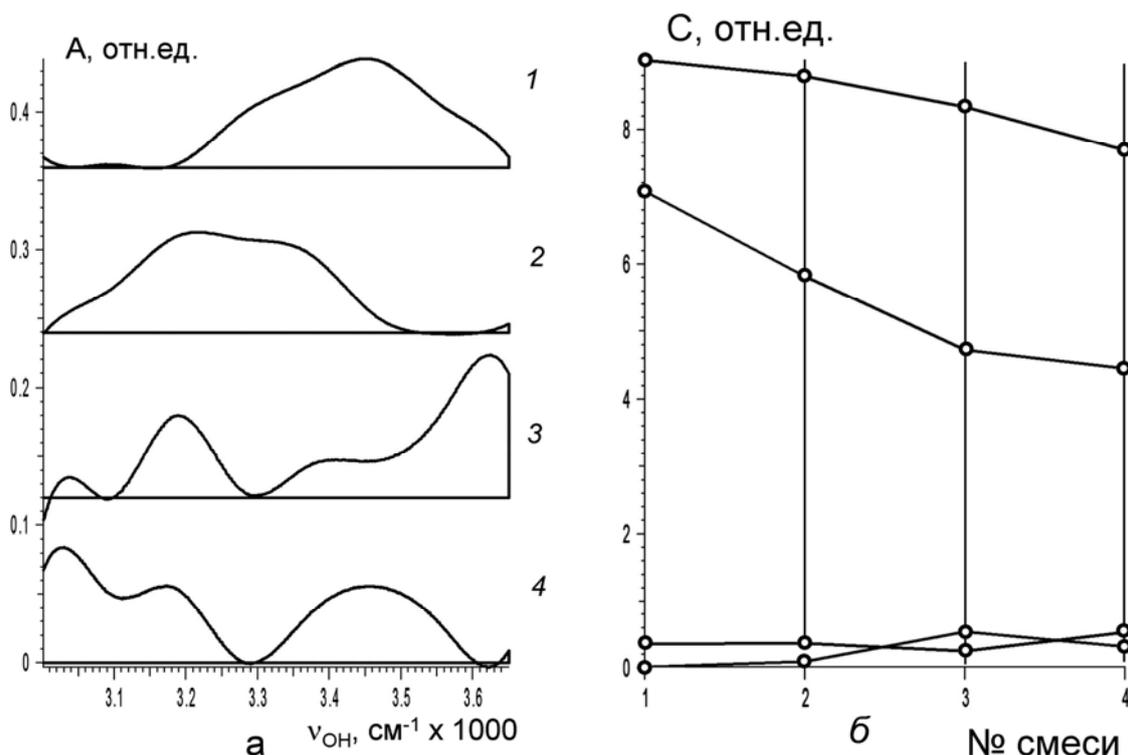


Рисунок 1.18. а): спектры компонентов, б) относительные вклады компонентов в смеси.

Можно увидеть, что спектральный контур ОН-колебаний жидкой воды состоит из нескольких спектральных полос (Рисунок 1.18а) и соответствует не менее чем 4 основным фрагментам структуры. Полученные результаты характеризуют лишь отдельные фрагменты, а не структуру в целом. Следовательно, нельзя сделаны определенные выводы об особенностях самих фрагментов. Тем не менее, можно сделать некоторые общие выводы. Подавляющее большинство молекул воды в жидком состоянии остаются симметричными. Одна половина этих молекул образует прочные водородные связи с окружением, примерно в 1,5 раза сильнее, чем связи второй половины "слабо связанных" молекул. Единственная возможность сосуществования фрагментов заключается в их разделенном размещении внутри структурных кластеров при распределении асимметричных молекул на их границах. Общее количество асимметричных молекул тем меньше, чем больше размер кластера.

Как уже было сказано выше, структура фрагментов, состоящих из молекул с несимметричным окружением, действительно существует в жидкой воде в

незначительном количестве. Судя по их найденной доле в 1/15, размер кластера должен иметь усредненный размер около  $10^{-6}$  см (примерно 100 молекул на кластер). Следует отметить, что если доля несимметричных молекул (6%) недооценена (в сравнении с вероятным значением 15%), то кластер должен содержать еще несколько десятков тысяч молекул воды. Таким образом, полученные результаты свидетельствуют о кластерной структуре колебаний сетки водородных связей в жидкой воде.

Для повышения информативности разложения следует увеличить количество смесей, участвующих в разложении. Однако, точное измерение контура  $\nu_{OH}$  является сложной задачей, не решенной с удовлетворительной точностью до сих пор.

### 1.4.3.2 Анализ спектров при наличии межмолекулярных взаимодействий: смеси диметилформаида с гексафторэтиленом

Анализ спектров равновесных смесей, в которых компонентами являются соединения, существующие только в равновесии с другими компонентами оказывается сложной задачей, решение которой часто возможно только с применением численных методов. В данном разделе мы рассмотрим пример численного анализа смесей сильно взаимодействующих компонентов. Такой анализ не позволяет выделить истинные спектры изолированных компонентов, но, по крайней мере, разделить наборы спектральных полос по группам, ведущим себя как компоненты при варьировании состава смесей и/или условий их существования.

На Рисунке 1.19 показан фрагмент 11 спектров смесей при разных относительных концентрациях гексафторэтилена. В этих смесях ожидалось образование 3 типов гетероассоциатов с соотношением гексафторэтилен : ДМФА = 1 : 1, 4 : 1 и 10 : 1.

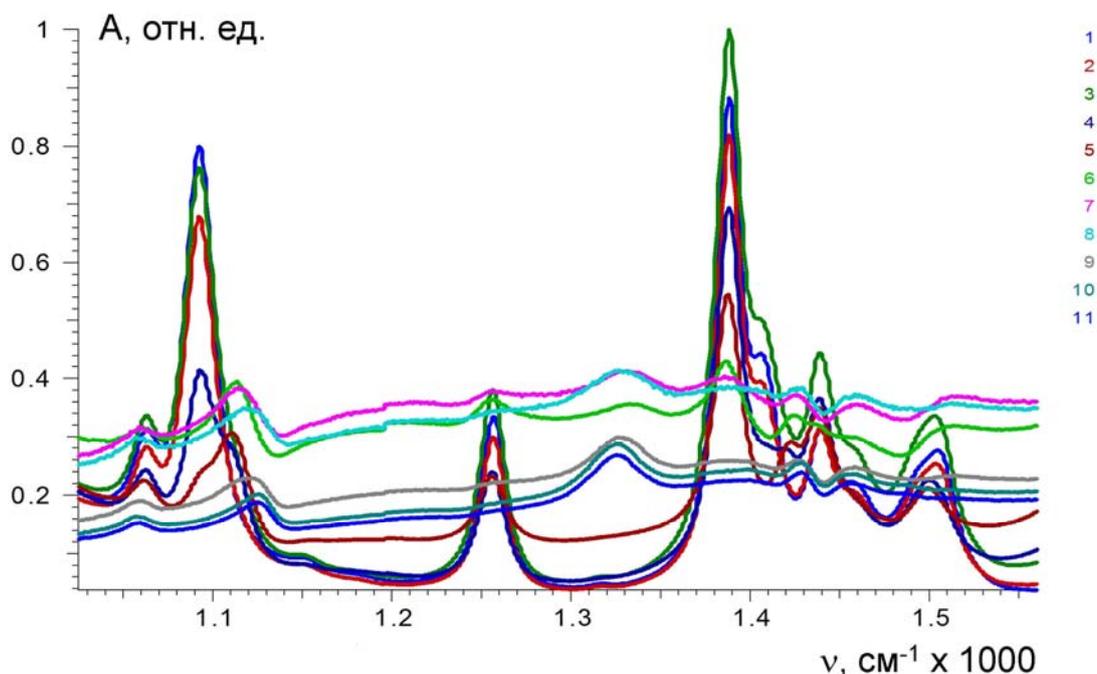


Рисунок 1.19 – ИК спектры смесей гексафторэтилена с ДМФА.

Анализ сингулярных векторов матрицы данных показал присутствие не менее 11 компонентов – критерий Дарбина-Ватсона не превышал  $0.7 < DW < 1.7$  для всех последовательностей. На основании того, что число компонентов равно числу смесей можно предположить, что в системе либо присутствует значительно большее число ассоциатов, чем ожидаемое 3, либо межмолекулярные взаимодействия делают систему нелинейной. Сингулярные векторы показаны в виде спектров на Рисунке 1.20. Форма контуров несет информацию об изменении формы спектральных полос при взаимодействиях компонентов. Из Рисунка видно, что вклад последних трех векторов в данные относительно мал, и 11-й вектор практически соответствует шумам измерений. Разложения на 10 и 8 компонентов показывают (Рисунке 1.21), что контуры индивидуальных спектров близки между решениями и соответствуют трем характерным формам (спектры 2, 3 и 6 в 8-компонентном решении). Остальные компоненты содержат полосы, которые соответствуют сдвигам основных полос при взаимодействиях компонентов смеси. Один из контуров (4-й) соответствует нескомпенсированному фону. Интересно отметить, что 6-й компонент 8-компонентного решения оказался в смеси с 7-м и 8-м контурами, что показывает его дальнейшее расщепление в 10-компонентном решении.

Пример анализа данного набора спектров показывает, что форма полос поглощения 3-х основных компонентов варьируется при изменении их относительных концентраций и остальные вычисленные компоненты содержат информацию о групповом изменении формы полос, не соотносясь непосредственно с химическими веществами, но предоставляя сведения о наличии и характере межмолекулярных взаимодействий.

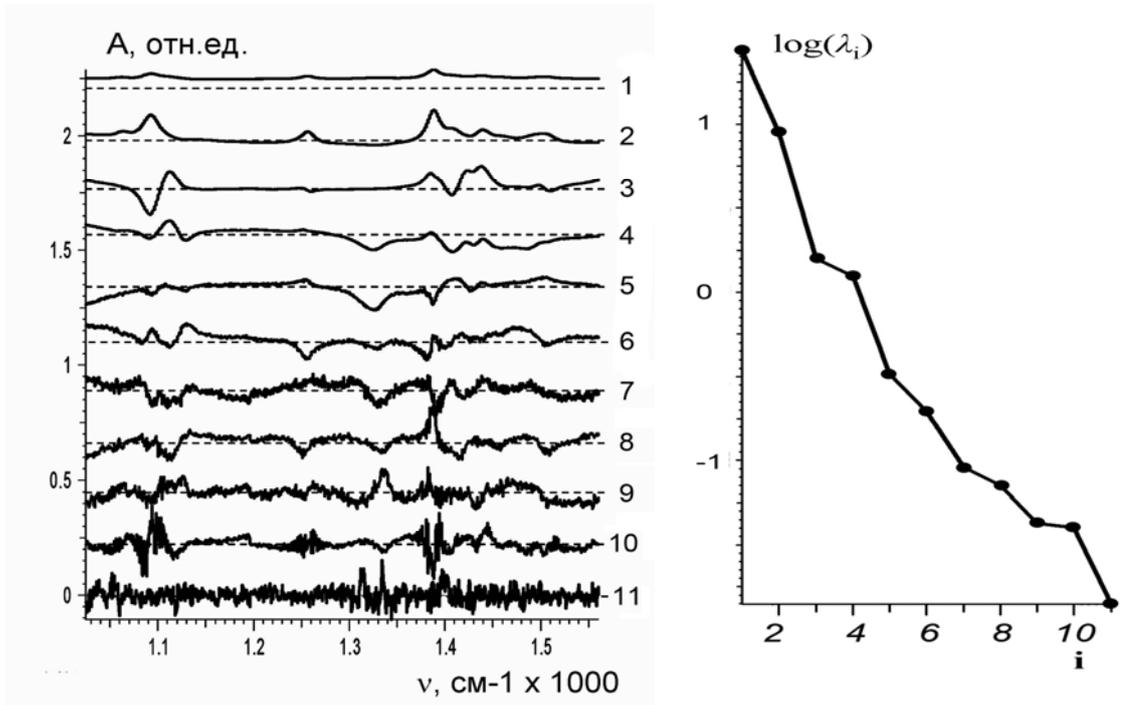


Рисунок 1.20 – Сингулярные векторы матрицы данных, показанных на Рисунке 1.19. Справа – спектр логарифмов сингулярных чисел.

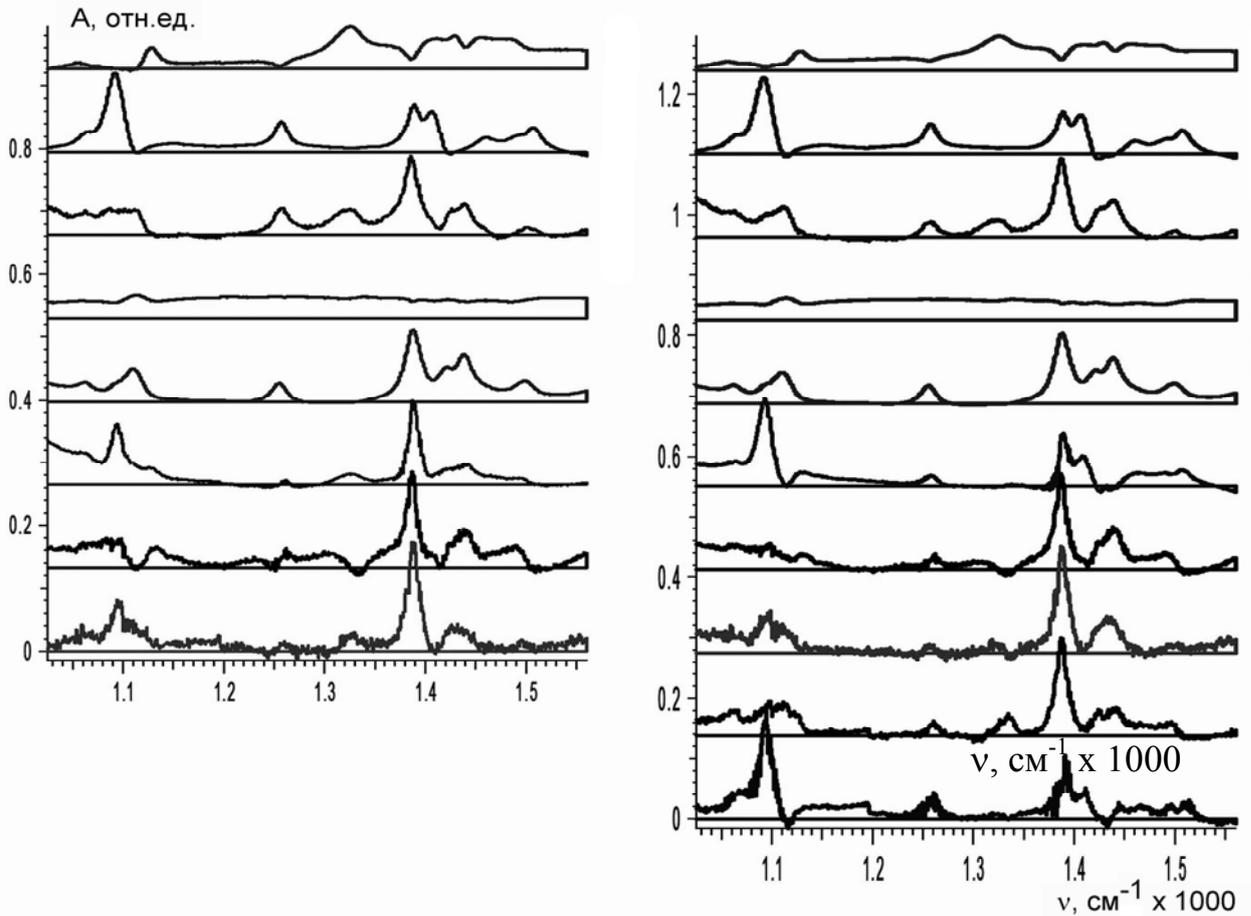


Рисунок 1.21 – Разложение набора данных Рисунка 1.19 на 8 (слева) и 10 компонентов (справа).

### 1.4.3.3 Анализ устойчивости решения

Приведенные примеры демонстрируют, что вопрос о том, каково максимально допустимое число компонентов в смесях, чтобы их можно было устойчиво разложить на спектры частично неперекрываемых компонентов часто не имеет однозначного ответа. На успех решения задачи разложения влияют такие факторы, как величина обусловленности концентрационной матрицы (чем лучше разделение компонентов при получении набора родственных смесей, тем меньше  $cond(C)$  и тем лучше сходимость итерационных методов поиска), число обусловленность матрицы  $D$  (чем более непохожи и неперекрываются спектры компонентов, тем меньше  $cond(D)$  и лучше работают все методы поиска решения), уровень спектральных ошибок, применяемый алгоритм поиска и т.д. Число обусловленности матрицы определяется отношением ее максимального и минимального сингулярных чисел из разложения вида (1.2):

$$cond(C) = \frac{\lambda_{\max}}{\lambda_{\min}}, \quad (1.54)$$

и чем оно больше, тем ближе матрица к вырожденной и хуже обусловлена задача разложения.

Так, авторы [22,24] отмечают, что метод Аленцева-Фока позволяет раскладывать смешанные спектры в среднем не более, чем 3-4 компонентов, метод Self-Modeling [21] - 2-3 компонентов. Техника минимизации целевой функции (1.50) позволила нам сравнительно легко решать задачи разложения 5-8 компонентных смесей различной природы. По-видимому, реальной границей следует считать 8-10 компонентов при  $cond(D) = 50-100$  и  $cond(C) = 10 - 50$ .

Анализ устойчивости решения, то есть оценивание стабильности контуров спектров компонентов, предлагается проводить путем сравнения решений, полученных при варьировании числа искомым компонентов, числа используемых при их конструировании сингулярных векторов, а также сравнение решений, полученных по неполным наборам экспериментальных данных. Последний способ основан на построении всех возможных комбинаций из  $M_i$  спектров

смесей из  $M$  полученных в эксперименте,  $K \leq M_t < M$ , и попарном сравнении спектров между получаемыми решениями. Критериями сравнения могут быть все рассмотренные ранее критерии сходства двух спектров. Данный подход является на практике весьма эффективным, в отличие от методов оценивания, основанных на анализе дисперсии решения. Дело в том, что оценки дисперсии решения основаны на теоремах распространения ошибок, соответствующие критерии есть оценки сверху, что приводит к получению, с точки зрения практика, бессмысленно больших величин. В следующем разделе это свойство оценок будет продемонстрировано на примере решения тривиальной задачи расчета спектров компонентов при известной матрице концентраций  $C$ .

## 1.5 Использование сведений об относительных концентрациях компонентов в смесях

Известные априори величины относительных концентраций компонентов в анализируемых смесях сводят задачу анализа к простому решению системы линейных уравнений при  $K = M$  или методом наименьших квадратов при  $K < M$ :

$$D = C \cdot X, \quad C^T \cdot D = C^T \cdot C \cdot X, \quad (1.55)$$

где  $D$  –  $M \times N$  матрица спектров смесей,  $X$  –  $K \times N$  матрица искоемых спектров компонентов,  $C$  –  $M \times K$  матрица известных относительных концентраций.

Обратим внимание, что матрицы в (1.55) теперь будут рассматриваться в транспонированном виде, так как решение относится ко всем компонентам в одной спектральной точке и спектры будут представлены не столбцами матриц  $D$  и  $X$ , а их строками.

Относительные концентрации могут быть получены, например, в результате анализа известных физических процессов, приводящих к изменению вкладов компонентов в смеси. При этом получают оценки элементов  $C_{jk}$  в виде параметрической модели, связывающей оценки концентраций с некоторыми известными экспериментальными параметрами, такими, например, как время, начальные концентрации, рН среды, температура и т.п., при варьировании которых можно получать наборы родственных смесей.

### 1.5.1 Литературный обзор

Существует ряд работ по анализу смесей, в которых концентрации компонентов изменяются в зависимости от рН или температуры среды по известным аналитическим выражениям, содержащим, однако, неизвестные параметры, характеризующие химическое равновесие, например, [106]. В этой работе уравнения для определения концентраций 2-компонентных смесей были

$$c_1 = \frac{1}{1 + 10^{k-pH}} \quad (1.56)$$

$$c_2 = 1 - c_1,$$

для 3-х значений рН, соответствующих 3-м смешанным спектрам. Константу равновесия  $k$  здесь находят из условий однозначности решения системы (1.55), т.е. равенства ранга расширенной матрицы рангу невырожденной матрицы коэффициентов. В [107] нахождение константы  $k$  совмещено с поиском относительных концентраций и самих спектров компонентов, что значительно усложняет численную процедуру разложения и, как отмечено в работе, не всегда приводит к успеху.

В [58] предложено проводить разложение спектров флуоресценции двухкомпонентных смесей органических соединений (антраценового ряда) путем моделирования матрицы относительных концентраций в уравнении (1.55) так, что

$$C_{jk} = C_{jk}^0 \cdot e^{\left(\frac{j \cdot \Delta t}{\tau_k}\right)}, \quad (1.57)$$

где  $\Delta t$  - интервал времени, прошедшей с момента начала эксперимента при начальной концентрации  $k$ -го компонента в  $j$ -ой смеси  $C_{jk}^0$  и  $\tau_k$  - время высвечивания компонента. Параметры  $C_{jk}^0$  и  $\tau_k$  варьируют с помощью итерационной процедуры минимизации так, чтобы вычисляемые по уравнению (1.55) спектры компонентов  $X$  наилучшим образом воспроизводили спектры смесей, т.е. минимизировали величину квадратичной невязки

$$T = \sum_{j=1}^M \sum_{i=1}^K \left( D_{ji}^{эксн} - D_{ji} \right)^2 \quad (1.58)$$

где

$$D_{ji} = \sum_{k=1}^K C_{jk} \cdot X_{ki}. \quad (1.59)$$

Индекс  $i$  есть номер точки в спектре. Если при исследовании равновесных смесей величины констант равновесия не определены, их также включают в число неизвестных. При этом, однако, решение задачи разложения становится неоднозначным [108,109].

В процессе физико-химического разделения смесей органических соединений часто образуются фракции одинакового качественного, но различного количественного состава. При этом вещества оказываются достаточно для получения молекулярных спектров каждой фракции. В то же время по аналитическим хроматограммам фракций можно быстро оценить относительные концентрации составляющих их компонентов, когда в силу ряда технических причин невозможно получить индивидуальные спектры. Этот способ оценки матрицы относительных концентраций предложен А. И. Гончаровым и И. П. Петровым в (Гончаров, А.И., Волков В.В., Петров И.П., Гречушников Б.Н., Калинкина И.Н. Метод варьирования концентраций компонентов в анализе сложных смесей органических соединений / А.И. Гончаров, В.В. Волков, И.П. Петров, Б.Н. Гречушников, И.Н. Калинкина // Журнал аналитической химии. - 1982. - Т. 37. - № 3. - С. 470-473.)

В случае ошибок в исходной информации, система уравнений (1.55) является несовместной, при этом ошибки, приводящие к нарушению аддитивности содержатся как в правых частях, так и в  $C$ . Наиболее распространенный метод решения таким систем - метод наименьших квадратов (**МНК**) [60,110-113], который, вообще говоря, когда возмущенной оказывается матрица коэффициентов, дает не только смещенные, но и несостоятельные оценки [114,115], минимизирующие функционал невязки (1.58). Задача (1.55)

является задачей пассивного эксперимента [114], в то время как МНК предполагает случай активного эксперимента (элементы  $C$  полагают известными точно и вся экспериментальная ошибка сосредоточена в правых частях (1.55), т.е. в матрице спектров смесей).

В свою очередь, для решения плохо обусловленных задач используют различные регуляризирующие алгоритмы [116-119]. Методы построения таких алгоритмов для случая пассивного эксперимента рассмотрены в [112, 114]. В [112] предложена общая тактика решения плохо обусловленных систем линейных алгебраических уравнений в условиях возмущения исходных данных. Несколько измененная в нашей работе, она заключается в следующих шагах. На первом этапе получают оценки евклидовых норм экспериментальных ошибок в матрице коэффициентов:

$$\varepsilon_C = \|C - \tilde{C}\|_E,$$

где оценка разности берется между теоретически точной и экспериментальной матрицами коэффициентов, исходя из априорной информации о задаче. Затем вычисляют сингулярное разложение

$$\tilde{C} = U \cdot \Lambda \cdot V^T$$

и анализируют величину сингулярных чисел (диагональных элементов  $\Lambda$ ). Если выполняются оба предложенных в [112] неравенства вида

$$\begin{aligned} \|\Lambda^+\|_2 \cdot \|\Lambda\|_E \cdot (\varepsilon_C + 2.9 \cdot (M + K) \cdot p^{-t+1}) &\leq 0.1, \\ \|\Lambda^+\|_E \cdot \|\Lambda\|_E \cdot (\varepsilon_C + 2.9 \cdot (M + K) \cdot p^{-t+1}) &\leq 0.1, \end{aligned} \quad (1.60)$$

где  $\Lambda^+$  - псевдообратная матрица,  $\|\Lambda^+\|_2 = \min \lambda_k$ ,  $\lambda_k$  - сингулярные числа, а слагаемое  $\varepsilon_C + 2.9 \cdot (M + K) \cdot p^{-t+1}$  отражает машинную точность вычислений ( $p$  - основание системы счисления,  $t$  - число значащих цифр в машинном представлении действительных чисел,  $M$  - число строк,  $K$  - число столбцов матрицы), то считают, что матрица концентраций имеет полный ранг и задачу (1.55) параметризуют, т.е. используя дополнительную информацию о точной

задаче (например, величину оценки  $\varepsilon$ ), решая систему (1.55) методами регуляризации, применяя алгоритмы поиска неотрицательных решений.

Основной проблемой в задаче разложения при известных оценках концентраций, как и при поиске компонентов другими методами, остается оценивание надежности решения. Под надежным решением мы будем понимать набор спектров компонентов, для каждого из которых получены индивидуальные оценки разброса. Данному аспекту, как показал анализ литературы, не уделяется достаточного внимания, хотя с точки зрения исследования систем желательно иметь не столько интегральную оценку дисперсии решения (которая, как правило настолько завышена, что не имеет физического смысла), но и индивидуальные оценки устойчивости для каждого из контуров спектров компонентов. Интегральную оценку лучше всего дает число обусловленности матрицы коэффициентов (1.54), устанавливающее верхнюю границу возможного роста нормы ошибки  $\delta X$  в решении по сравнению с нормами ошибок в исходных данных  $\delta C$  и  $\delta D$  [112]:

$$\frac{\|\delta X\|}{\|X\|} \leq \text{cond}(C) \cdot \frac{\|\delta C\|}{\|C\|}, \quad \frac{\|\delta X\|}{\|X\|} \leq \text{cond}(C) \cdot \frac{\|\delta D\|}{\|D\|}, \quad (1.61)$$

### 1.5.2 Тактика поиска спектров компонентов с анализом устойчивости и надежности решения

Рассмотрим теперь разработанную нами процедуру нахождения спектров компонентов с использованием оценок их относительных концентраций. В нашей работе для решения (53) использован метод наименьших квадратов по нескольким соображениям. Численные эксперименты показывают, во-первых, что при решении плохо обусловленных задач методы пассивного эксперимента не позволяют получать решения с меньшей дисперсией ошибок, чем МНК. Во-вторых, величина смещения решения в МНК при встречающихся на практике уровнях экспериментальных ошибок оказывается меньше его дисперсии. В-третьих, методы пассивного эксперимента позволяют получить малую величину смещения только при значительном переопределении системы (34), что не всегда выполнимо. В-четвертых, эти алгоритмы, как правило, итерационные и требуют дополнительных усилий по обеспечению их сходимости в окрестности решения.

Как показала наша практика решения модельных задач, методы параметризации позволяют получить более близкие к истинным спектры компонентов по сравнению с решением МНК только в случае очень плохо обусловленных задач ( $cond(C) > 500$  при уровне ошибок в матрице  $\delta C$  порядка 1-3 отн.%). К сожалению, и в этом случае параметризованные решения плохо воспроизводят истинные спектры компонентов. Если  $cond(C)$  не превышает 150-200, то параметризация, хотя и позволяет уменьшить норму ошибки в решении для отдельной спектральной точки, не обеспечивает меньшей величины относительной ошибки между точками в спектре, которая и ответственна за искажение спектрального контура. Так как вид спектральных кривых в большинстве случаев трудно выразить аналитически, то довольно сложно связать отдельные точки в индивидуальных спектрах между собой для формирования целевой функции при минимальной ошибке в спектральных образах. На Рисунке 1.22а,б показано решение задачи для второй модельной задачи, рассмотренной в предыдущих разделах с концентрационной матрицей, определенной по данным

аналитической хроматографии, имеющей число обусловленности  $cond(C) = 512$ , сделанное методами НК и регуляризации. Параметризованное решение, соответствующее уравнению

$$\tilde{D} = (\tilde{C} + \alpha E)\tilde{X} \quad (1.62)$$

где  $\alpha = \frac{\varepsilon C}{\|\tilde{C}\|} = 0.03$  – параметр регуляризации при условии  $\|\tilde{C}\| = 1.0$ ,  $E$  – единичная матрица. Такая регуляризация искусственно уменьшает число обусловленности матрицы коэффициентов и тем самым стабилизирует решение, увеличивая его смещение, но уменьшая дисперсию решения согласно оценкам (1.61).

Оказалось, что лучшие результаты для задачи (1.55) с матрицами любой обусловленности можно получить, решая ее методом взвешенных наименьших квадратов

$$(\tilde{C} \cdot W)^T \cdot \tilde{D} \cdot W = (\tilde{C} \cdot W)^T \cdot \tilde{C} \cdot W \cdot \tilde{X}, \quad (1.63)$$

в котором диагональные элементы матрицы весов  $w_j$  для условных уравнений  $j = 1, \dots, M$  назначают равными обратным величинам суммарных квадратичных невязок для каждого из уравнений после получения первичного решения  $\tilde{X}$  одним из рассмотренных выше методов:

$$w_j = \left[ \sum_{k=1}^K \sum_{i=1}^L (\tilde{X}_{ki} \cdot \tilde{C}_{jk} - \tilde{D}_{ji})^2 \right]^{-1}. \quad (1.64)$$

Из величин  $w_j$ , отмасштабированных по максимальному значению к 1, образуют диагональную матрицу весов и решают новую систему уравнений (1.63) также методом НК. Этот способ дает на практике тем лучшие результаты, чем сильнее переопределена матрица  $C$ . При этом, если делают независимые измерения спектров смесей, их лучше не объединять, а вводить в систему (1.63) независимо друг от друга. К сожалению, вследствие большого количества источников ошибок в исходной информации, часто систематических, трудно статистически обосновать конкретный способ вычисления весов  $W$ , в том числе и конкретный вид формулы (1.64).

Каким бы методом не решать задачу разложения, получаемые спектры компонентов в случае плохой обусловленности матрицы  $C$  довольно сильно искажены. Поэтому нами были разработаны способы оценки надежности получаемого решения [A2].

Число обусловленности  $cond(C)$  является мерой максимального роста нормы ошибки в решении по сравнению с нормой ошибки в исходной информации только при малых возмущениях, когда матрица ошибок  $C_e$  в представлении

$$\tilde{C} = C + C_e \quad (1.65)$$

имеет настолько малую норму, что при всех ее реализациях соответствующая матрица  $\tilde{C}$  остается полного ранга. В реальных задачах  $C_e$  имеет, как правило, настолько большие элементы, что  $\tilde{C}$  может менять свой ранг и тогда использовать  $cond(\tilde{C})$  как меру неустойчивости (или ненадежности) решения нельзя, т.к. это число в значительной мере уже случайное [112]. Однако, практика показывает, что  $cond(\tilde{C})$  все-таки может служить эмпирическим критерием общей надежности (или изменчивости) решения, т.к. степень различия между рассчитанными и истинными спектрами компонентов возрастает с ростом  $cond(C)$ . Разумеется, этот критерий применим только в том случае, если ошибки все еще достаточно малы, так как с ростом амплитуды ошибок  $C_e$  величина  $cond(\tilde{C})$  уменьшается из-за некоррелированного.

Кроме общей оценки изменчивости решения в условиях возмущения исходных данных, можно оценивать относительную изменчивость каждого спектра компонента в отдельности. Это особенно важно, т.к. для основных и малоинтенсивных компонентов соответствующие спектры-решения искажены в разной степени. Нами предложены две процедуры получения таких оценок, основанные на сравнении решений-вариантов с решением, полученным для исходной задачи (1.55) (в дальнейшем - опорным решением). Будем называть вариантами либо решения систем линейных уравнений, получаемых из исходной путем отбрасывания некоторых условных уравнений (*первая процедура*), либо

параметризованные решения полной системы (1.55), соответствующие различным значениям параметра  $\alpha$  из соотношения (1.62) (*вторая процедура*). При плохой обусловленности задачи (1.55) любые экспериментальные ошибки приводят к значительным искажениям спектров-решений. Смоделировать все эти ошибки достаточно сложно из-за того, что часто неизвестны законы распределения всех составляющих случайных ошибок измерений и вклады систематических погрешностей.

В этом случае обычно применяемый способ оценки ошибок в решении методом моделирования псевдослучайных ошибок в исходных данных и сравнения возмущенных решений-вариантов не только неприменим, но и необоснован, так как позволяет оценивать устойчивость решения не точной задачи, а уже искаженной, которая в данном случае необоснованно считается точной. В нашей работе мы предложили способ формирования вариантов путем отбрасывания части исходных уравнений. Это позволяет варьировать все виды реализованных в эксперименте ошибок: шумы детектора спектрометра, дрейф прибора, нарушения спектральной аддитивности из-за взаимодействия компонентов в смесях, ошибки в определении элементов  $\tilde{C}$ ,  $\tilde{D}$  и т.п. Формирование систем уравнений меньшей переопределенности, чем исходная, приводит, естественно, к некоторому возрастанию ошибок в решении и изменению числа обусловленности концентрационной матрицы варианта  $\tilde{C}_B$ . Опыт решения модельных задач показал, что условие  $cond(\tilde{C}_B) < 2 cond(\tilde{C})$  в большинстве случаев оказывается приемлемым. Параметризация (1.62), как уже отмечалось, не позволяет улучшить отдельные решения по сравнению с МНК, но, во всяком случае, при изменении величины  $\alpha$  спектры компонентов также изменяются. Методика исследования устойчивости элементов решения путем варьирования величины параметра регуляризации была нами успешно применена при решении несовместных переопределенных систем линейных уравнений в задаче расчета колебательных спектров ионов  $\text{H}_5\text{O}_2^+$  и  $\text{H}_3\text{O}_2^-$  в ангармоническом приближении [A10].

Модельные расчеты показали, что спектры-решения, относительно мало изменяющиеся от варианта к варианту, оказываются ближе к истинным спектрам компонентов, чем сильно меняющиеся. Критериями различия (или сходства) спектральных контуров могут служить все критерии, рассмотренные в разделе 1.3.5.

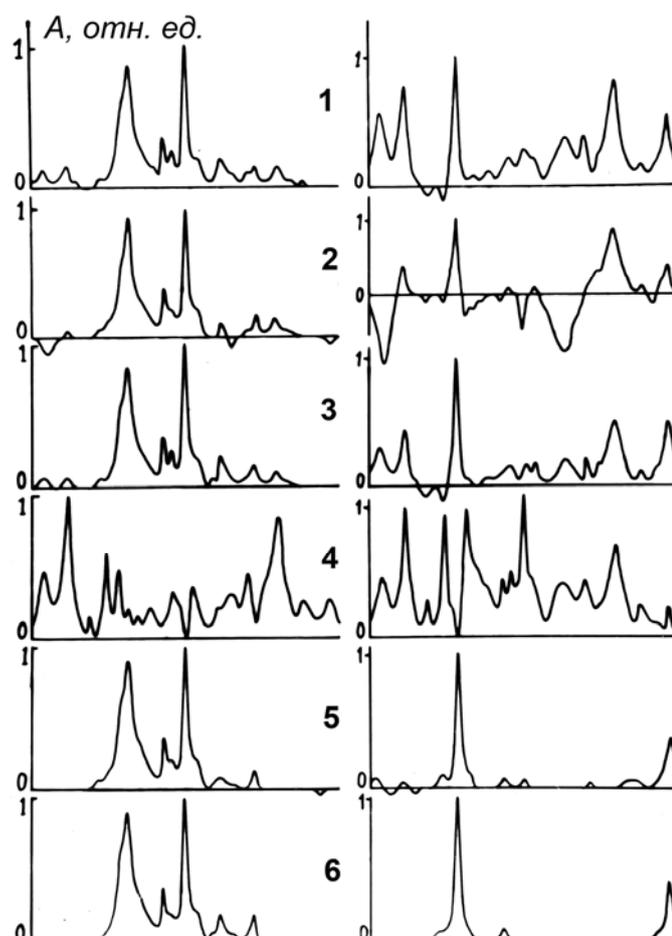


Рисунок 1.22 – Разложение ИК-спектров 6-компонентных смесей при априори известной оценке относительных концентраций. 1: решение МНК, 2: регуляризованное решение, 3: метод взвешенных НК, 4: спектры ошибок, 5: исправленные решения, 6: истинные спектры веществ: а) – изооктана (устойчивый компонент,  $\theta = 12.5^\circ$ ,  $r = 0.97$ ), б)–бензола (неустойчивый компонент,  $\theta = 35.0^\circ$ ,  $r = 0.77$ ). Пояснения см. в тексте.

Чтобы оценить относительную надежность получаемых индивидуальных спектров в отдельности, их ранжируют для каждого из вариантов в порядке возрастания (или убывания) выбранного критерия сходства и вычисляют средние значения рангов по всей совокупности частных решений. Эти средние значения и считают показателями относительной изменчивости спектров-решений задачи разложения, если отдельные ранжировки хорошо согласуются между собой. Мерой согласия  $J$  ранжировок могут служить  $J/2$  коэффициентов ранговой корреляции [65] для всех возможных парных сочетаний ранжировок. Для наших целей лучше подходит коэффициент конкордации  $W_C$  [65], являющийся непараметрическим статистическим критерием сходства.

Пусть получены ранжировки для  $J$  вариантов  $K$ -компонентной смеси, величины рангов равны  $1, 2, \dots, K$ . Чтобы вычислить  $W_C$ , суммируют все  $J$  рангов для каждого спектра компонента, вычитают из каждой суммы среднее значение  $\frac{J \cdot (K + 1)}{2}$  и суммируют квадраты полученных разностей. Коэффициент конкордации  $W_C$  вычисляют как

$$W_C = \frac{12 \cdot (\text{сумма квадратов разностей})}{J^2 \cdot K \cdot (K^2 - 1)}. \quad (1.66)$$

При полном совпадении ранжировок  $W_C = 1.0$ , а чем хуже они согласуются, тем  $W_C$  ближе к 0.

Если предположить, что ранжировки взаимно независимы, то величина  $H = \frac{(J-1) \cdot W_C}{1 - W_C}$  распределена примерно как  $F_{n_1, n_2}$  с  $n_1 = (K-1)$ ,  $n_2 = (K-1)(J-1) - 2$  степенями свободы. Путем сравнения величин  $F$  и  $H$  можно оценить значимость  $W_C$  [65].

В Таблице 1.6 приведены ранжировки для модельной 6-компонентной задачи (Рисунок 1.4). Видно, что ранги компонентов сохраняются примерно одинаковыми от варианта к варианту.  $W_C$  для данного примера оказался равным 0,78;  $H = 39.2$ ;  $F_{n_1, n_2} = 2.4$ , если использовать для сравнения спектров величину

угла между векторами, вычисляемого из (1.35), и  $W_C = 0.83$ ;  $H = 54.5$ ;  $F_{n_1, n_2} = 4.8$  для коэффициента корреляции (1.36). Для этих данных значимость коэффициента конкордации оказалась чрезвычайно высокой, т.е. ранжировки безусловно справедливы и на их основании можно делать выводы о "надежности" найденных контуров спектров. Однако, констатация только того факта, что один из спектров менее надежен, чем другой, все еще недостаточна для практики. Значительно полезнее оказывается более детальная информация о "степени ложности" отдельных полос (или участков) в получаемых спектрах.

Таблица 1.6. Результаты ранжирования спектров-решений вариантов по степени возрастания их изменчивости относительно опорного решения.  $\theta$  обозначает использование в качестве критерия различия величины угла между векторами,  $r$  - коэффициента корреляции.

№ компонента	Ранги для 12 выборочных решений	Средний ранг	Результующий ранг	Критерий сравнения
1	1 1 3 2 3 1 3 1 3 2 3 1	2.0	2	$\theta$ (35)
2	5 5 6 5 6 6 6 5 6 5 6 6	5.8	6	
3	6 6 4 6 4 4 5 6 4 4 4 4	4.8	5	
4	4 4 5 4 5 5 4 3 5 6 5 5	4.6	4	
5	3 3 2 3 2 3 2 4 1 1 1 2	2.3	3	
6	2 2 1 1 1 2 1 2 2 3 2 3	1.8	1	
1	1 1 2 2 2 2 2 1 3 2 3 1	1.8	2	$r$ (36)
2	5 5 6 5 6 5 6 4 6 4 6 6	5.3	5-6	
3	6 6 5 6 5 6 5 6 5 5 5 4	5.3	5-6	
4	3 3 4 4 4 4 4 3 4 6 4 5	4.0	4	
5	4 4 3 3 3 3 3 5 2 3 2 3	3.2	3	
6	2 2 1 1 1 1 1 2 1 1 1 2	1.3	1	

Так как различные способы решения систем линейных уравнений сводятся к построению тех или иных линейных комбинаций правых частей (в нашем случае - спектров смесей), то спектры-решения представляют собой в отсутствие ошибок в матрице  $\tilde{D}$  линейные комбинации решений невозмущенной задачи, т.е. истинных, но неизвестных точно спектров компонентов. На практике

искажения матрицы  $\tilde{D}$  могут быть значительно меньше ошибок в  $\tilde{C}$ , что позволяет нам условно считать невозмущенной. Тогда техника вычисления вариантов дает возможность не только ранжировать получаемые спектры компонентов по степени их изменчивости (ненадежности), но и оценивать относительную изменчивость отдельных деталей их контуров.

Предлагаемый нами метод основан на накапливании в виде "спектров" абсолютных значений разностей между спектром компонента опорного решения (т.е. решения, полученного по полному набору данных) и спектрами из соответствующих решений-вариантов, нормированных перед вычитанием по одному из пиков, который считают присущим спектру данного компонента. Обычно в качестве такого пика берут наиболее интенсивную полосу в опорном решении (если нет других предположений), максимально неперекрываемую, к тому же, с полосами других компонентов. Для поиска характеристических полос можно использовать также методы факторного анализа [1, 42] или построение отношений спектральных контуров [39] и отбирать спектральные точки по максимальному соответствию строк матрицы  $\tilde{D}$  строкам  $\tilde{C}$  (для каждого компонента).

Накопленные абсолютные значения разностей можно рассматривать как своего рода "спектры ошибок" для каждого из компонентов. В местах расположения характеристических полос в спектре ошибок будут небольшие амплитуды, в местах ложных полос - выбросы. На Рисунке 1.22(4) приведены спектры ошибок для рассматриваемой нами в качестве примера модельной 6-компонентной задачи.

Полученные данные об относительной надежности отдельных участков спектров компонентов можно использовать в процедуре "исправления" спектральных контуров. Разработанная нами процедура основана на предположении о малости ошибок спектроскопических измерений - тогда спектры смесей можно считать достаточно точными линейными комбинациями спектров компонентов. На Рисунке 1.22(5) результат представлен результат такой коррекции для двух из 6 компонентов в смеси. Процедура исправления состоит в

исключении тех спектральных участков, которым соответствуют большие относительные амплитуды в спектрах ошибок (таким участкам приписывают нулевые или достаточно малые интенсивности) с последующим проектированием исправленных векторов-спектров  $\mathbf{t}$  на подпространство спектров смесей (т.е. на пространство столбцов  $\tilde{D}$ , Рисунок 1.1) с помощью преобразования проектирования (1.45), рассмотренного в разделе 1.4.2. Полученная проекция  $\mathbf{p}$  есть линейная комбинация векторов-спектров смесей и, следовательно, удовлетворяет условию аддитивности, т.е. формально может считаться спектром компонента. При этом спектр-проекция максимально близок к исходному вектору (в евклидовой метрике), т.е. гарантируется минимальность возмущения спектра-решения при таком преобразовании. Последнее, в свою очередь, означает, что все сделанные исправления будут учтены в максимально полно. Контролировать правильность исключения ложных участков можно по величине угла или коэффициента корреляции между векторами  $\mathbf{t}$  и  $\mathbf{p}$ . Если проекция достаточно сильно отличается от испытуемого спектра (на практике в качестве границы обычно можно брать  $\theta > 15^\circ$  или  $r < 0,95$ ), то это означает, что среди исключенных полос в спектре-решении оказались и присущие данному компоненту. Такие полосы можно узнать из визуального сравнения контуров  $\mathbf{t}$  и  $\mathbf{p}$ , т.к. исключенные "правильные" полосы вновь возникают в спектре проекции. Кроме того, большое различие между  $\mathbf{t}$  и  $\mathbf{p}$  может свидетельствовать о неполном исключении ложных полос. Этот случай наиболее сложен, так как тогда часть исключаемых действительно ложных полос может не полностью исчезнуть в спектре проекции. Но при повторном исключении они, в отличие от "своих" полос, обычно исчезают почти полностью вместе с теми полосами, которые на первом этапе не были идентифицированы как ложные.

Вопрос о достижимой степени близости исправленных и истинных спектров компонентов непосредственно связан с условиями, аналогичными условиям спектральной неперекрываемости Аленцева-Фока (речь о которых шла в разделах 1.4.1 и 1.4.2). В случае строгого выполнения этих условий и при

отсутствии ошибок в матрице  $\tilde{D}$ , можно найти спектры компонентов, совпадающие с истинными [A1].

Наибольший интерес представляет случай, когда упомянутые условия не выполнены, т.е. спектры компонентов не содержат участков, в которых они неперекрываются. Тогда при использовании рассмотренной тактики исключения ложных полос можно найти более "простые", чем на самом деле, спектры компонентов. Однако, в большинстве случаев выполняются условия, напоминающие условия неперекрываемости Аленцева-Фока: в индивидуальных спектрах имеются участки с несовпадающими по положению максимумов полосами, принадлежащими разным компонентам. При этом тактика исключения ложной полосы состоит в сглаживании результирующего спектрального контура в окрестности ее максимума. Часто оказывается, что форма спектральных полос в спектрах компонентов заранее точно не известна. В этом случае не удается построить какие-либо однозначные критерии для исключения ложных пиков, перекрытых со своими соседями, но не совпадающих в максимумах. Этим и обуславливается строгая неоднозначность решения задачи в общем случае и сложность формализации алгоритма.

Дополнительными критериями допустимости исправления решений являются неотрицательность элементов соответствующей концентрационной матрицы  $C$ , соответствие формы кривых распределения концентраций компонентов между смесями данному методу получения этих фракций, а также непревышение величины элементов разностной матрицы  $D^{\#}$  уровня экспериментальных ошибок в  $\tilde{D}$ . Преобразование проектирования позволяет осуществлять входной контроль исходной информации на отсутствие больших ошибок, проявляющихся в несоответствии левых и правых частей системы (1.55). Общий подход состоит в проектировании векторов более искаженной матрицы на соответствующее пространство менее искаженной. Так, проектирование на позволяет исправить спектральные точки с выпадающе большими нарушениями аддитивности.

Как показала практика проектирования, оно не позволяет сколько-нибудь

существенно исправить матрицу концентраций, но по соответствию столбцов матрицы  $\tilde{C}$  столбцам  $\tilde{D}$  можно находить характеристические участки в спектрах компонентов (или даже окна прозрачности). В некоторых случаях матрицу можно исправить, корректируя кривые распределения компонентов по смесям-фракциям, если имеются какие-либо сведения о форме этих контуров (монотонность, аналитические выражения для них и т.п.).

Резюмируя материал первого раздела, можно сказать следующее.

1. Задача поиска спектров компонентов по наборам аддитивных спектров родственных смесей математически плохо обусловлена и при отсутствии априорной информации о составе смесей неоднозначна.

2. Даже в случае выполнения условий неперекрываемости индивидуальных спектров решение зависит от выбора значений параметров целевой функции.

3. В этих условиях неприменимы обычные статистические оценки дисперсии решений, основанные на использовании оценок дисперсии исходных данных, так как ошибки включают и неизвестные априори систематические отклонения, нарушающие аддитивность.

4. В таком случае, оценку надежности решения предлагается проводить путем сравнительного анализа контуров спектров компонентов, полученных при варьировании:

- стартового приближения,
- параметров целевой функции (как набора, так и значений),
- набора исходных данных (например, получая решения по неполным наборам спектров, перебирая все комбинации).

5. В качестве критериев оценки качества (надежности, стабильности) решения можно использовать величину дисперсии коэффициентов корреляции (или других критериев попарного сходства или различия) между спектрами компонентов; чисел обусловленности матриц компонентов. Можно применять статистические оценки автокорреляции в элементах векторов, представляющих

собой разности между соответствующими компонентами в наборах решений.

Невозможно предложить абсолютно универсальный алгоритм разложения и, как отмечено в разделе 1.1 и в работе [49], выбор наилучшего решения остается предметом экспертной оценки.

Разработанные алгоритмы и программы анализа аддитивных спектров смесей были адаптированы для более общих случаев и применены для исследования многокомпонентных данных малоуглового рентгеновского и нейтронного рассеяния, что рассмотрено во второй части работы.

## Часть 2

### **Определение структурных параметров наноразмерных объектов по данным малоуглового рентгеновского и нейтронного рассеяния**

Необходимость изучения связей "физико-химические свойства - состав - структура" предполагает применение, в числе прочих, методов исследований, дающих информацию о пространственном строении объектов на нанодатомном уровне. Эти эксперименты часто оказываются довольно сложными из-за невозможности выделения отдельных компонентов таких "наносистем" в изолированном виде и невозможности применять специальную подготовку образцов. В таких случаях приходится получать и анализировать информацию "*in vivo*", и в этом смысле изучаемые объекты являются многокомпонентными смесями. В данной работе мы ограничимся одним их высокоэффективных структурных методов исследования - малоугловым рассеянием рентгеновских лучей и нейтронов (МУР), возможности которого неизмеримо выросли в последние годы благодаря появлению ярких лабораторных и синхротронных источников излучения, станций на непрерывных и импульсных нейтронных реакторах. Широкое применение этих источников, в совокупности с новыми методиками анализа данных рассеяния и построения структурных моделей, сделали малоугловое рассеяние одним из самых эффективных аналитических методов анализа наноразмерных структур. Серьезное преимущество МУР заключается в возможности исследования объектов без какой-либо специальной подготовки, что зачастую оказывается определяющим фактором при изучении растворов, гелей и других аналогичных систем. В работе рассмотрены две области нанодиагностики, где прогресс в малоугловом эксперименте и новейшие методики интерпретации данных рассеяния особенно рельефно проявились в последние годы. Эти направления - анализ строения биологических макромолекул в растворе и структурные исследования полимерных и композитных систем с наночастицами и кластерами различной природы. Развитые в работе методы интерпретации данных рассеяния будут продемонстрированы примерами практических биологических и нанотехнологических приложений. Определение размерных параметров частиц предполагает нахождение их формфакторов, поэтому методам поиска функции формы будет отведено значительное место в рассмотрении.

## 2.1 Литературный обзор

В структурных исследованиях применяются методы рентгеновской дифракции и рассеяния, основанные на эффектах упругого взаимодействия рентгеновских лучей со связанными электронами в образце. Для структурного анализа используется достаточно жесткое излучение с длиной волны  $\lambda$  порядка  $\sim 0.1$  нм, что близко к межатомным расстояниям. Для многих кристаллических объектов методы рентгеновской кристаллографии позволяют определять их структуру в виде трехмерного распределения электронной плотности  $\rho(\mathbf{r})$ . Дифракция на кристаллических объектах дает трехмерную картину рентгеновской интенсивности  $I(\mathbf{s})$ , состоящую из набора рефлексов вдоль определенных направлений вектора рассеяния  $\mathbf{s}$ . Вектор рассеяния задает координаты точки измерения интенсивности и представляет собой разность между волновыми векторами рассеянной  $\mathbf{k}_s$  и исходной  $\mathbf{k}_0$  волн, а его длина (модуль) используется в качестве угловой меры:

$$\mathbf{s} = \mathbf{k}_s - \mathbf{k}_0, \quad |\mathbf{k}_s| = |\mathbf{k}_0| = \frac{2\pi}{\lambda}, \quad \text{то } |\mathbf{s}| = s = 2|\mathbf{k}_0| \sin \theta = \frac{4\pi \sin \theta}{\lambda}, \quad (2.1)$$

где  $\lambda$  – длина волны,  $2\theta$  - угол рассеяния.

Рефлексы от упорядоченной структуры можно измерять до больших углов рассеяния, что собственно и дает возможность восстановления трехмерной структуры при атомном разрешении после решения фазовой проблемы [120], рассмотрению которой посвящено большое число работ. Разупорядоченные объекты рефлексов, как правило, не дают, и измеряемая картина рассеяния – это гладкая изотропная функция  $I(s)$ . В 1938 г. французский ученый А. Гинье обнаружил, что центральная часть дифракционной картины содержит довольно сильное рассеяние при наличии в веществе высокодисперсных зерен с размерами в десятки нанометров [121]. Эти работы и положили начало методу малоуглового рентгеновского рассеяния (**МУР**), который позволяет исследовать вещества, содержащие неоднородности с размерами в диапазоне  $1 - 10^3$  нм. Чем больше размер рассеивающего объекта, тем в меньшем угловом интервале сосредоточено

рассеянное излучение (номинальное разрешение  $d$  определяется соотношением  $d = 2\pi/s$ ). Соответственно, рассеяние на малые углы (меньше нескольких градусов) несет информацию о "крупномасштабных" (по отношению к длине волны излучения  $\lambda$ ) рассеивающих объектах, и МУР дает структурную информацию с разрешением до 1-2 нм. Для исследования надатомной структуры оказывается достаточным регистрировать дифракционную картину в области малых углов рассеяния, откуда и получил название сам метод. Зачастую его применение – единственный способ получения прямой структурной информации о системах с хаотическим распределением неоднородностей плотности.

Измерение интенсивности рентгеновского МУР представляет собой сложную техническую задачу, поскольку приходится регистрировать слабое рассеяние вблизи мощного первичного пучка излучения. В малоугловых дифрактометрах использование двумерного детектора не обязательно, так как регистрируется только радиальная картина интенсивности рассеяния, и поэтому во многих установках применяют линейные позиционно-чувствительные детекторы. Общая схема установки МУР показана на Рисунке 2.1. Основное ее отличие от дифрактометров, предназначенных для исследования кристаллических и поликристаллических образцов часто заключается в специальной схеме коллиматора, позволяющего начинать измерения с очень малых углов  $0.1^\circ - 0.03^\circ$ . Это позволяют делать так называемые блочные коллиматоры, самым известным из которых является коллиматор Кратки. Их конструкции подробно рассмотрены в литературе [122].

Применение МУР к исследованию биологических объектов начало интенсивно развиваться в 60-70-е гг. Размеры этих объектов лежат в области от нескольких нанометров до нескольких сотен нанометров и растворы таких биополимеров близки к идеальной системе идентичных наночастиц. В это время на модельных расчетах удалось показать, что интенсивность малоуглового рассеяния сильно зависит не только от размера частицы и ее формы, но и от внутренней структуры наночастицы.

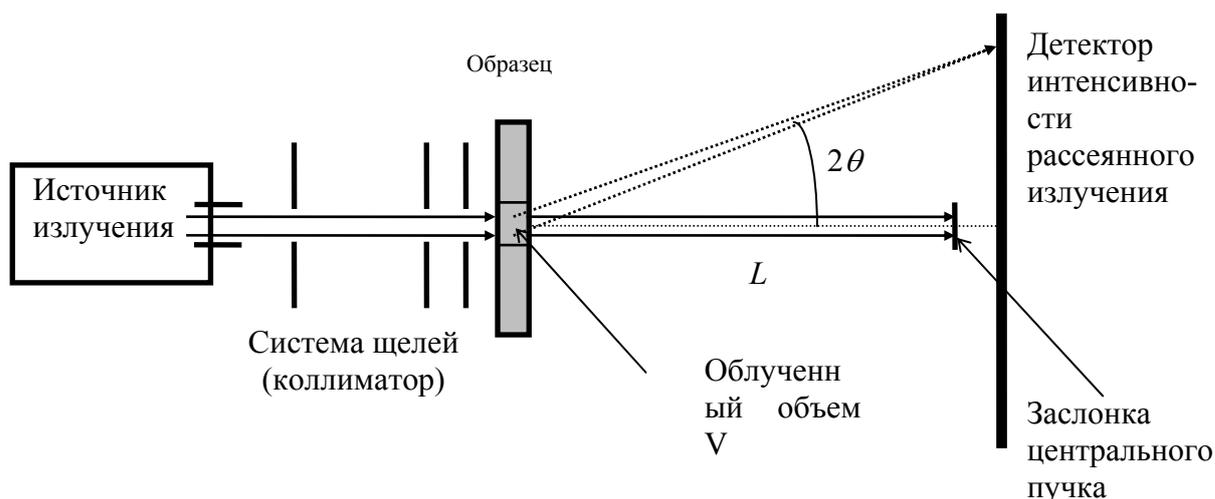


Рисунок 2.1 – Общая схема установки по измерению интенсивности рассеянного излучения.

По данным малоуглового рассеяния были построены структурные модели ряда белков, рибосом и нескольких бактериальных вирусов – бактериофагов. Для последних объектов удалось определить форму белковой оболочки бактериофагов Сд и Т7, размер отростка и особенности размещения РНК внутри вируса. Для осуществления этих исследований были разработаны и построены первые автоматические рентгеновские малоугловые дифрактометры и разработаны программы обработки данных малоуглового рассеяния и построения моделей строения различных наночастиц. Высокая информативность определения структуры бактериофага Т7 [123,124] спустя 25 лет была подтверждена при использовании новейшего низкотемпературного электронно-микроскопического метода получения трехмерных изображений [125].

В 1986-1987 гг. в СССР и США вышла в свет монография Л.А. Фейгина и Д.И. Свергуна по рентгеновскому и нейтронному малоугловому рассеянию, которая до сего времени широко используется исследователями в данной области [126]. Более современный короткий обзор возможностей метода рассмотрен в [А58].

Важнейшим фактором развития МУР стали также новые методики анализа данных рассеяния и поиска структурных моделей. Экспериментальные данные МУР позволяют напрямую определять общие характеристические параметры исследуемых систем, например средний размер частиц, их массу, анизотрию.

Построение трехмерных моделей рассеивающих объектов представляет собой сложнейшую задачу из-за ограниченного количества информации, содержащейся в данных рассеяния. В сравнении с рентгенограммами монокристаллов, где измеряются трехмерные наборы рефлексов, МУР дает картины рассеяния, усредненные по ориентациям наночастиц, а для полидисперсных систем и по размерам рассеивающих объектов. Решение обратной задачи рассеяния, т.е. однозначное восстановление по экспериментальным данным трехмерной структуры объекта, в общем случае невозможно. Тем не менее в ряде работ 1970-80-х гг. (обзор которых дан в [126]) было показано, что при определенных ограничениях из данных рассеяния удастся определять приближенные трехмерные модели. Основной прогресс был достигнут в конце 1990-х гг., когда появились новые подходы к интерпретации данных МУР изотропными монодисперсными системами [126]. Эти подходы, реализованные в виде компьютерных программ, доступных всему научному сообществу, получили широкое применение для анализа данных МУР, сначала от биологических систем, а затем и от сложных полидисперсных органических, неорганических и композитных систем.

В большинстве практически важных приложений, рассеивающий объект можно представить как совокупность частиц, внедренных в матрицу, причем рассеивающая способность частиц отличается от матрицы (частицы имеют контраст по отношению к матрице). Рассеивающая способность частиц для рентгеновских лучей описывается распределением электронной плотности  $\rho(\mathbf{r})$ , где  $\mathbf{r}$  – координатный радиус-вектор. В предположении, что матрица имеет постоянную электронную плотность  $\rho_s$ , эффективная рассеивающая способность частицы дается разницей  $\Delta\rho(\mathbf{r})=\rho(\mathbf{r})-\rho_s$ . Амплитуда упругого рассеяния от одной частицы в матрице определяется преобразованиями Фурье

$$A(\mathbf{s}) = \mathfrak{F}[\Delta\rho(\mathbf{r})] = \int_V \Delta\rho(\mathbf{r}) \exp(i\mathbf{s}\mathbf{r}) d\mathbf{r}, \quad (2.2)$$

где вектор рассеяния  $\mathbf{s}$  есть разность между волновыми векторами рассеянной и падающей волн (2.1). В эксперименте измеряется не амплитуда, а интенсивность

рассеяния  $I(s)=A(s)A^*(s)$ , которая пропорциональна числу фотонов, рассеянных в направлении  $s$ .

В малоугловых исследованиях разупорядоченных объектов чрезвычайно важным является случай **монодисперсной системы**, когда все частицы идентичны, но имеют хаотическую ориентацию в образце. Классическим примером такой системы являются высокоочищенные растворы биологических макромолекул, где частицы (белки или их комплексы) производятся по одной и той же генетической программе и имеют одинаковую конформацию. При исследовании разбавленных растворов эффектами межчастичной интерференции можно пренебречь, и кривая рентгеновского МУР после вычета рассеяния матрицей (в данном случае, растворителем) отвечает рассеянию одной частицей, усредненной по всем ориентациям  $I(s)=\langle I(s) \rangle_{\Omega}$ , где  $\Omega$  - телесный угол в обратном пространстве:

$$I(s) = \left\langle A^2(s) \right\rangle_{\Omega} = 4\pi \int_{r=0}^{\infty} r^2 \langle \rho(r) \otimes \rho(r) \rangle_{\omega} \frac{\sin(sr)}{sr} dr, \quad (2.3)$$

$\otimes$  означает операцию свертки,  $\omega$  - телесный угол в прямом пространстве.

По такой кривой изотропного рассеяния удастся без привлечения априорной информации определять некоторые интегральные характеристики исследуемых частиц (так называемые инварианты).

Инвариантами называют структурные характеристики систем, которые могут быть в явном виде выражены через интенсивность МУР. Вычисление инвариантов при правильно проведенном эксперименте не требует никакой априорной информации о частицах в системе. Геометрические характеристики частиц можно определять из кривой рассеяния в относительных единицах, а расчет весовых параметров требует абсолютизации измерений.

Классическим параметром (который был впервые определен еще в основополагающей работе А. Гинье [127]) является радиус инерции частицы  $R_g$ . Этот параметр может быть найден по начальному участку кривой рассеяния (для малых значений  $s$ ) с помощью формул (где  $n=3$  для сферических образований, 2 для радиуса инерции вытянутой частицы относительно оси, 1 для радиуса

инерции плоской фигуры поперечного сечения дискообразных частиц относительно ее длинной оси сечения):

$$I(s) = I(0)e^{-\frac{s^2}{n} \cdot Rg^2}, \quad \ln I(s) = \ln[I(0)] - \frac{s^2}{n} \cdot Rg^2. \quad (2.4)$$

Из тангенса угла наклона прямолинейного участка зависимости  $\ln I(s)$  от  $s^2/n$  (графика Гинье) определяют величину  $Rg$ , которая характеризует средний размер частицы, а также интенсивность рассеяния в нулевой угол  $I(0)$ , которая пропорциональна квадрату молекулярной массы, или произведению квадрата контраста средней электронной или массовой плотности на квадрат объема  $\Delta\rho^2 \cdot V^2$ . Объем  $V$  и площадь поверхности  $S$  вычисляют с использованием нормированного инварианта Порода [122]:

$$Q_0 = \int_0^{\infty} i(s)s^2 ds, \quad i(s) = I(s)/I(0)$$

$$V = \frac{2\pi^2}{Q_0}, \quad \frac{S}{V} = \frac{\pi}{Q_0} \lim[s^4 i(s)], \quad s \rightarrow \infty \quad (2.5)$$

Фурье-преобразование интенсивности рассеяния позволяет определить характеристическую функцию  $\chi(r)$  и функцию распределения по расстояниям (функцию парных расстояний)  $p(r) = r^2 \chi(r)$ :

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

$$p(r) = \frac{r^2}{2\pi^2} \int_0^{\infty} s^2 I(s) \frac{\sin(sr)}{sr} dr. \quad (2.6)$$

Эти функции, помимо информации об анизотрии частицы, позволяют оценивать максимальный размер  $D_{max}$  из условия  $p(r) = 0$  при  $r > D_{max}$ . В практических исследованиях,  $p(r)$  рассчитывают не прямым интегрированием интенсивности (что приводит к сильным эффектам обрыва) а с помощью так называемого косвенного фурье-преобразования [126,128,169,170]. Функции (2.6) позволяют определять радиус инерции частицы с большей, чем по формуле (2.4)

точностью, так как интегрирование ведется не по начальному участку кривой рассеяния, соответствующему (2.4), а по спектру Фурье, рассчитанному по всей доступной из эксперимента области углов рассеяния [121]:

$$R_g^2 = \frac{\int_0^D r^4 \gamma(r) dr}{2 \int_0^D r^2 \gamma(r) dr} = \frac{\int_0^D r^2 p(r) dr}{2 \int_0^D p(r) dr} . \quad (2.7)$$

Сферическое усреднение кривой интенсивности рассеяния, возникающее из-за хаотической ориентации частиц, приводит к значительной потере информации, содержащейся в дифракционных данных. Интерпретация одномерных данных рассеяния с использованием трехмерных моделей – чрезвычайно трудная задача. Традиционные методы анализа кривых МУР изотропными монодисперсными системами основывались на расчете инвариантов, а также включали моделирование путем аппроксимирования формы частицы трехпараметрическими геометрическими телами и комбинацией из нескольких таких тел. Для построения более сложных моделей применяли метод проб и ошибок, с использованием априорной информации о строении частиц полученной другими методами (электронная микроскопия, различные физико-химические измерения) [129-134]. Для сравнения моделей применяли как сравнение с данными интенсивности рассеяния, так и сравнение функций парных расстояний [135]. Как правило, строился ряд предполагаемых моделей исследуемых частиц и сравнение рассеяния на них с экспериментальной кривой. Этот подход иногда позволял подобрать модель, наилучшим образом отвечающую как данным рассеяния, так и априорной информации, но не говорил ничего о единственности решения.

В случае **полидисперсных систем** (т.е. систем неидентичных частиц) можно выделить два наиболее важных случая. В одном из них, объект содержит несколько различных по форме типов невзаимодействующих частиц с произвольной структурой. Интенсивность рассеяния от такой смеси может быть

записана в виде линейной комбинации (если не учитывать межчастичную интерференцию, что с достаточной для практики точностью справедливо в случае разбавленных растворов)

$$I(s) = \sum_{k=1}^K v_k I_k(s), \quad (2.8)$$

где  $v_k > 0$  и  $I_k(s)$  – объемная доля и интенсивность рассеяния от  $k$ -го типа частиц (компонент), соответственно, а  $K$  - число компонент. Ясно, что располагая только экспериментальными данными рассеяния от смеси, нельзя реконструировать структуры отдельных компонентов и параметры, которые могут быть извлечены, зависят от дополнительной информации. Если число компонент и их интенсивности рассеяния известны априори, можно определить объем фракций  $v_k$  в линейной комбинации (6) методом неотрицательных линейных наименьших квадратов [A23]. Такие ситуации часто встречаются при анализе равновесных систем (растворы олигомерных белков, процессы сборки самоорганизующихся объектов, фазовые переходы в микроэмульсиях).

Другой тип полидисперсности характерен для систем, где частицы имеют схожие формы и отличаются только размерами. Такие системы удобно описывать объемной функцией распределения  $D(R) = N(R)V(R)$ , где  $N(R)$  это число частиц с характерным размером  $R$  в системе, а  $V(R)$  - объем одной частицы данного размера. Интенсивность рассеяния в этом случае определяется интегралом (для разбавленных систем)

$$I(s) = (\Delta\rho)^2 \int_{R_{\min}}^{R_{\max}} D(R)V(R)i_0(sR)dR, \quad (2.9)$$

где  $i_0(sR)$  – квадрат «**формфактора**», т.е. нормированная интенсивность рассеяния частицей (при  $i_0(0)=1$ ),  $R_{\min}$  и  $R_{\max}$  - минимальный и максимальный размер частиц соответственно (под формфактором иногда понимают также нормированную амплитуду рассеяния от тела заданной формы). Биологические объекты редко дают такой вид полидисперсности, но зато данное уравнение часто применяется к мицеллам, микроэмульсиям, блок-сополимерам, металлическим

наночастицам и др. В большинстве практических случаев предполагается, что форм-фактор частицы известен (в частности, для изотропной системы частицы обычно можно считать сферическими). Тогда объемную функцию распределения функций  $D(R)$  можно определить с помощью косвенного преобразования аналогично тому, как решается интегральное уравнение (2.6).

В конденсированных системах и концентрированных растворах эффектом межчастичной интерференции пренебрегать нельзя. В этом случае обычно предполагают, что взаимодействия происходят только между частицами, имеющими сферическую форму (для других форм аналитические выражения еще не получены) и интенсивность рассеяния записывают в виде

$$J(s) = I(s) \cdot T(s), \quad (2.10)$$

где  $T(s)$  - структурный фактор, описывающий интерференционные эффекты. Этот множитель, не меняя интегральной интенсивности, определяет форму кривой рассеяния [121,122].

Анализ данных рассеяния от сложных систем предполагает, как уже было отмечено, знание формфакторов. Поэтому их определению в работе было уделено особое внимание.

Далее приведен обзор новых методов анализа данных МУР, развитых в последнее десятилетие. Эти методы позволяют определять трехмерные структуры частиц в монодисперсных (и даже умеренно полидисперсных) системах при разрешении до 1 нм.

Для извлечения информации о трехмерной структуре исследуемых объектов в теории малоуглового рассеяния широко применяется аппарат описания малоугловой дифракции с помощью сферических гармоник.

Основным достоинством сферических гармоник является то, что сферическое усреднение дифракционной картины существенно упрощается при использовании мультипольного разложения. Действительно, рассмотрим рассеяние отдельно взятой частицей, состоящей из  $N$  атомов. Для атомных координат  $r_k = (r_k, \omega_k) = (r_k, \theta_k, \varphi_k)$ ,  $k = 1, 2, \dots, N$  и соответствующих атомных форм-факторов  $f_k(s)$ , трехмерная амплитуда рассеяния:

$$A(\mathbf{s}) = \sum_{k=1}^N f_k(s) \exp(i\mathbf{s}\mathbf{r}_k). \quad (2.11)$$

Функцию  $\exp(i\mathbf{s}\mathbf{r})$  можно представить в виде [122]:

$$\exp(i\mathbf{s}\mathbf{r}) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^l i^l j_l(sr) Y_{lm}^*(\omega) Y_{lm}(\Omega), \quad (2.12)$$

где  $j_l(sr)$  - сферические функции Бесселя, а  $Y_{lm}(\Omega)$  - сферические гармоники.

Таким образом, амплитуда рассеяния записывается в виде

$$A(\mathbf{s}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l A_{lm}(s) Y_{lm}(\Omega), \quad (2.13)$$

где функции  $A_{lm}(s)$  называют парциальными амплитудами:

$$A_{lm}(s) = 4\pi i^l \sum_{k=1}^N f_k(s) j_l(sr_k) Y_{lm}^*(\omega_k). \quad (2.14)$$

При возведении (2.14) в квадрат и сферическом усреднении по  $\Omega$ , благодаря ортогональным свойствам сферических гармоник все перекрестные члены исчезают, что приводит к простому выражению для интенсивности малоуглового рассеяния, с точностью до  $2\pi$ :

$$I_L(s) = \sum_{l=0}^L \sum_{m=-l}^l |A_{lm}(s)|^2, \quad (2.15)$$

где величина обрыва ряда  $L$  (максимальный индекс гармоники) определяет разрешение представления структуры частицы.

Впервые это описание было предложено в [136,137], и в настоящее время оно составляет базис для наиболее эффективных методик структурного анализа данных МУР.

Математический аппарат сферических гармоник оказался весьма эффективным и для параметризации структуры частицы в прямом пространстве. При этом можно параметризовать как распределение электронной плотности внутри частицы, так и ее форму в предположении, что частица однородна.

В первом случае рассеивающую плотность представляют в виде ряда

$$\rho(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \rho_{lm}(r) Y_{lm}(\omega) = \sum_{l=0}^{\infty} \rho_l(\mathbf{r}), \quad (2.16)$$

где  $Y_{lm}(\omega)$  - сферические гармоники,  $\omega$  – телесный угол в прямом пространстве, а  $\rho_l(\mathbf{r})$  – парциальные мультипольные плотности. Функции  $A_{lm}(s)$  и  $\rho_{lm}(r)$  связаны преобразованием Ханкеля:

$$A_{lm}(s) = i^l \sqrt{\frac{2}{\pi}} \int_0^{\infty} \rho_{lm}(r) j_l(sr) r^2 dr, \quad (2.17)$$

$$\rho_{lm}(r) = (-i)^l \sqrt{\frac{2}{\pi}} \int_0^{\infty} A_{lm}(s) j_l(sr) s^2 ds, \quad (2.18)$$

и для расчета интенсивности используется выражение (2.15).

В [138] показан способ расчета распределения электронной плотности в аксиально-симметричных частицах, основанный на представлениях (2.15-2.18) при ограничении ряда (2.16) до некоторого значения  $l = L$  и наложения ограничения на максимальный диаметр структуры, который является одним из инвариантов. В работе [139] показано, что при этих условиях решение обратной задачи единственно в случае сферически симметричных частиц и метод применен к определению структуры бактериофага T7.

Во втором случае, если предполагать частицу однородной, с помощью сферических гармоник можно параметризовать форму ее оболочки  $F(\omega)$ , где каждому угловому направлению  $\omega$  ставится в соответствие расстояние от центра до границы частицы. Распределение плотности внутри частицы

$$\rho(r) = \begin{cases} 1, & 0 \leq r < F(\omega) - \Delta \\ F(\omega) - r / \Delta, & F(\omega) - \Delta < r \leq F(\omega) \\ 0, & r > F(\omega) \end{cases} \quad (2.19)$$

учитывает возможную размытость границ в реальных объектах ( $\Delta$  - ширина переходного слоя частица-матрица). Функция оболочки параметризуется с помощью сферических гармоник

$$F(\omega) = \sum_{l=0}^L \sum_{m=-l}^l f_{lm} Y_{lm}(\omega) \quad (2.20)$$

где  $f_{lm}$  - комплексные числа, а пространственное разрешение характеризуется параметром обрыва ряда  $L$ :  $\delta r \approx \pi R_0 / (L + 1)$ , где  $R_0$  - радиус эквивалентной сферы. Структура частицы представляется таким образом с помощью  $(L+1)^2$  параметров, а для расчета модельной интенсивности рассеяния по формуле (2.15) используют выражение для парциальных амплитуд при заданной границе  $p_{\max}$ :

$$A_{lm}(s) = (is)^l \cdot \left(\frac{2}{\pi}\right)^{1/2} \times \sum_{p=0}^{p_{\max}} \left( (-1)^p \cdot f_{lm}^{l+2p+3} \cdot \left\{ 2^p \cdot p! \cdot (l+2p+3) \cdot [2(l+p)+1]!! \right\}^{-1} \cdot s^{2p} \right), \quad (2.21)$$

$$\text{где } f_{lm}^{(q)} = \int_{\omega} [F(\omega)]^q \cdot Y_{lm}^*(\omega) \cdot d\omega.$$

Использование математического аппарата сферических гармоник позволило создать ряд высокоэффективных алгоритмов и программ моделирования формы и строения низкого разрешения наночастиц в монодисперсных системах. Данные программы в настоящее время широко используются во всем мире [А34,А39].

В конце 90-х гг. был разработаны методы *ab initio* восстановления формы с использованием **модели конечных объемных элементов**. Поиск трехмерной формы частицы осуществляется с использованием эвристических методик глобальной минимизации суммарной квадратичной невязки между экспериментальной кривой рассеяния и теоретической, рассчитываемой от модели структуры, представленной набором плотноупакованных шариков достаточно малого объема. Поиск решения (пространственного размещения шариков в модели) ведут с помощью методов глобальной минимизации, например, основанных на моделировании отжига (simulated annealing) [140] или на генетических алгоритмах [141]. Способ, основанный на методе моделирования отжига оказался исключительно эффективным из-за использования описанного выше аппарата сферических гармоник и ряда алгоритмических решений, в частности, из-за отказа от вариации на каждом шаге поиска всех параметров модели, а варьировании только одного, выбираемого случайным образом. При

этом экспериментальные данные рассеяния приближаются с наложением ограничений (компактность и неразрывность) на возможное решение.

Основная идея моделирования конечными элементами, предложенного в [140], заключается в следующем. Область поиска, (например, сфера с диаметром, равным максимальному размеру частицы  $D_{max}$ ) заполняется  $N$  ( $\sim 10^3$ ) плотноупакованными шариками радиуса  $r_0 \ll D_{max}$ . Каждому шарика приписывается индекс  $X_j$ , обозначающий к какой фазе принадлежит данный объемный элемент ( $X_j=0$  означает растворитель,  $X_j=1$  – частица) При фиксированных положениях шариков форма и структура модели конечных объемных элементов полностью описывается конфигурационным вектором  $X$ , содержащим  $N \approx (D/r_0)^3$  компонент. В методе [140] интенсивность рассеяния такой моделью быстро рассчитывается с помощью приведенного выше мультипольного разложения.

Техника моделирования отжига заключается в старте с произвольной модели  $X_j$  и ее случайных модификаций до тех пор, пока модель не станет приближать данные. В отличие от метода Монте-Карло, в котором в качестве текущей модели используется всегда лучшая из найденных, и пробная вариация производится относительно нее, в методе отжига в качестве текущей модели используют модель, которая не обязательно является лучшей, а используется модель, принятая после предыдущей вариации с некоторой вероятностью, несмотря на то, что она может быть хуже лучшей. Таким образом, в методе Монте-Карло используют две модели - лучшая (которая варьируется на следующем шаге) и пробная, полученная вариацией лучшей модели. Пробная модель отвергается, если она хуже и принимается в качестве лучшей в противном случае, после чего цикл повторяют. В методе моделирования отжига используют три модели - лучшая, текущая и пробная. Текущая модель является опорной и сравнение новой, пробной модели производят относительно нее. Вариацию проводят относительно текущей модели и получают пробную. Вероятность принять пробную модель в качестве новой текущей зависит от некоторого параметра, который называют в литературе "температурой"  $T$ . Большая

температура означает высокую вероятность, вычисляемую как  $\exp(-\Delta/T)$ , принять в качестве новой текущей модели худшую относительно текущей, если  $\Delta$  (величина изменения целевой функции) больше 0 (решение хуже). Но если пробная модель оказывается лучше, т.е.  $\Delta < 0$ , то она всегда принимается в качестве новой текущей. В начале поиска температуру выбирают достаточно высокой, чтобы частота принятия худших решений в качестве текущих превышала частоту обновления лучших моделей. Это заставляет программу "бродить" по области поиска и, в случае попадания модели в окрестности локального минимума, в конце концов выбираться из него. По мере роста числа испытаний температуру снижают. Чаще всего используют экспоненциальный закон,  $T_{\text{next}} = T_{\text{curr}} * F$ ,  $F$  (фактор отжига) = 0.9 - 0.95, а значение целевой функции, общего вида (2.22), уменьшается по аналогии с уменьшением внутренней энергии системы по мере снижения температуры, из-за чего метод и получил название "моделирование отжига" [142].

Поскольку для определения формы используются данные рассеяния, соответствующие низкому разрешению, искомая модель также должна иметь низкое разрешение по отношению к  $r_0$ . Поэтому на модель накладываются также условия связности и компактности, так что целевая функция, минимизируемая в процессе восстановления модели низкого разрешения, имеет вид:  $f(X) = \chi^2 + \alpha P(X)$ . Здесь,  $\chi^2$  это невязка

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^N \left[ \frac{I(s_j) - I_{\text{exp}}(s_j)}{\sigma(s_j)} \right]^2, \quad (2.22)$$

где  $I_{\text{exp}}(s)$  экспериментальная интенсивность, определенная в  $N$  точках  $s_j, j=1, \dots, N$ , и  $\sigma(s_j)$  соответствующие стандартные отклонения, а  $P(X)$  – штраф за некомпактность модели ( $\alpha > 0$  – его вес, задаваемый пользователем).

Этот подход в разных вариантах использован в ряде программ, находящихся в свободном сетевом доступе (например, DALAI\_GA, DAMMIN, DAMMIF, GASBOR и др.) [142, A53].

В настоящее время определение формы частиц по данным МУР – стандартная процедура, и на некоторых синхротронных линиях (например на станции ЕМБЛ Х33 на синхротроне DESY) пользователи получают форму автоматически, в режиме реального времени, спустя ~2 мин. после самого измерения.

Другим чрезвычайно важным подходом для интерпретации данных МУР трехмерными моделями является моделирование жесткими телами, или **метод молекулярной тектоники**. Он получил свое развитие прежде всего в биологических исследованиях, где современные проекты «структурной геномики» позволяют определять тысячи структур индивидуальных макромолекул и их фрагментов с помощью рентгеновского анализа и ЯМР [143-145].

Однако, самые важные клеточные функции белков совершаются не отдельными белками, а макромолекулярными комплексами. Такие комплексы обычно слишком велики для исследования методом ЯМР и, кроме того, они часто обладают внутренней структурной гибкостью, затрудняя тем самым их кристаллизацию. Поэтому необходимы альтернативные экспериментальные методики и подходы при анализе данных для анализа макромолекулярных ансамблей, и МУР предоставляет такие возможности.

Кривые рассеяния от растворов чувствительны к изменениям четвертичной структуры макромолекул. Таким образом, этот метод оказывается чрезвычайно полезным для анализа макромолекулярных комплексов. Подробные модели могут быть построены в случае, когда известны с высоким разрешением структуры отдельных доменов или субъединиц, составляющих комплекс [146].

Суть алгоритма молекулярной тектоники можно проиллюстрировать на примере комплекса, состоящего из двух субъединиц (А и В). Амплитуды рассеяния от субъединиц, расположенных в центре с соответствующей ориентацией, обозначим как  $A(s)$  и  $B(s)$  соответственно. Произвольный комплекс может быть образован фиксированной первой субъединицей и вращающейся со сдвигами второй субъединицей. Вращение описывается углами Эйлера  $\alpha$ ,  $\beta$  и  $\gamma$ , а

сдвиг вектором  $u = (u_x, u_y, u_z)$ , так что вся операция описывается шестью параметрами.

Если обозначить через  $C(s)$  амплитуду рассеяния от смещенной второй субъединицы, то рассеяние от комплекса может быть выражено как

$$I(s) = I_A(s) + I_B(s) + 2 \left\langle A(\mathbf{s}) C^*(\mathbf{s}) \right\rangle_{\Omega}. \quad (2.23)$$

Используя разложение (2.15) и учитывая ортогональные свойства сферических гармоник, уравнение (2.23) можно свести к следующему выражению:

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^l \left\{ |A_{lm}(s)|^2 + |B_{lm}(s)|^2 + 2 \operatorname{Re} [A_{lm}(s) C_{lm}^*(s)] \right\}. \quad (2.24)$$

При условии, что структуры субъединиц известны, можно вычислить амплитуды рассеяния и парциальные функции  $A_{lm}(s)$  и  $B_{lm}(s)$  из атомных моделей, как показано в [146,147]. Парциальные функции  $C_{lm}(s)$  вращающейся и сдвинутой второй субъединицы выражаются аналитически:

$$C_{lm}(s) = \sum_{k=-l}^l D_{mk}^l(\alpha, \beta, \gamma) C_{lk}^0(s, \mathbf{u}), \quad (2.25)$$

где  $D_{mi}^l(\alpha, \beta, \gamma)$  обозначают элементы матрицы конечных вращений.

Таким образом, моделирование жесткими телами заключается в определении шести параметров второй субъединицы, минимизирующей невязку (2.22) между экспериментальной и расчетной кривой рассеяния. Этот принцип легко обобщается на случай большего числа субъединиц, а также для симметричных частиц (в последнем случае, достаточно уточнять расположение субъединиц в анизотрической части комплекса).

Метод молекулярной тектоники получил в настоящее время широкое распространение для анализа строения биологических комплексов и макромолекулярных машин (например, обзор [143]). Этот подход, наряду с определением формы, оказывается возможным применять для трехмерного моделирования наночастичных систем различной природы.

## 2.2 Изучение формы наночастиц и биомолекул в растворе по данным малоуглового рассеяния

### 2.2.1 Проблема однозначности

Диаграмма функциональных связей между прямыми и обратными пространствами может быть представлена в виде, показанном на Рисунке 2.2.

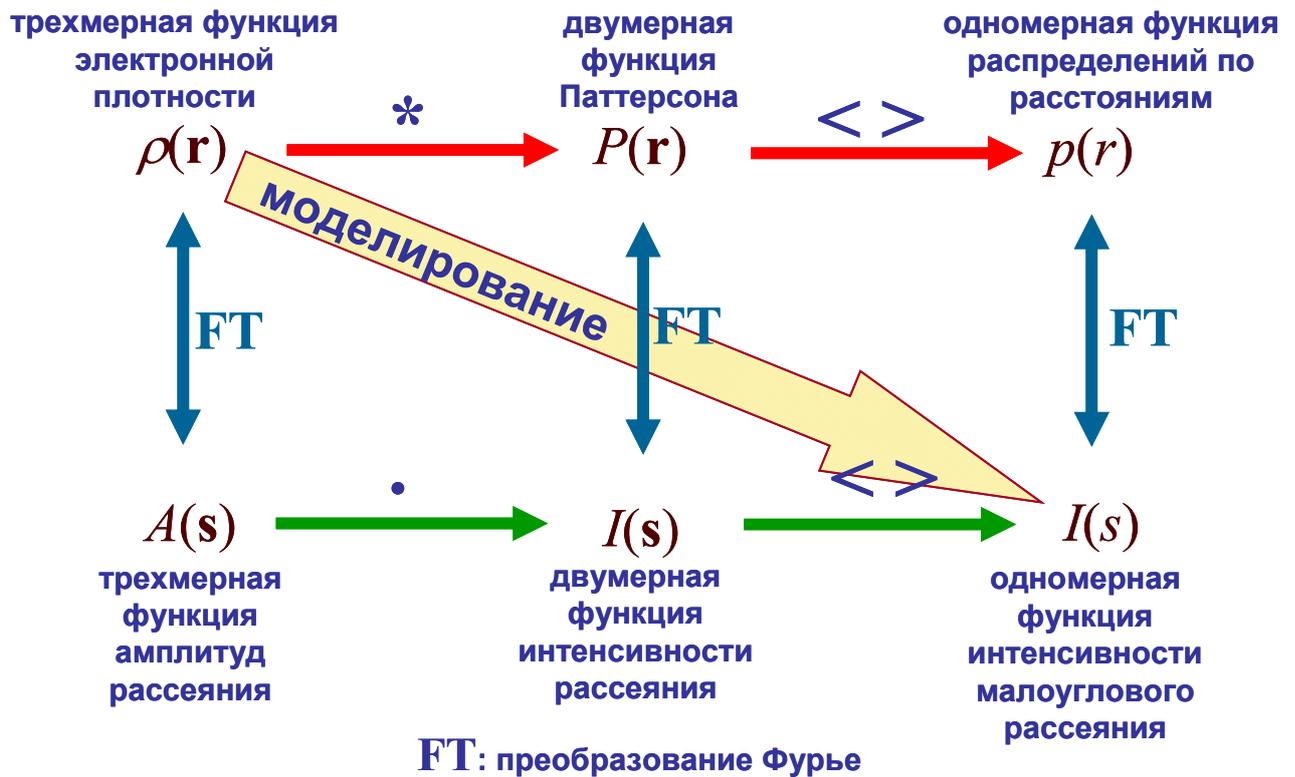


Рисунок 2.2 - Однозначность функциональных связей между функциями структуры объекта и функциями в обратном пространстве (амплитудами и интенсивностями рассеянного на образце излучения). Направления стрелок указывают однозначность преобразований: (\*) - свертки, (<>) - усреднение по телесному углу, (·) - комплексное возведение в квадрат, (FT) - преобразование Фурье.

Из этой диаграммы видно, что однозначно получить функцию структуры по данным дифракции или рассеяния какого-либо излучения, то есть решить обратную задачу, в общем случае невозможно. Но всегда есть возможность, зная модель структуры, однозначно рассчитать теоретическую функцию амплитуды

или интенсивности рассеяния и, сравнивая ее с экспериментальными данными, проверить справедливость модели и оценить величину ошибки в обратном пространстве. Минимизируя эту ошибку путем модифицирования функции структуры с помощью какой-либо процедуры, мы сможем найти модель объекта, которая не противоречит экспериментальным данным. Однако, и это не означает, что такая модель единственна в случае малоуглового рассеяния, интенсивность которого, как видно на диаграмме, связана со структурой максимально опосредованно. Этот факт специально отмечали многие исследователи, однако в литературе трудно найти примеры двух разных по структуре тел, малоугловое рассеяние от которых было бы идентично. Идентичность кривых рассеяния означает, что телам должны соответствовать одинаковые функции парных расстояний  $p(r)$  (формула 2.6), из чего следует идентичность максимальных размеров  $D_{max}$  и радиусов инерции  $R_g$ , которые определены формулой (2.7).

Кроме того, тела должны иметь одинаковые объемы, если учесть идентичность интенсивностей рассеяния в нулевой угол:

$$I(0) = |A(0)|^2 = 4\pi \int_0^D p(r) \cdot dr = \rho^2 V^2 \quad (2.26)$$

Это, конечно, ограничивает число возможных форм рассеивающих частиц, но они могут различаться в деталях. В литературе существует пример расположения 5 точек на плоскости двумя разными способами, при идентичных наборах парных расстояний между вершинами графов и, дополнительно, идентичности наборов площадей треугольников [148]. Эксперимент с опубликованным примером показал, что, если вывести центральную точку из плоскости на одинаковое расстояние (что обеспечивает равенство объемов полученных пирамид) и, полагая тела однородными по плотности, численно рассчитать функции расстояний  $p(r)$  с уменьшающимся шагом по пространству, можно увидеть, что кривые распределений стремятся к совпадению. Формы этих моделей показаны на Рисунке 2.3.

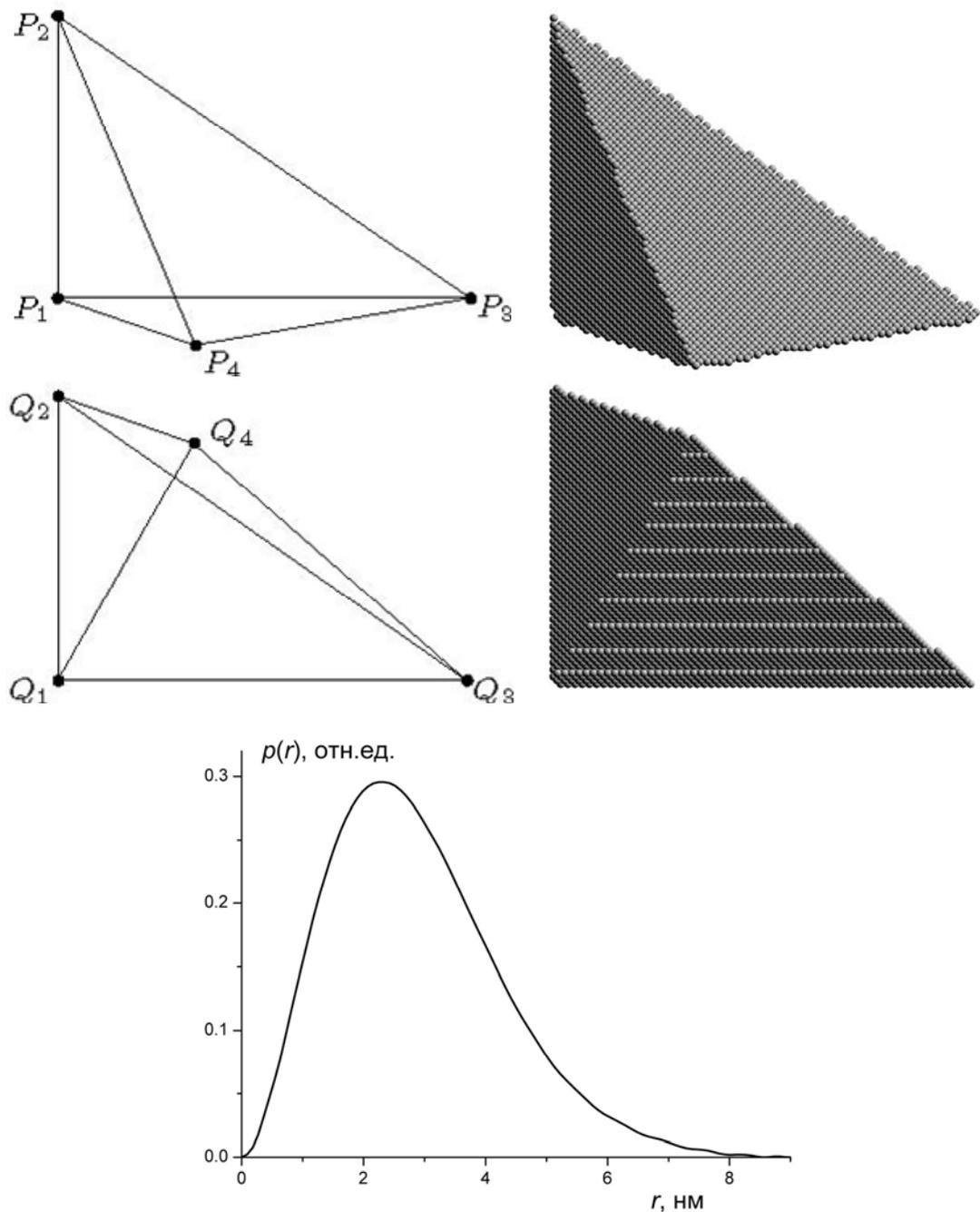


Рисунок 2.3 – Пример двух пирамидальных тел одинаковой высоты относительно плоскости Рисунка (точки  $P_4$  и  $Q_4$  есть вершины), обладающих идентичными функциями парных расстояний  $\rho(r)$ . Слева - опубликованная модель на плоскости. Справа показаны соответствующие объемные шариковые модели, составленные примерно из 20000 шариков, с размером  $Q_3-Q_1 = P_3-P_1 = 9$  нм. Внизу - вид функций парных расстояний, рассчитанных по шариковым моделям (совпадают с точностью толщины линии на графике).

На практике проблема поиска формы частицы по данным малоуглового рассеяния усложняется еще тем, что соответствующая численная задача оказывается плохо обусловленной в окрестности минимума целевой функции, что

означает существование различных решений, совпадающих по теоретической интенсивности рассеяния в пределах ошибок расчетов и возможностей алгоритмов минимизации. Поэтому в работе особое внимание было уделено выбору наиболее эффективных для решения задач поиска формы частиц алгоритмов минимизации нелинейных функций и разработке программ, адаптирующихся к текущей точности расчетов.

Плохая обусловленность задачи поиска связана, в частности, с невысокой информативностью данных малоуглового рассеяния, связанной и с небольшим доступным диапазоном углов, и с усреднением интенсивности рассеяния. Формально [126], число независимых параметров, описывающих кривую малоуглового рассеяния, можно выразить числом интервалов длиной  $\pi / D_{max}$ , укладываемых в угловой диапазон измерений, где  $D_{max}$  - максимальное значение расстояния, при котором функция парных расстояний  $p(r)$  еще не равна 0.  $D_{max}$  по сути является периодом высшей гармоники в (2.6), то есть спектр функции интенсивности рассеяния ограничен и поэтому к ней применима теорема Котельникова (в англоязычной литературе — теорема Найквиста — Шеннона) или теорема отсчетов [149,150]), которая утверждает следующее: если аналоговый сигнал имеет ограниченный по частоте спектр, то он может быть восстановлен однозначно и без потерь по своим дискретным отсчетам, взятым с частотой, строго большей удвоенной верхней частоты. Этой теореме соответствует формула точного разложения функции  $f(t)$  со спектром ширины  $D$ :

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2D}\right) \frac{\sin(2\pi Dt - n\pi)}{(2\pi Dt - n\pi)}$$

и шагом узлов  $\leq 1/2D$ , запись которой аналогична формуле представления функции интенсивности малоуглового рассеяния [122]:

$$s \cdot I(s) = \sum_1^{\infty} s_n \cdot I(s_n) \cdot \left\{ \frac{\sin[D_{max}(s - s_n)]}{D_{max}(s - s_n)} - \frac{\sin[D_{max}(s + s_n)]}{D_{max}(s + s_n)} \right\}. \quad (2.27)$$

Эта формула говорит только о максимально допустимом угловом шаге измерения данных. Если аналоговым сигналом является кривая интенсивности рассеяния, то максимально допустимы шагом ее представления будет  $\Delta s = \pi / D_{max}$ . В этом

рассмотрении предполагается, что спектр  $p(r)$  вычислен по кривой интенсивности рассеяния  $I(s)$ , заданной на полубесконечности. На практике,  $I(s)$  задана на ограниченном участке  $[s_{\min}, s_{\max}]$ . Для случая рассеяния от монодисперсных систем интенсивность может быть экстраполирована до  $s = 0$  с удовлетворительной точностью по соотношению Гинье (2.4). Точность этой аппроксимации зависит от точности определения радиуса инерции частицы  $R_g$  по (2.7). На практике, как показал опыт автора и отмечается в [122], ошибка в определении  $R_g$  при отсутствии агрегации и других примесей в образце, не превышает 1-2% и кривая  $I(s)$  достаточно надежно экстраполируется в нулевой угол. На больших углах интенсивность рассеяния от систем частиц падает по сравнению с нулевым углом на 2-4 порядка и имеет степенную тенденцию спада (степень 2-4, в зависимости от анизотрии формы частицы [121]), то есть интегральный вклад в рассеяние на больших углах очень мал, им можно пренебречь и рассчитывать спектр  $p(r)$  по ограниченному участку данных. Соответственно, число точек, которые соответствуют представлению сигнала  $I(s)$  без потерь, то есть позволят восстановить его с ошибкой, не превосходящей уровень экспериментальных шумов (который составляет обычно 0.5-2%, будет

$$N_S = \frac{(s_{\max} - s_{\min}) \cdot D_{\max}}{\pi} + 1. \quad (2.28)$$

Лузати [151] называет число  $N_S - 1$  числом степеней свободы, содержащемся в наборе данных рассеяния. Разумеется, точность восстановления по (2.27) будет зависеть и от эффекта обрыва  $I(s)$  на больших углах.

Это качественное рассуждение приводит к оценке числа независимых параметров в данных рассеяния, равным  $N_S$ . В литературе величину  $N_S$  часто называют числом **шенноновских каналов**. На практике диапазон экспериментальных данных рассеяния составляет от 4-5 до 20-25 каналов. Такая низкая информативность данных МУР приводит к тому, что восстановление формы или структуры частицы возможно только при небольшом числе параметров. И при этом приходится искать ответы на следующие вопросы.

- 1). Какие требования должны отвечать параметры структурной модели?

2). Каким должно быть число параметров модели? Как оно должно соотноситься с числом каналов в данных  $N_S$ ?

3). Насколько устойчиво решение, то есть насколько оно зависит от стартового приближения и ошибках в исходных данных?

При этом надо различать два типа ошибок. Первый тип - случайные шумы измерений. Влияние этих ошибок можно уменьшать, увеличивая статистику измерений. Практика автора показывает, что, если есть такая возможность, лучше увеличивать число точек - угловых отсчетов. Это позволяет одновременно контролировать однородность чувствительности детектора и вовремя распознавать связанные с неоднородностью систематические аппаратные ошибки. Второй тип ошибок связан с неомогенностью изучаемой системы. Так, образцы белков часто содержат агрегаты молекул, то есть в растворе присутствует некоторая доля ассоциатов - димеры, агломераты из многих молекул. Рассмотрение таких случаев будет дано в следующих разделах.

Структурные инварианты, рассмотренные в разделе 2.1, такие, как радиус инерции, объем частицы, площадь поверхности - однозначно определяются из данных рассеяния и могут рассматриваться как однопараметрические структурные модели. Точность определения инвариантов и параметризация данных МУР подробно рассмотрена в работах [152-154]. В этих работах число  $N_S$  непосредственно связано с точностью определения инвариантов по заданному интервалу данных. Однако, для описания пространственного строения частиц необходимо уже больше параметров. Если к параметрам модели предъявлять какие-либо требования, то представляется естественным выбирать модель с минимальным их числом. Тогда мы приходим к моделям с ортогональными вкладами структурных компонентов, то есть к разложениям по ортогональным гармоникам. Разложение должно обладать свойством сходимости, то есть увеличение числа слагаемых должно все более точно описывать структуру, а обрыв ряда будет ограничивать пространственное разрешение модели. Тем самым мы имеем возможность выбирать компромисс между разрешением и числом параметров. Один из таких подходов рассматривается в следующем разделе.

Что касается числа **независимых** параметров, то на первый взгляд оно не должно превышать число каналов  $N_S$ . В случае ортогональных моделей это число легко подсчитывается. Однако, такой подсчет ни о чем не говорит, так как задача определения формы (или структуры) частицы нелинейна, связь между структурой (например, описываемой функцией электронной плотности  $\rho(\mathbf{r})$ ) и интенсивностью рассеяния определяется самосверткой функции структуры:

$$\gamma(r) = \langle P(\mathbf{r}) \rangle_{\omega} = \frac{1}{2\pi^2} \int_{s=0}^{\infty} s^2 I(s) \cdot \frac{\sin(sr)}{sr} ds, \quad (2.29)$$

$$\gamma(r) = \langle P(\mathbf{r}) \rangle_{\omega} = \frac{1}{4\pi} \int_{\omega=0}^{4\pi} P(\mathbf{r}) \cdot d\omega = \langle \rho(\mathbf{r}) \otimes \rho(-\mathbf{r}) \rangle_{\omega}, \quad (2.30)$$

где  $\gamma(r)$  -- корреляционная функция, представляющая собой усредненную по телесному углу  $\omega$  в прямом пространстве самосвертку функции электронной плотности, или усредненную по всем ориентациям функцию Паттерсона  $P(\mathbf{r})$ . Нахождение модели структуры в общем случае (мы не будем рассматривать частные случаи сферически симметричных частиц, для которых показано, что определение радиальной функции электронной плотности однозначно [138,139,156,157]) при этом может быть сформулирована в терминах задачи нелинейных наименьших квадратов. Тем не менее, число  $N_S$  можно использовать для построения эмпирических критериев максимального числа параметров модели, как это будет показано дальше. В книге [126] отмечается, что связать число параметров модели с числом степеней свободы  $N_S$  затруднительно, и детальность модели частицы в значительной степени зависит от привлекаемой дополнительной информации (например, наличие элементов симметрии в структуре).

В общем случае необходимо различать **однозначность** решения задачи определения структурной модели и **устойчивость** решения. И неоднозначность, и неустойчивость приводят к тому, что находимые модели частиц отличаются друг от друга, если варьировать угловой диапазон данных рассеяния, проводить поиск с разных стартовых приближений, изменять параметры самого метода

минимизации. Но все структурные модели для данного набора данных должны обладать одинаковыми значениями инвариантов -  $R_g$ ,  $V$ ,  $S$ , что существенно сужает область неопределенности. Для получения окончательного ответа необходимо проводить серию численных экспериментов с варьированием условий поиска, анализировать отдельные решения с точки зрения "физического смысла", отбирать группы похожих решений (например, по  $R$ -фактору) и т.п. В итоге, анализ данных малоуглового рассеяния оказывается довольно сложной экспертной задачей, хотя нередко и удается получить окончательный ответ без проведения большого количества численных экспериментов.

Устойчивость решения можно также охарактеризовать числом обусловленности решаемой задачи, которое было рассмотрено в первой части, и которое всегда можно вычислить для конечно-разностной аппроксимации матрицы вторых производных целевой функции в любой точке гиперповерхности функции.

## 2.2.2 Моделирование формы наночастиц ортогональным рядом

В данном разделе будет рассмотрен один из примеров "оптимального" (с точки зрения числа параметров) представления модели формы частицы, предложенного в работах Г. Штурмана и Д.И. Свергуна [136,137,139]. Разложение функции формы частицы (2.20) можно рассматривать как случай составного тела, оболочка которого представлена суммой (смесью) ортогональных составляющих, то есть приближение формы однородной частицы рядом (2.20). На Рисунке 2.4 наглядно показан принцип разложения.

Гармоника с нулевым индексом представляет собой сферу примерно того же объема, что и частица, три гармоники с первым индексом 1 описывают сдвиг центра тяжести частицы из начала координат (и поэтому формально они несущественны и могут быть исключены из числа независимых параметров), остальные гармоники определяют деформацию сферической оболочки, тем детальнее, чем больше их первый индекс.

$$F(\omega) = f_{00} A_{00}(s) + f_{11} A_{11}(s) + f_{20} A_{20}(s) + f_{22} A_{22}(s) + \dots$$

$$\rho \quad r$$

$$F(\omega) - \text{функция оболочки}$$

Рисунок 2.4 - Наглядное представление разложения функции формы частицы в ряд по некоторым сферическим гармоникам  $A_{lm}$ .  $f_{lm}$  - искомые коэффициенты разложения. Число независимых параметров модели =  $(L+1)^2 - 6$ .

Для поиска параметров разложения  $f_{lm}$  в (2.20) по данным рассеяния была разработана программа SASHA, основанная на алгоритмах расчета модельной интенсивности, описанных в [136,137,139]. В программе вначале был использован программный пакет нелинейной минимизации OPTIS, разработанный автором для

решения задач разложения спектров смесей, рассмотренных в разделе 1.4.3. Целевая функция в программе представляет функционал, состоящий из слагаемого, соответствующего задаче наименьших квадратов, и нескольких штрафных членов, накладывающих ограничения на решение:

$$\min_{f_{lm}} \left\{ \frac{\sum_{i=1}^N [(I_{\text{exp}}(s_i) - \xi \cdot I(s_i)) \cdot W(s_i)]^2}{\sum_{i=1}^N [I_{\text{exp}}^2(s_i) \cdot W^2(s_i)]} + \sum_{k=1}^K g_k^2 \cdot W_k^2 \right\}, \quad (2.31)$$

где  $W(s)$  - весовая функция, ослабляющая относительный вклад в суммарную невязку начального участка кривой рассеяния и учитывающая оценки экспериментальных ошибок  $\sigma(s_i)$ :  $W(s_i) = s_i^2 / \sigma(s_i)$ , где  $i=1, \dots, N$ ,  $N$  - число точек в экспериментальной кривой рассеяния;  $\xi$  - вспомогательный МНК множитель, совмещающий кривые интенсивности рассеяния перед вычислением разности

$$\xi = \frac{(I_{W, \text{exp}} \cdot I_W)}{\|I_W\|^2}, \quad I_W(s_i) = I(s_i) \cdot W(s_i), \quad (2.32)$$

где  $I_{W, \text{exp}} \cdot I_W$  означает скалярное произведение векторов взвешенных экспериментальной и теоретической интенсивностей рассеяния;  $g_k$  - набор штрафных членов за отрицательность функции формы, уход центра тяжести тела из начала координат и за негладкость формы (в этом случае число штрафов  $K=3$ ). Штрафные члены  $g_k$  умножены на соответствующие весовые коэффициенты  $W_k$ , величины которых были подобраны экспериментально, на основе решения различных тестовых задач. Нормировка первого слагаемого - квадратичной невязки - на квадрат нормы взвешенной экспериментальной интенсивности

$$\sum_{i=1}^N [I_{\text{exp}}^2(s_i) \cdot W^2(s_i)] \quad (2.33)$$

обеспечивает независимость поиска от абсолютных значений интенсивности. Из литературы по методам минимизации известно, что целевые функции со штрафными слагаемыми как правило плохо обусловлены. Тем не менее,

использование штрафов необходимо для ограничения области допустимых решений путем введения дополнительных критериев приемлемости, которые невозможно аналитически выразить через параметры модели в рамках решаемой задачи наименьших квадратов.

Основные штрафные члены выглядят следующим образом:

а) штраф за отрицательность функции формы  $F(\omega)$ :

$$g_1 = \int_{\substack{\omega, \\ \forall F(\omega) < 0}} F(\omega) \cdot d\omega ; \quad (2.34)$$

б) штраф за уход центра тяжести однородного тела  $[x_c, y_c, z_c]$  из начала координат:

$$\begin{aligned} g_2 &= x_c, \\ g_3 &= y_c, \\ g_4 &= z_c ; \end{aligned} \quad (2.35)$$

в) штраф за “сложность” поверхности частицы, здесь  $S$  = площадь поверхности,  $R_g$  - радиус инерции,  $V$  - объем частицы:

$$g_5 = \frac{S \cdot R_g}{V} . \quad (2.36)$$

Вычисление интенсивности рассеяния по формулам (2.15 - 2.21) более чем на порядок быстрее, чем непосредственное вычисление по функции формы, численно заданной на угловой пространственной сетке по формулам (2.15, 2.17), где  $r$  - расстояние от начала координат до точек на поверхности.

Это важно на практике, при реализации вычислений на сравнительно маломощных персональных компьютерах. Но наиболее важное преимущество параметризации (2.19, 2.21) - в возможности ограничения пространственного разрешения простым обрывом ряда (2.20) до некоторого значения  $l=L$ . Из-за ортогональности членов ряда такой обрыв влияет на общую форму частицы минимальным образом. Пространственное разрешение  $\delta$  соответствующей формы определяется формулой

$$\delta = \frac{\pi R_0}{(L+1)}, \quad (2.37)$$

где  $R_0$  - радиус эквивалентной сферы. Число параметров в таком представлении функции формы равно, как уже указывалось,  $(L+1)^2$ .

Необходимость введения ограничения на пространственное разрешение связана с тем, что мелкими деталями формы можно описать вклад рассеяния от малоразмерных внутренних неоднородностей частицы и получить, тем самым, неверную структурную интерпретацию. В программе SASHA вклад от внутренних неоднородностей Д.И. Свергнуном было предложено приближенно моделировать аддитивным вкладом рассеяния от набора высших сферических гармоник с максимальным значением индекса  $L$ , в 5 раз превышающим максимальный индекс, используемый для описания функции формы. Коэффициенты этих гармоник рассчитываются с применением генератора псевдослучайных чисел внутри эллипсоида, аппроксимирующего форму частицы. Параметры эллипсоида, в свою очередь, определяются по начальному участку кривой рассеяния путем решения вспомогательной обратной задачи рассеяния с использованием известных формул, рассмотренных в [122,126], методом наименьших квадратов. Возможность моделирования неоднородностей предусмотрена в программе SASHA, однако на практике ею

При дальнейшем развитии программы [A14] в качестве процедуры минимизации была выбрана наиболее эффективная реализация алгоритма Левенберга-Маркардта для решения задач нелинейных наименьших квадратов - алгоритм NL2SOL [89,90,93,94]. В этой процедуре стандартная схема Левенберга-Маркардта усилена пересчетом матрицы Гессе целевой функции по схеме переменной метрики. При этом программа минимизации является мало чувствительной к плохому стартовому приближению и большой величине остаточной невязки в точке минимума, однако весьма чувствительна к выбору длины шагов по переменным функции, которые используются для вычисления первых производных. Поэтому схема NL2SOL была расширена блоком адаптации программы к ошибкам вычислений и определении на основании полученных

оценок оптимальной для шага приращения для каждого искомого параметра задачи. Соответствующие блоки программы практически идентичны примененным в пакете OPTIS. Исследование нескольких десятков модельных задач показало, что число обусловленности, определяемое как число обусловленности (определяемое уравнением 1.52) матрицы вторых производных целевой функции (2.31) в текущей точке поиска, может достигать  $10^4 - 10^6$ , что делает задачу близкой к вырожденной и, соответственно, приводит к большим девиациям решения, согласно оценкам (1.61), примененным к матрице вторых производных целевой функции.

Алгоритм NL2SOL был выбран для решения задачи поиска формы частиц в программе SASHA не случайно. Во-первых, примененная в нем схема алгоритма Левенберга-Марквардта, разработана для решения задач нелинейных наименьших квадратов вида

$$\min_x \left\{ F(\mathbf{x}) = \sum_{i=1}^N [f_i(\mathbf{x})]^2 = \|\mathbf{f}(\mathbf{x})\|_2^2 \right\}, \quad (2.38)$$

где  $\mathbf{f}(\mathbf{x})$  - нелинейная векторная функция, имеющая в нашем случае вид (под аргументом  $\mathbf{x}$  понимается набор искоемых коэффициентов разложения  $f_{lm}$ ):

$$\mathbf{f}(\mathbf{x}) = \left[ \left\{ \frac{(I_{\text{exp}}(s_i) - \xi \cdot I(s_i)) \cdot W(s_i)}{\sqrt{\sum_{i=1}^N [I_{\text{exp}}^2(s_i) \cdot W^2(s_i)]}} \right\}_{i=1, \dots, N}, \quad \left\{ g_k \cdot W_k \right\}_{k=1, \dots, 5} \right]. \quad (2.39)$$

Другими словами, элементы функции есть значения нормированных невязок и отдельные, необъединенные штрафные члены со своими весами. Веса для  $g_k$  задают в виде констант, выбираемых из соображения примерного равенства членов (2.39) в конце поиска решения.

Векторный характер  $\mathbf{f}(\mathbf{x})$  обуславливает высокую эффективность конечно-разностного пересчета аппроксимации матрицы вторых производных функции (2.31), что и определяет высокую скорость сходимости к минимуму. Поэтому в программе SASHA штрафные члены не объединяются, а добавляются к  $\mathbf{f}(\mathbf{x})$  как

дополнительные элементы "невязки", не возведенные в квадрат. Такой прием значительно улучшил сходимость процедуры поиска решения в программе SASHA по сравнению с традиционными подходами метода штрафов, в которых минимизируется скалярная функция.

Согласно алгоритму Левенберга-Марквардта [89,93], направление поиска  $\mathbf{p}_m$  на  $m$ -ом шаге вычисляется через матрицу первых производных  $J_m$  векторной функции  $\mathbf{f}$ , сумма квадратов элементов которой должна быть минимизирована:

$$\left( J_m^T \cdot J_m + \lambda_m \cdot I \right) \cdot \mathbf{p}_m = -J_m^T \cdot \mathbf{f}_m, \quad (2.40)$$

где  $\lambda_m$  - некоторое неотрицательное число, с помощью которого регулируют сходимость метода. В качестве следующей точки выбирают

$$\mathbf{x}_{m+1} = \mathbf{x}_m + \mathbf{p}_m. \quad (2.41)$$

Член  $J_m^T \cdot J_m$  есть аппроксимация матрицы вторых производных  $H(x_m)$  функции (2.31) в точке  $m$ , за исключением поправочного члена:

$$H(x_m) = J(x_m)^T \cdot J(x_m) + Q(x_m) \quad (2.42)$$

В случае линейной задачи  $Q(x_m) = \sum_{i=1}^m \mathbf{f}(x_i) \cdot H(x_i)$ . В нелинейном случае

поправочный член можно вычислять по схемам алгоритмов переменной метрики, что делает поиск адаптирующимся к резким изменениям формы поверхности отклика целевой функции. В алгоритме NL2SOL предложено использовать схему пересчета матрицы  $Q$  по формуле Бройдена-Флетчера-Голдфарба-Шанно для матрицы вторых производных [89]:

$$Q(\mathbf{x}_{m+1}) = Q(\mathbf{x}_m) - \frac{1}{\mathbf{p}_m^T \cdot W_m \cdot \mathbf{p}_m} \cdot W_m \cdot \mathbf{p}_m \cdot \mathbf{p}_m^T \cdot W_m + \frac{1}{\mathbf{y}_m^T \cdot \mathbf{p}_m} \cdot \mathbf{y}_m \cdot \mathbf{y}_m^T, \quad (2.43)$$

где  $\mathbf{p}_m = \mathbf{x}_{m+1} - \mathbf{x}_m$  - вектор длины шага линейного поиска на  $m$ -ой итерации,  $\mathbf{y}_m = J(x_{m+1})^T \cdot \mathbf{f}(x_{m+1}) - J(x_m)^T \cdot \mathbf{f}(x_m)$  - вектор разности градиентов функции, вычисленных на последовательных шагах.

Длину шага линейного поиска вдоль направления  $p_m$  можно выбирать разными способами, рассмотренными, например, в [89]. В случае задачи поиска формы частицы оказалось, что наиболее эффективен выбор длины из условия минимума вдоль направления поиска  $p_m$ . Для поиска минимума был применен относительно простой и достаточно эффективный алгоритм Брента [158], обеспечивающий не более 5-10 пробных вычислений функции на каждой итерации. В качестве критерия останова поиска вдоль  $p_m$  был применен критерий приблизительного равенства ошибок вычисления градиента  $J(x_m)^T \cdot f(x_m)$ , связанных с конечно-разностной схемой, и ошибок, которые можно назвать "шумами вычисления" целевой функции, зависящими не только от ошибок округления, но и ошибок вычислений, связанных с численной реализацией формул расчета интенсивности малоуглового рассеяния по (2.15, 2.21). Так как аналитически все такие ошибки в нелинейном случае учесть практически невозможно, в алгоритме SASHA была реализована схема расчета погрешности вычислений по таблицам конечных разностей, примененная в пакете OPTIS, расширенная для случая векторных функций. Получаемые оценки ошибок вычислений используются программой для назначения оптимальных величин шагов для поиска и вычисления производной целевой функции. Применение всех модификаций алгоритма NL2SOL сделало поиск функции формы достаточно эффективным: для получения окончательного решения программе требуется не более 50 итераций (против нескольких тысяч или десятков тысяч с случае использования методов переменной метрики для минимизации скалярной функции (2.31), причем часто решение оказывается при этом недопустимо далеко от области минимума).

### 2.2.2.1 Численное исследование неоднозначности и стабильности решения

Число независимых параметров, описывающих форму, определенную рядом (2.20) по  $L$  сферическим гармоникам, есть  $N_L=(L+1)^2 - 6$ . Уменьшение на шесть параметров обусловлено произвольностью положения частицы (три координаты центра тяжести тела) и ее ориентации (три угла поворота). Пространственное разрешение построенной таким образом модели можно оценить величиной  $\pi R_0 / (L+1)$ , где  $R_0$  - радиус эквивалентной сферы, то есть сферы равного с частицей объема. С ростом  $L$  разрешение улучшается, но одновременно возрастает число неизвестных в параметризации (2.20), что осложняет устойчивое восстановление формы частицы, то есть независимость решения от стартового приближения и параметров метода поиска. Надежность результата определяется, очевидно, соотношением между  $N_L$  и величиной  $N_S$  (2.28), определяющей информационное содержание экспериментальных данных. Аналитически определить эту зависимость не представляется возможным в силу нелинейного характера соотношений, связывающих набор параметров  $\{f_{lm}\}$  с модельной интенсивностью рассеяния  $I_L(s)$  (2.15, 2.20). Поэтому с этой целью был выполнен ряд модельных численных экспериментов по принципу “мультистарта”, хорошо известному в теории нелинейной оптимизации [89,90,92]. Суть его заключается в том, что производится многократный запуск оптимизационного алгоритма с различными начальными значениями, которые выбираются случайно в разных областях пространства возможных решений. Если при этом каждый раз оптимум достигается на одном и том же векторе параметров (или с малым их разбросом), то можно с определенной вероятностью утверждать, что получаемое решение однозначно, и соответствует глобальному минимуму целевой функции.

Общая схема модельных экспериментов была следующей [A9]. Вначале конструировали оболочку однородного тела, форма которой описывалась набором сферических гармоник  $\{f_{lm}\}$ ,  $-l \leq m \leq l$ ;  $l \leq L$  для заданной величины  $L$ . Для этого тела рассчитывали модельную интенсивность рассеяния по формуле (2.15), при этом длину ряда брали не менее  $l=L+5$  во избежание численных

эффектов обрыва. Частные амплитуды  $A_{lm}$  рассчитывали по формуле (2.23) до  $p_{max}=40$ , так как учет слагаемых более высокого порядка уже не приводил к увеличению точности из-за ошибок округления машинной арифметики. С целью проверки устойчивости используемых алгоритмов восстановление формы проводилось по данным модельной интенсивности искаженным случайным шумом с относительной интенсивностью 10%, а также по данным интенсивности рассеяния без участка Гинье (что соответствует экспериментальным измерениям с отброшенным участком, в котором значительно влияние рассеяния от примесей агрегатов частиц), искаженным 3%-ным относительным шумом и систематическими погрешностями (имитация реальных условий малоуглового эксперимента - ошибки вычитания рассеяния от растворителя). При этом на вход программы подавали участки кривой интенсивности различной длины, содержащие различное число информационных каналов  $N_S$ . В целом проведенные модельные эксперименты позволяют утверждать, что при  $L \leq 3$  и  $N_L \leq 1.5 N_S$  полученные решения однозначны (с точностью до общего вида формы и энантиоморфного преобразования частицы, не меняющего интенсивности рассеяния) в классе моделей, описываемых функцией формы (2.20).

### Эксперимент 1.

Построенные модельные частицы, от которых рассчитывали теоретическую интенсивность рассеяния показаны на Рисунке 2.5. Частица 1 имеет спадающий спектр вкладов сферических гармоник в разложении (2.20), во II-ой слегка доминирует вторая гармоника, в III-ей - третья, что видно по форме оболочек. Численные значения отдельных вкладов  $f_{lm}$  здесь несущественны, поэтому не приводятся. Максимальный индекс гармоник в разложении составлял  $L = 3$ . Максимальный диаметр для всех тел был  $D_{max} = 18.3$  нм.

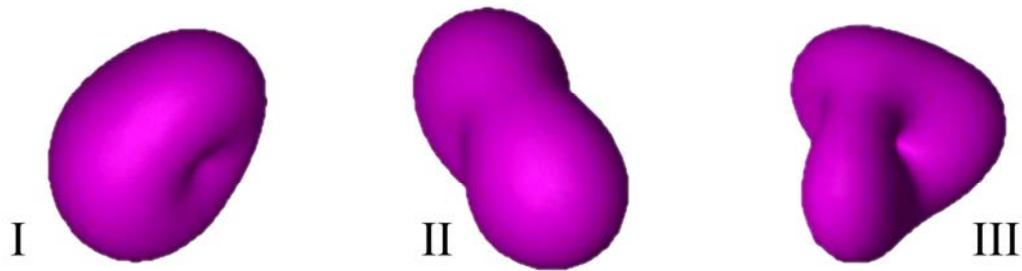


Рисунок 2.5 – Модельные оболочки однородных частиц при  $L = 3$ , построенные для изучения однозначности восстановления формы. Пояснения в тексте.

Далее от построенных моделей рассчитывали интенсивность МУР, пользуясь соотношениями (2.21 и 2.15) и использовали эти данные в качестве экспериментальных. С целью снижения влияния точности машинной арифметики, значения интенсивности сохраняли при 15-16 значащих цифрах в мантиссе. Псевдослучайных шумов в этом эксперименте не добавляли.

Перед началом экспериментов была проведена оценка машинной точности вычислений по формулам разложений 2.15-2.21 и по формуле расчета интенсивности непосредственно по квазиравномерной сетке точек на поверхности оболочки частицы  $F(\omega)$ , применяя численное интегрирование по соотношению

$$A_{lm}(s) = i^l \sqrt{\frac{2}{\pi}} \int_{\omega} Y_{lm}(\omega) \cdot d(\omega) \int_0^{F(\omega)} j_l(sr) \cdot r^2 \cdot dr . \quad (2.44)$$

Число точек на поверхности выбирали порядка 1000, что обеспечило абсолютную точность расчетов не хуже  $1.0^{-7}$ . Увеличение числа точек далее не приводило к улучшению точности. Расчет по дискретной сетке из нескольких сотен или тысяч точек на поверхности оболочки требует в десятки и сотни раз больше времени, чем по формулам степенных разложений, но обеспечивает большую точность на больших углах рассеяния.

Под точностью здесь имеется в виду не характеристика отклонения интенсивности (от правильного значения), обусловленного приближенностью используемой формализации, а флуктуация результатов, которая обусловлена совокупностью ошибок округления, накапливаемых в ходе всех вычислений. Автор называет эту ошибку "шумом вычислений". В ходе разработок программного обеспечения автор уделял особое внимание малости ошибок

расчетов всех специальных функций, суммирования рядов, численного интегрирования и других промежуточных вычислений. Шум вычислений пагубно сказывается на эффективности работы программ минимизации, так как при малых вариациях целевой функции расчет ее градиентов становится бессмысленным и поиск прекращается тем дальше от минимума, чем больше шумы. Аналитически предсказать шум вычислений невозможно, так как если такие оценки и существуют для каждого шага в цепочке расчетов, они являются, как правило, оценками сверху и практического смысла не имеют. Но можно провести оценку "по факту", проводя вариации параметров модели и анализируя отклик программы. Для этой цели автором был разработан ряд программ, которые исследуют целевую функцию задачи, проводя малые вариации ее аргументов или параметров модели. Получаемые значения целевой функции записываются в вектор  $F_0$ . Достаточное для практики число вариаций составляет 15-20 по каждому параметру. Величина вариации параметра  $\delta_x$  выбирается из приближенного условия оптимального приращения для конечно-разностной аппроксимации третьей производной целевой функции при отсутствии дополнительной информации [89]

$$\delta_x = (1 + x) \cdot \sqrt[3]{\frac{3 \cdot \delta_F}{F'''}} , \quad \delta_F = EPS, \quad F''' = 1.0 , \quad (2.45)$$

где  $EPS$  - машинное эpsilon, или минимальное положительное число, которое при добавлении его к единице дает результат, отличный от единицы. Для ANSI арифметики двойной точности, используемой в расчетах оно составляет  $2.22 \cdot 10^{-16} = 2^{(-52)}$ . Надо отметить, что алгоритм расчета шумов вычислений мало чувствителен к выбору величины приращения, его можно варьировать в пределах 2-3х порядков или даже более. Переносимость программ на разные платформы была обеспечена разработкой модулей расчетов машинно-зависимых констант для произвольной машинной архитектуры и типа используемого компилятора, учитывающих эффекты денормализации точности представления чисел в процессоре. Собственная разработка была необходима, так как известные аналоги

подобных программ оказались несостоятельны на современных процессорных архитектурах.

Полученные при вариации параметра  $x$  значения целевой функции заносят в первый столбец таблицы конечных разностей. В следующие столбцы записывают величины разностей двух последовательных элементов предыдущего столбца, получая столбцы первых, вторых, третьих и т.д. разностей. В [159] доказана теорема о том, что если в столбце  $n$ -ых разностей знаки последовательных элементов чередуются (то есть коэффициент корреляции значений последовательных элементов стремится к  $-1$ ), то их величины обусловлены дисперсией случайных вариаций в первом столбце исходных значений. Зная номер разности, из величины дисперсии  $\delta_n$  соответствующего столбца можно провести обратный расчет дисперсии шумов расчетов функции

$$\delta_F = \frac{\delta_n}{\sqrt{\frac{(2n)!}{(n!)^2}}} . \quad (2.46)$$

Следует отметить, что данный алгоритм был успешно применен автором во всех разработанных программах поиска минимума нелинейных функционалов для оценки величин приращений аргументов при вычислении градиентов функций и, как будет рассмотрено далее, для адекватной автоматической оценки начальной виртуальной температуры в методе "моделирования отжига". Это сделало программы адаптирующимися к целевой функции и избавило пользователя от трудоемкой оценки многих критичных для процесса поиска параметров, в том числе и критериев качества решения и остановки процедуры.

Таким образом удалось оценить точность вычислений интенсивности рассеяния по двум подходам. В качестве примера на Рисунке 2.6 приведено сравнение относительных шумов вычислений интенсивности в зависимости от угла рассеяния. Из рисунка видно, что относительная ошибка интенсивности, рассчитываемой по (2.15-2.21) нарастает по степенному закону, будучи очень малой в малых углах и становясь недопустимо большой в средних. Это можно считать преимуществом метода, так как среднеугловая часть данных МУР (при

числе шенноновских каналов более 4-8) отвечает уже значительному вкладу рассеяния от малоразмерных внутренних неоднородностей частицы и поэтому для определения формы низкого разрешения ее не используют. Нарастание ошибки при использовании степенных разложений обусловлено ухудшением сходимости ряда (2.21) с ростом  $s$ . Расчет по формуле (2.44) менее пригоден для определения формы как в силу большей ошибки вычислений при малых  $s$ , так и значительного времени расчетов.

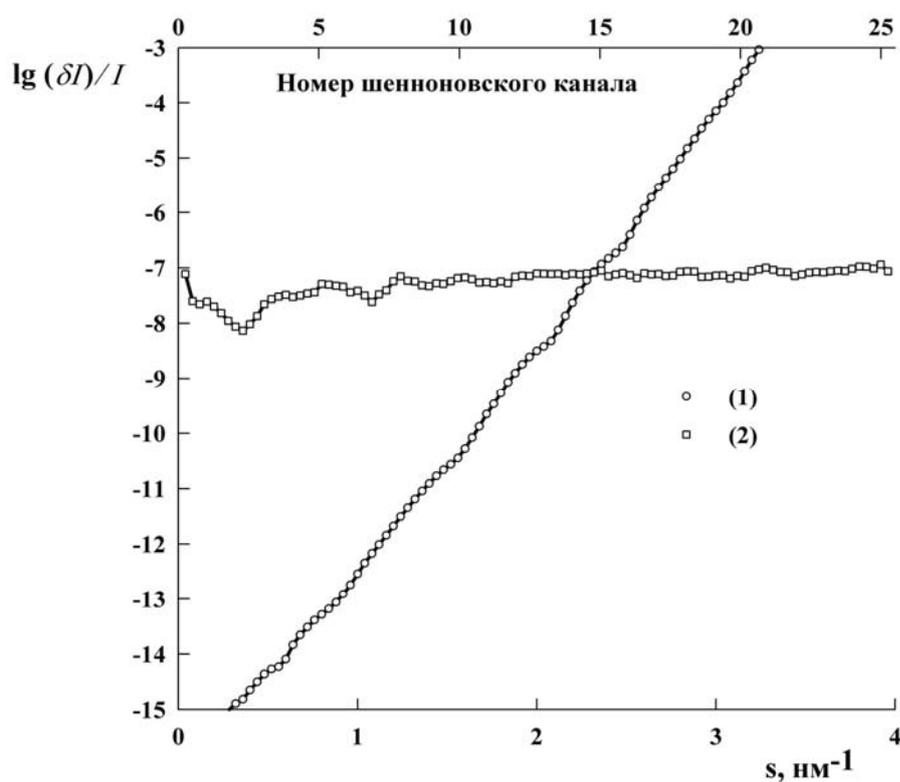


Рисунок 2.6 – Относительная ошибка вычислений интенсивности рассеяния от тела I (Рисунок 2.5): (1) с использованием степенного ряда (2.21) (возрастающая зависимость) и (2) численного интегрирования (2.44) (горизонтальная кривая).

Анализ полученных зависимостей шумов измерений привел к выводу, что восстановление формы оболочки частицы в однородном приближении ее структуры теряет устойчивость при числе параметров модели, превышающем  $1.5 \cdot N_S$  ( $N_S$  - число шенноновских каналов). Пример потери устойчивости будет показан намного позднее.

Вернемся к моделям, рассмотренным в начале данного раздела. При вычислении интенсивностей степенной ряд (2.21) ограничивали  $p_{\max} = 40$ , так как

увеличение его длины не приводило к улучшению точности. Ряд (2.15) обрывали на  $l = L+5$ , где  $L$  - число гармоник в представлении формы частиц. Это сделало пренебрежимо малым влияние ошибок обрыва ряда. Весовую функцию  $W^2(s_i)$  в формуле (2.31) брали, как предложено в [167]:

$$W(s) = s^2, \quad (2.47)$$

применение которой уменьшает диапазон интенсивностей при расчете невязки и делает программу более чувствительной к невязке при больших  $s$ . Для подгонки брали разные участки исходных данных, как показано на Рисунке 2.7.

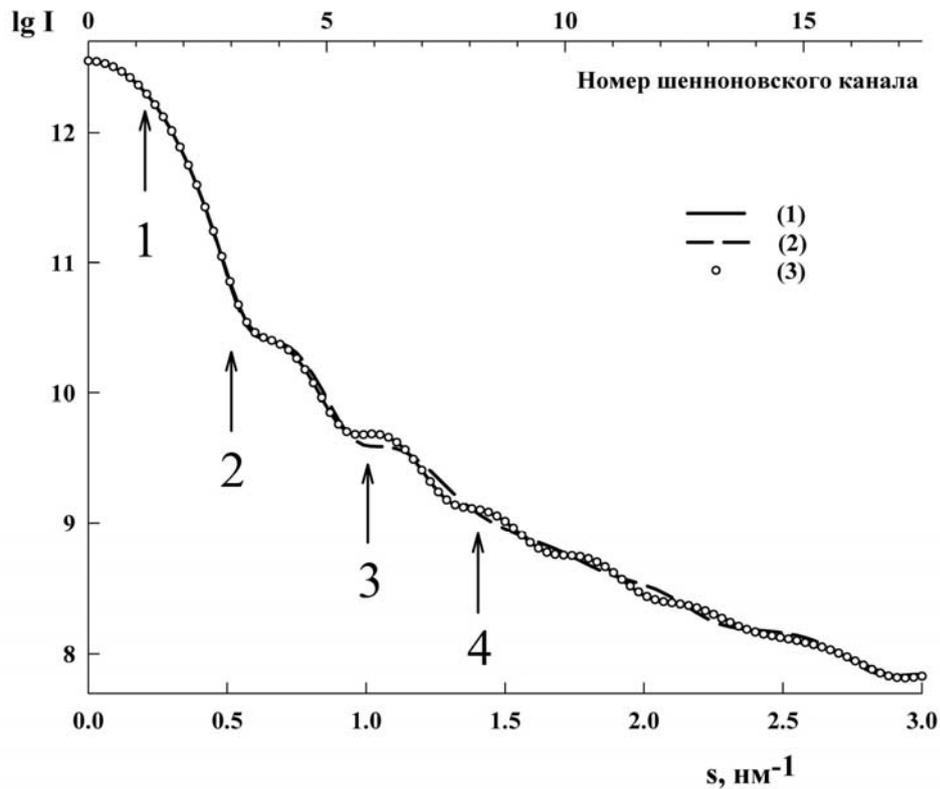


Рисунок 2.7 – Модельные кривые МУР от частицы I Рисунка 2.5. (1) - исходная кривая, используемая в качестве экспериментальной. (2) - МУР от частицы, восстановленной по диапазону 1. (3) - восстановление по диапазону 2, кривая практически совпадает с исходными данными. Восстановления формы по диапазонам 3 и 4 приводят к еще более точному восстановлению исходной кривой. Все кривые от найденных моделей рассчитаны с превышением углового диапазона подгонки с целью демонстрации их предсказательного характера. Величина шенноновского канала (2.28) рассчитана для максимального диаметра оболочки 18.3 нм.

Во всех случаях модельных расчетов в качестве начального приближения формы брали сферу.

Для характеристики степени схождения модельных тел (mod) с восстановленными (rest) использовали величину  $R$ -фактора

$$R_{\omega}^2 = \frac{\sum_{\omega} [F_{\text{mod}}(\omega) - F_{\text{rest}}(\omega)]^2}{\sum_{\omega} [F_{\text{mod}}(\omega)]^2} \quad (2.48)$$

суммирование ведется по квазиравномерной сетке направлений при числе соответствующих узлов на поверхности оболочки от 100 до 800. Значения  $R$ -факторов для рассматриваемых в разделе численных экспериментов сведены в Таблицу 2.1.

Таблица 2.1.  $R_{\omega}$ -факторы схождения модельных и восстановленных форм частиц.

модель		Диапазон $s$ (обозначения на рисунках)						Псевдо-реальные диапазон $s$			
		1		2		3		4			
		Точные данные	Данные с 10% шумом	Точные данные	Данные с 10% шумом	Точные данные	Данные с 10% шумом	Шум 10% отн.			
L= 3	I	7.4	10.4	4.2	6.1	1.4	5.1	5.9	5.3	5.9	6.3
	II	9.9	33.0	0.2	5.0	0.5	1.4				
	III	11.1	27.1	7.9	44.0	5.4	10.2				
L= 4	I	9.2	13.3	12.4	11.6	3.9	9.2	10.0	10.4	10.0	10.4
	II	10.7	17.2	8.3	25.3	3.8	10.0				

Неожиданным, даже удивительным, оказался факт хорошего и устойчивого восстановления форм при очень малом угловом диапазоне данных - менее 1.5 шенноновских интервала и числе независимых параметров моделей 16 - 6=10 (область 1 на Рисунке 2.7). При этом число точек для расчета невязки в этом диапазоне было всего 5. После этого неудивительно улучшение  $R$ -фактора для больших диапазонов. Для дополнительной проверки устойчивости

восстановления был поставлен отдельный численный эксперимент, который заключался в добавлении к целевой функции (2.31) дополнительного штрафного члена за приближение рабочей модели к точной. Этот штраф вычислялся как обратная величина к (2.48) с соответствующим весовым вкладом, который варьировался от эксперимента к эксперименту от больших до малых значений, таких, что вклад штрафа в (2.48) составлял от 100 до 1%. При большом штрафе программа была не в состоянии найти модель, удовлетворительно восстанавливающую кривую рассеяния. По мере уменьшения штрафа находимое решение приближалось к точному телу. Таким образом, для тел на Рисунке 2.5 численно была показана единственность восстановления, несмотря на значительное превышение числа параметров над числом шенноновских каналов. Решения показаны на Рисунке 2.8.

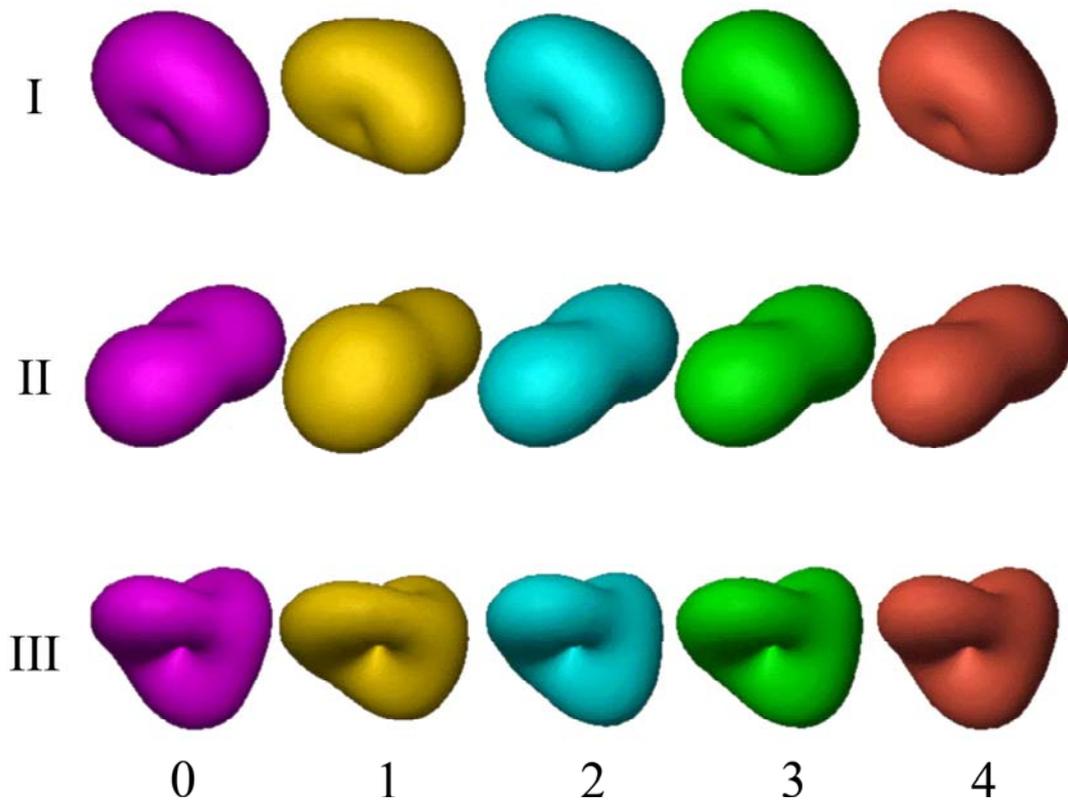


Рисунок 2.8 – Восстановление формы частиц, представленной до  $L = 3$ , по разным участкам кривой интенсивности МУР. Столбец 0: точные модели. Столбцы 1-4: восстановление по участкам 1-4 Рисунка 2.7. Римскими цифрами обозначены формы тел, соответственно Рисунку 2.5.

Перекрестные восстановления (старт с другого тела) привели к тем же устойчивым решениям. Разумеется, полученный результат не является общим для всех возможных форм и условий эксперимента.

### **Эксперимент 2.**

В численных экспериментах по восстановлению формы, описываемой рядом (2.20) до  $L=4$  ( $25-6=19$  независимых параметров,  $D_{\max} = 20.3$  нм) брали разные стартовые приближения. Кривые рассеяния и результаты восстановлений формы до  $L = 4$  показаны на Рисунках 2.9 - 2.10. Верхний ряд фигур на Рисунке 2.10 соответствует старту со сферической формы, нижний - с анизометричной формы с преобладанием высших гармоник. Больше, чем в предыдущем эксперименте, число параметров модели привело к большей нестабильности решения. Малый угловой диапазон 1 привел к решениям с большим отклонением от точного тела, чем в случае с  $L = 3$  (Таблица 2.1, Рисунок 2.10, столбец 1.). Однако, увеличение углового диапазона до значений, соответствующих реальным экспериментам (диапазон 3), сделало восстановление практически независимым от стартового приближения (столбец 3 в верхнем и нижнем ряду фигур). Внутренний диапазон 2 реального практического смысла не имеет, расчет сделан для проверки важности начального участка кривой рассеяния, в котором содержится информация об анизометрии формы и ее деталях низкого разрешения. Потеря этой информации действительно влияет на результат восстановления (столбец 2, строки А и Б).

### **Эксперимент 3.**

Для проверки стабильности решения от случайных шумов в данных, к исходным интенсивностям был добавлен псевдослучайные нормально распределенные числа с относительной дисперсией 10%. Результат для тел с  $L = 3$  (Рисунки 2.5 и 2.8) показан на Рисунках 2.11 и 2.12, для тел с  $L = 4$  - на Рисунках 2.12 и 2.13. Добавление случайного шума особенно сильно повлияло на качество восстановления при малых угловых диапазонах 1 и 2. Полный диапазон 3, соответствующий качественному малоугловому эксперименту, позволил программе восстановить форму с приемлемым качеством (Рисунок 2.12, столбец

4). Результаты сравнения решений с исходными моделями представлены в Таблице 2.1.

Результаты восстановления тела с  $L = 4$  представлен на Рисунке 2.13.

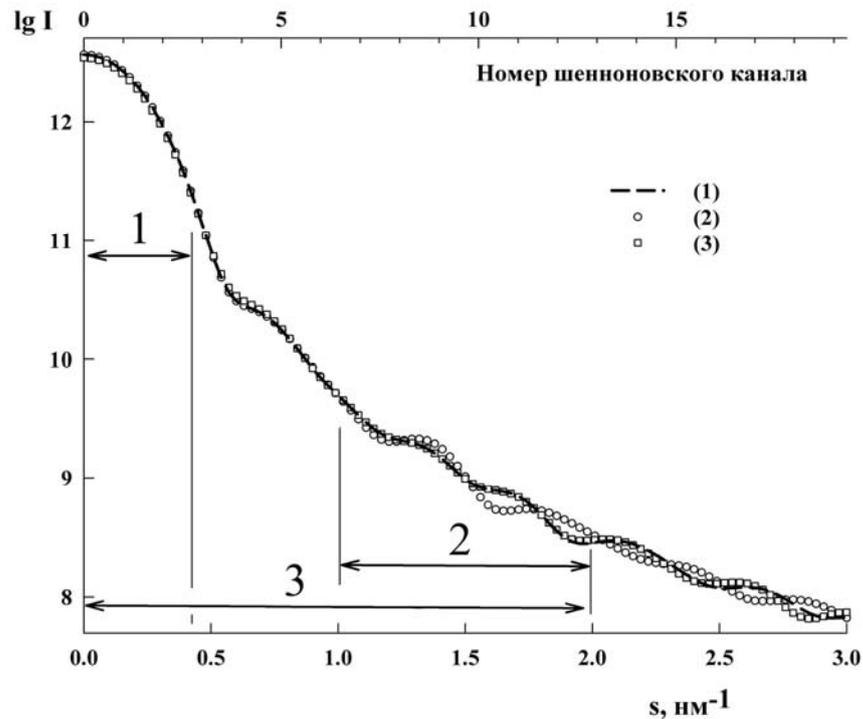


Рисунок 2.9 – Модельные кривые МУР от частицы. (1) - исходная кривая, используемая в качестве экспериментальной. (2) - МУР от частицы, восстановленной по диапазону 1. (3) - восстановление по диапазону 2, кривая практически совпадает с исходными данными. Кривые от найденных моделей рассчитаны с превышением углового диапазона подгонки с целью демонстрации их предсказательного характера. Величина шенноновского канала рассчитана для максимального диаметра оболочки 20.3 нм.

Шумы в данных влияют на восстановление из-за того, что их низкочастотные составляющие искажают форму кривой рассеяния. Этот эффект может быть уменьшен путем уменьшения шага по  $s$ , т.е. увеличения числа экспериментальных точек в шенноновском канале, однако не может быть полностью устранен.

Данный пример говорит о том, что при измерениях рассеяния желательно устанавливать частый угловой шаг отсчетов для улучшения статистики и возможности распознавания низкочастотной составляющей шума измерений.

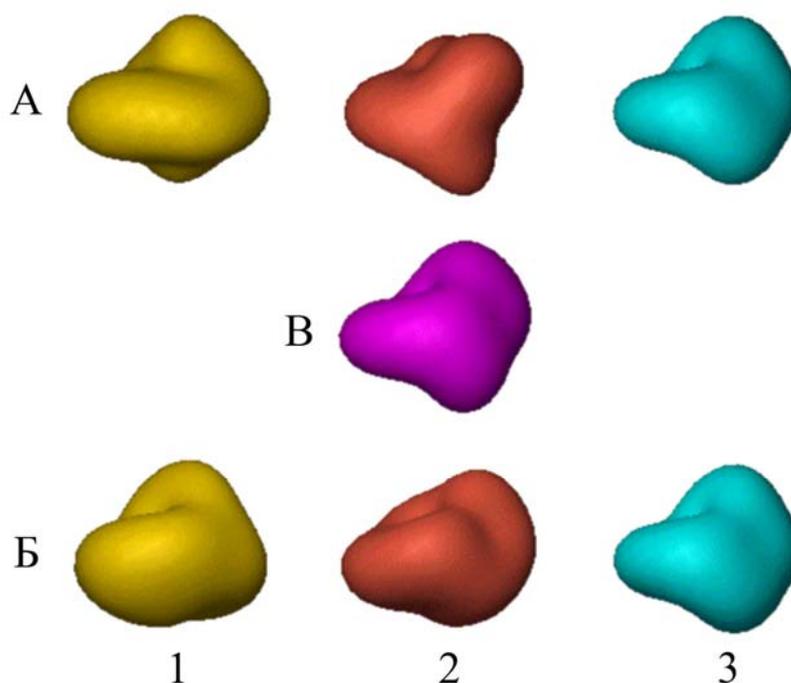


Рисунок 2.10 – Восстановление формы частицы В, представленной до  $L = 4$ , по разным участкам кривой интенсивности МУР. Столбцы 1 - 3: модели, восстановленные по диапазонам 1 - 3, соответственно. В центре показана точная модель. Строка А: старт со сферической формы, строка Б - старт с произвольной формы.

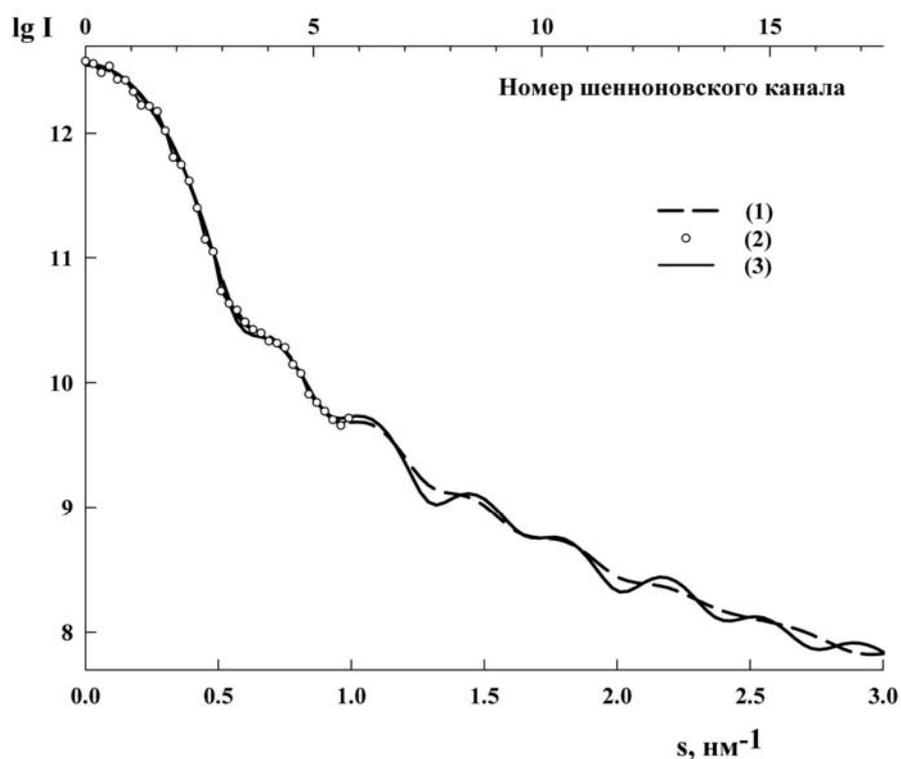


Рисунок 2.11 – Данные рассеяния для восстановления формы с  $L = 3$  при 10% случайном шуме в исходных данных. (1): исходная кривая интенсивности. (2): кривая (1) с 10% псевдослучайным шумом (взята в качестве экспериментальных данных). (3): рассеяние от найденной модели III Рисунок 2.12, столбец 0.

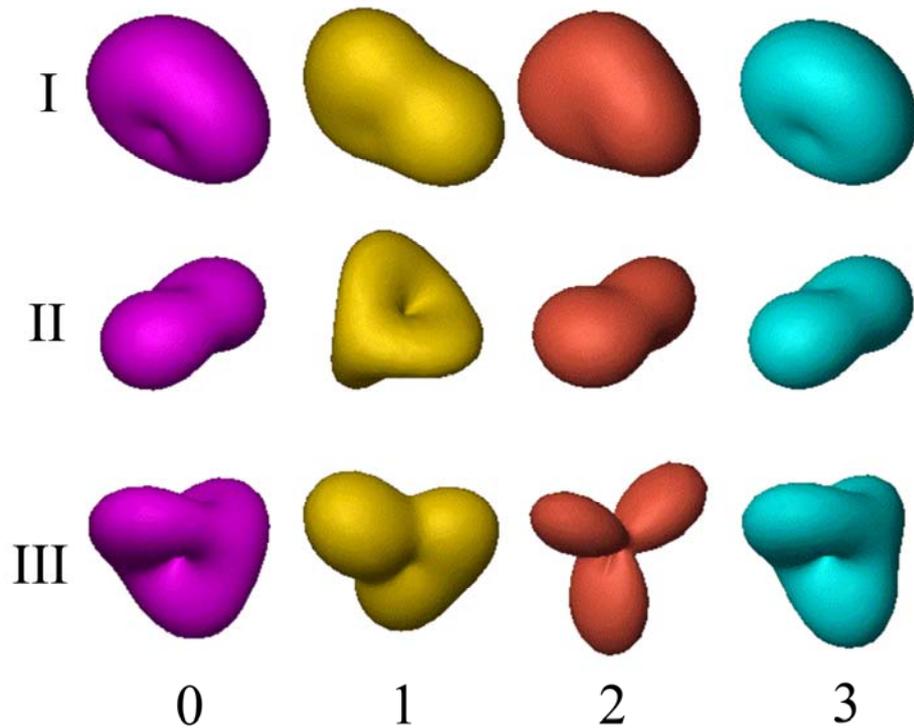


Рисунок 2.12 – Восстановление форм I - III (Рисунок 2.5) с  $L = 3$  при 10% случайном шуме в исходных данных. Столбец 0: исходные тела. Столбцы 1 - 3 соответствуют восстановлениям по диапазонам интенсивности 2 - 4 Рисунка 2.7. Диапазон 1 не брали для расчетов из-за полной непригодности в силу малого числа отсчетов.

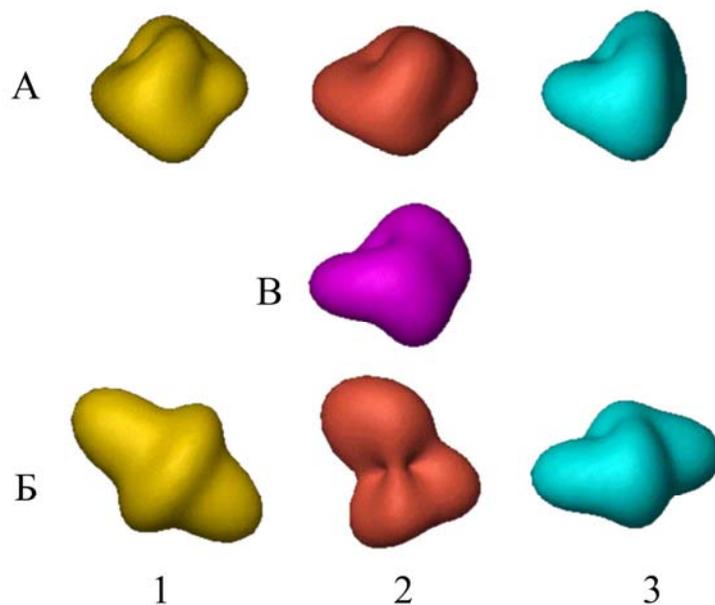


Рисунок 2.13 – Восстановление формы частицы В, представленной до  $L = 4$ , по разным участкам кривой интенсивности МУР при 10% случайном шуме в исходных данных. Столбцы 1 - 3: модели, восстановленные по диапазонам 1 - 3, соответственно. В центре показана точная модель. Строка А: старт со сферической формы, строка Б - старт с произвольной формы.

### Эксперимент 4

В реальных экспериментах по измерению рассеяния минимальный угол ограничен размерами сечения пучка излучения, который перекрывается перед детектором (Рисунок 2.1). Для моделирования реальных условий эксперимента теоретическую кривую рассеяния от модельного тела начинали с  $s = 0.1 \text{ нм}^{-1}$  и добавляли 3%-ый относительный псевдослучайный шум. Кроме того, к данным рассеяния была добавлена константа, равная  $10^{-4}$  от интенсивности в нулевой угол, которая моделировала ошибку вычитания фона рассеяния.

На Рисунках 2.14 и 2.15 показаны результаты эксперимента. Для тел с  $L = 3$  введенные в данные рассеяния искажения практически не повлияли на результаты восстановления, при  $L = 4$  искажения форм заметны, но R-фактор не превышал 10.4% (Таблица 2.1).

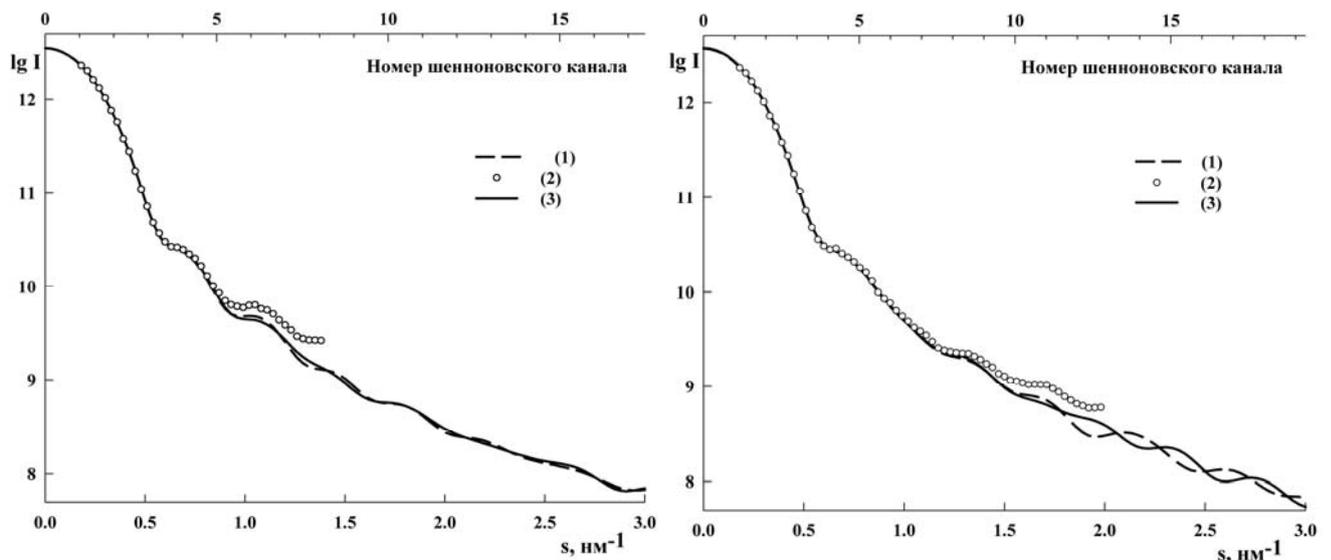


Рисунок 2.14 – Восстановление форм частиц в условиях, приближенном к экспериментальным. Слева - моделирование тела с  $L = 3$  (тело I на Рисунке 2.5). (1): идеальная кривая рассеяния. (2): кривая с шумом и добавленной константой, используемая в качестве экспериментальных данных. (3): рассеяние от восстановленной модели (Рисунок 2.15, верхняя строка, столбец 1). Справа - те же обозначения для тела B с  $L = 4$  (Рисунок 2.15, нижняя строка, столбец 1).

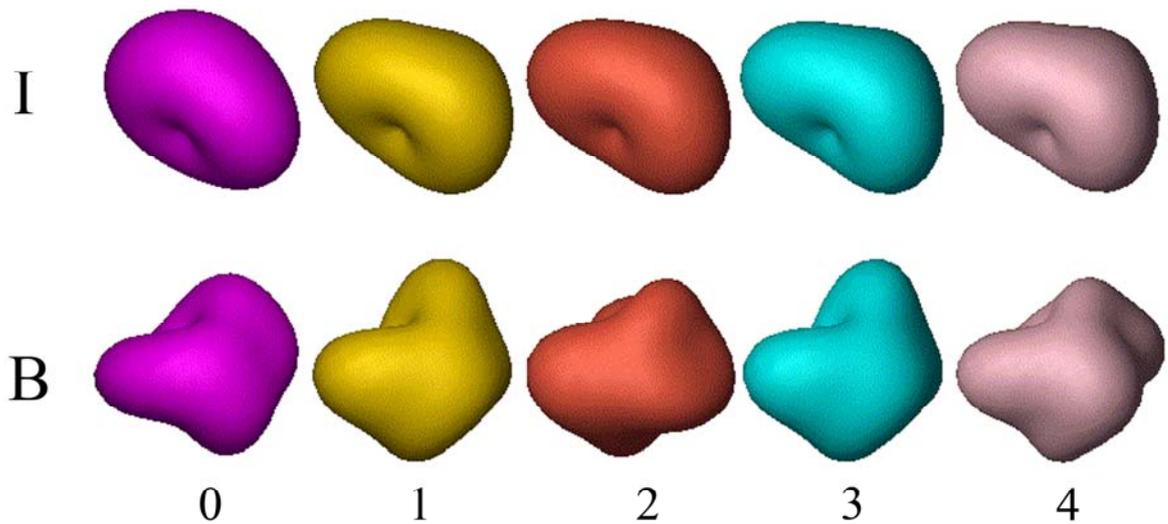


Рисунок 2.15 – Результаты моделирования форм частиц в условиях, приближенным к экспериментальным (Рисунок 2.14). Столбец 0: точные модели. Столбцы 1 - 4: результаты восстановления при разных стартовых приближениях.

### Эксперимент 5.

Большое число параметров модели (например,  $L = 5$ ) уже приводит к неоднозначным решениям. Это продемонстрировано на Рисунках 2.16 и 2.17, которые показывают, что рассеяние от трех разных в деталях форм практически совпадает в рабочем диапазоне (показан стрелкой), который использовали для их восстановления, при числе шенноновских каналов 12. Число независимых параметров моделей равно  $(L+1)^2 - 6 = 30$ .

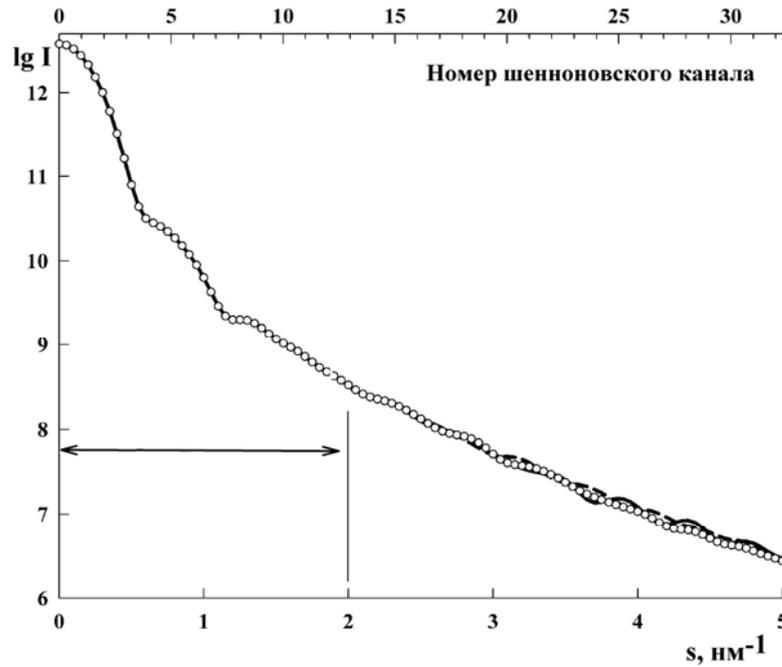


Рисунок 2.16 – Демонстрация неоднозначности данных рассеяния от разных тел с  $L = 5$  (30 независимых параметров, 13 шенноновских каналов).  $R$ -факторы приведены в Таблице 2.1. Оболочки тел показаны на Рисунке 2.17. Кривые рассеяния практически неразличимы в области, показанной стрелкой.

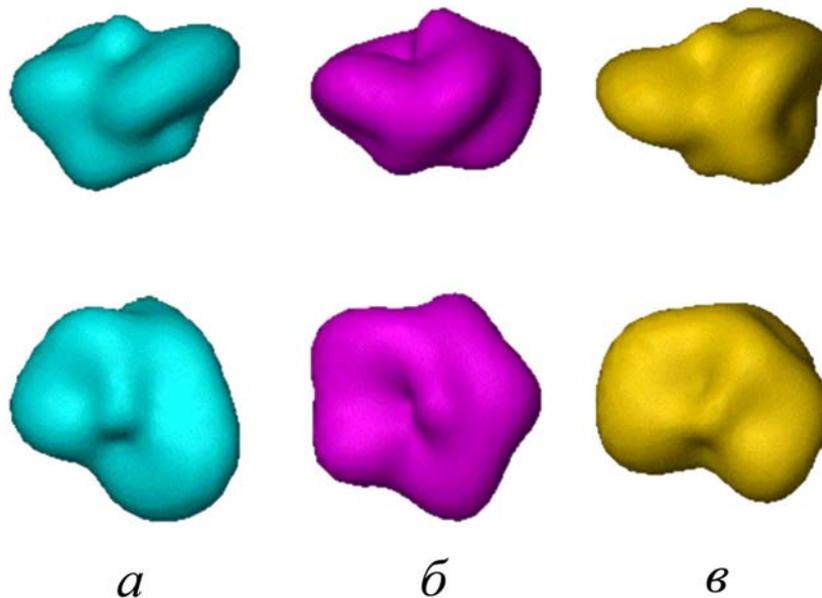


Рисунок 2.17 – Пример форм тел  $a$ ,  $б$  и  $в$ , обладающих практически идентичными кривыми рассеяния в диапазоне 13-ти шенноновских каналов (Рисунок 2.16).  
Верхняя строка: тела повернуты на  $90^\circ$  вокруг горизонтальной оси.

В целом, проведенные модельные эксперименты позволяют утверждать, что при  $L \leq 1.5N_S$  получаемое решение однозначно с приемлемой на практике точностью (с точностью до энантиоморфного преобразования частицы, не меняющего интенсивности рассеяния) в классе моделей, описываемых функцией формы (2.20). Увеличение диапазона данных не должно существенно уменьшать величину вариаций решений, так как за пределами 15-20 шенноновских каналов вклад в рассеяние определяется уже внутренними неоднородностями, на фоне спада кривой рассеяния от формы низкого разрешения по закону Порода [122] который определяется только двумя параметрами, величиной  $s$  и показателем степени  $p$  в  $s^p$ . Рассмотренный подход применим только для оценки общего вида формы частиц, детали формы с разрешением менее  $1/3 - 1/4$  от ее максимального диаметра оказываются ненадежными. Влияние диапазона данных на точность восстановления наглядно демонстрируется Рисунком 2.18

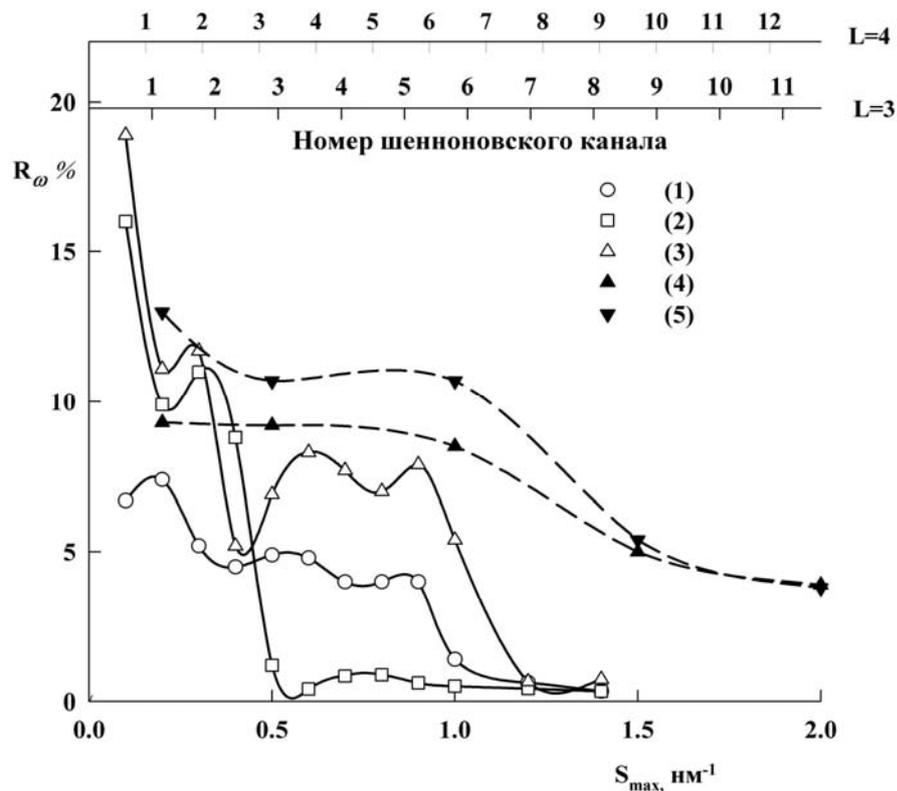


Рисунок 2.18 – Зависимость R-фактора для восстановленных тел от углового диапазона данных рассеяния. (1)-(3): для тел I - III с  $L = 3$  Рисунка 2.5. (4)-(5): для тела В с  $L = 4$  Рисунка 2.10, при двух стартах с различных начальных аппроксимаций. Положения значений шенноновских каналов различаются для двух групп тел в силу их различного максимального диаметра.

Привлечение дополнительной информации о задаче, например, наличие оси симметрии, может быть легко учтено в задаче или наложением дополнительных штрафов, или введением оси симметрии, которая значительно уменьшает число независимых параметров, отбрасывая гармоники, не обладающие такой симметрией. В последнем случае величину  $L$  можно увеличивать и, тем самым, повышать разрешение устойчивой модели.

Рассмотренные численные эксперименты были еще ограничены тем, что идеальное рассеяния от модельных тел рассчитывали по ограниченному ряду гармоник. На практике, конечно, такого ограничения не существует и это следует принимать во внимание при анализе экспериментальных данных, уменьшая доверие к относительно мелким деталям формы.

Недостатком данного метода является невозможность описывать отверстия и глубокие впадины на поверхности, так как прямая, выходящая из начала координат, должна пересекать параметризованную поверхность только один раз. Этому недостатка лишен метод моделирования малыми объемными элементами (шариками), который будет рассмотрен далее.

С целью повышения устойчивости результата алгоритм восстановления формы был модифицирован с учетом эффекта обрыва ряда (2.20). Сначала определяется трехосный эллипсоид с полуосями  $A, B, C$ , модельная интенсивность рассеяния от которого  $I_E(s)$

$$I_E(s_i) = \int_0^1 \int_0^1 \left\{ s_i \left[ A^2 \cos^2\left(\frac{1}{2}\pi x\right) + B^2 \sin^2\left(\frac{1}{2}\pi x\right)(1-y^2) + C^2 y^2 \right]^{1/2} \right\} \cdot dx \cdot dy$$

наилучшим образом аппроксимирует начальный отрезок экспериментальной кривой интенсивности рассеяния (соответствующий  $N_S = 2 - 3$ ). Разложив функцию формы найденного эллипсоида в ряд по сферическим гармоникам и удержав слагаемые с  $l \leq L$ , получают форму  $E_L$ . Положив  $w(s) = I_{E_L}(s) / I_E(s)$ , подают на вход алгоритма нелинейной минимизации взвешенные экспериментальные значения  $w(s_i)J(s_i)$ . Таким образом удается ослабить вклад коэффициентов высокочастотных сферических гармоник во входных данных. Модельные испытания модифицированного алгоритма по методу

“мультизапуска” показали, что он обеспечивает стабильное восстановление формы частиц при низком разрешении, не зависящее от стартового приближения.

Программа определения формы большинство необходимых установок делает сама, но пользователь может, например, уменьшить пространственное разрешение, установив в процессе диалога с программой значение  $L$ , равное 3, а не 4. Это приведет к значительному улучшению стабильности решения - формы частицы, но мелкие детали структуры, видимые в решении при  $L = 4$ , потеряются.

Со времени своего создания программа SASHA была применена исследователями во всем мире для оценки формы многих десятков, если не сотен белковых молекул и наночастиц другой природы. Программа входит в состав пакета анализа данных малоуглового рассеяния ATSAS [A39], открытый для свободного использования (Data analysis software ATSAS, <http://www.embl-hamburg.de/biosaxs/software.html>).

Изображения оболочек тел на рисунках были получены с помощью системы трехмерной графики ASSA [A15]. ASSA позволяет отображать и манипулировать одновременно 50 объектами, представленными как функциями оболочек (каркасная оболочка, затененная поверхность с различной степенью прозрачности), так и совокупностями атомов (или цепью  $C_{\alpha}$ -атомов). Кроме того, ASSA способна вызывать программы анализа данных, в частности, программу расчета интенсивности рассеяния отдельным объектом, а также программу определения формы. В последнем случае осуществляется поиск формы в реальном времени: в процессе уточнения формы последовательные приближения отображаются на экране в виде трехмерных тел, и одновременно показывается график согласия между экспериментальной и расчетной кривой. Каждая программа работает как независимый процесс, обмен информацией происходит через специально организованные информационные каналы. Пользователь в любой момент имеет возможность остановить процесс восстановления формы, проанализировать получающееся на данный момент решение, поменять параметры и либо продолжить, либо прекратить работу программы. Таким образом удастся визуализировать и сделать контролируемым процесс автоматического определения формы, обычно скрытый от глаз исследователя.

### 2.2.2.2 Пример определения форм белковых молекул в растворе по данным малоуглового рентгеновского рассеяния

Практическая значимость разработанного алгоритма была проверена на восстановлении по реальным экспериментальным данным формы ряда белков с известной структурой в кристалле (FKBP25mem [A7], лизоцим, гексокиназа, рибонуклеотид редуктаза, обратная транскриптаза, тиоредоксин редуктаза, миозин S1, энопирувил трансфераза и много других). Результаты модельных расчетов для большого количества белков были использованы при разработке метода и программы классификации белковых структур по данным рассеяния [A30]. Рисунки 2.19 и 2.20 иллюстрируют определение формы двух белков: гексокиназы и обратной транскриптазы вируса иммунодефицита человека в растворе и дают представление о разрешении, которое достигается прямым восстановлением формы по данным МУР [A22]. Молекулярные веса этих частиц: 52 и 105 КДа, соответственно. Измерения проводились на установке X-33 (EMBL, синхротрон DESY, Гамбург, Германия). На Рисунке 2.19 точками показаны экспериментальные данные, сплошными линиями - восстановленные кривые интенсивности рассеяния. Модели низкого разрешения в обоих случаях вычислялись при  $L=4$  (19 независимых параметров). На Рисунке 2.20 они изображены в виде прозрачных оболочек. Для сравнения показаны также атомные модели частиц [161].

Другим примером служит определение формы макромолекул белка SecA и конформационной гибкости димера в растворе. Димерные молекулы SecA являются удлиненными образованиями, 15 нм в длину и 8 нм в поперечном сечении. Поэтому длина димера примерно в 4 раза превышает толщину мембраны, в которую он встраивается. Присоединение к поверхности SecA макромолекулы АТФ или АДФ не изменяет радиус инерции. Мутант SecA, который катализирует несколько циклов гидролиза АТФ не претерпевает серьезных конформационных изменений, которые обнаруживало бы малоугловое рассеяние.

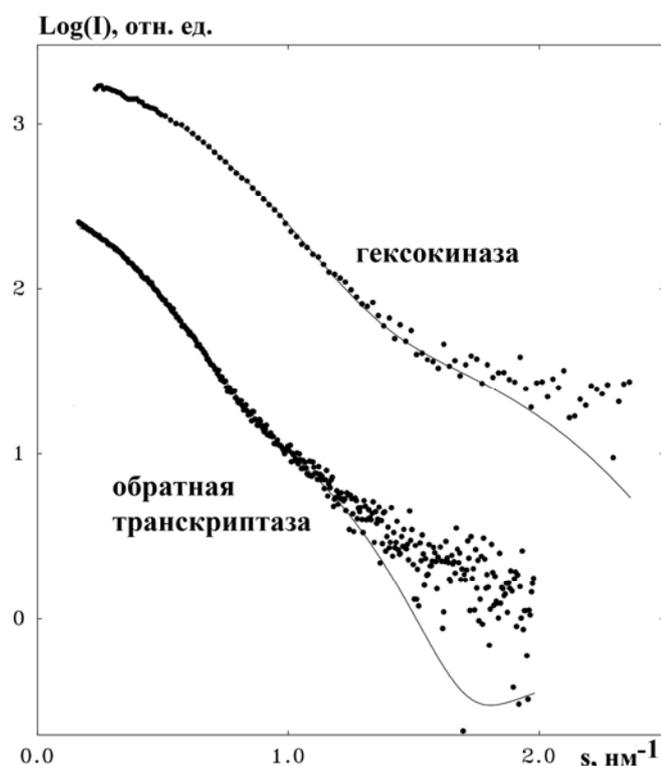


Рисунок 2.19 – Интерпретация данных МУР растворами гексокиназы (верхняя кривая) и обратной транскриптазы вируса HIV-1 (нижняя кривая). Экспериментальные данные изображены точками.

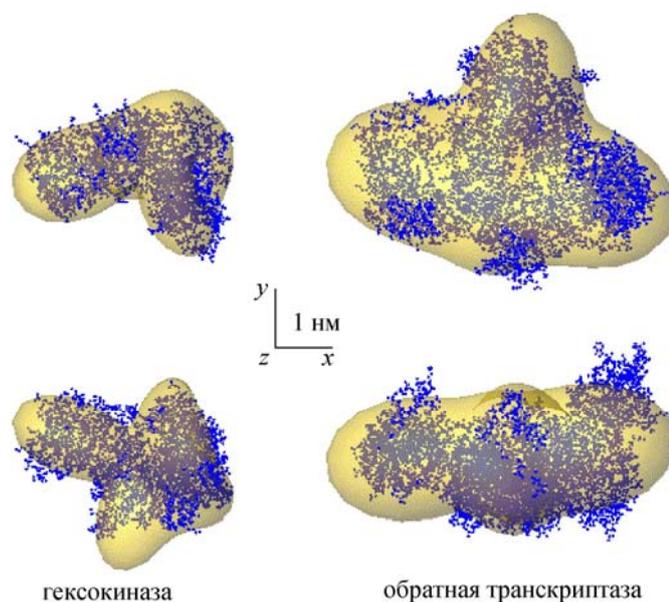


Рисунок 2.20 – Модели низкого разрешения (прозрачные оболочки) и атомные модели гексокиназы (слева) и обратной транскриптазы (справа). Верхний ряд: структуры, повернутые на  $90^\circ$  вокруг горизонтальной оси. Кристаллические атомные модели показаны точками.

Этот вывод был сделан на основании моделирования формы мономера и димера в растворе с помощью рассмотренного выше подхода. Более подробно

детали биохимического функционирования данного белка и условия приготовления образцов представлены в [A19] и не будут предметом рассмотрения в данном разделе. Представим здесь только финальный результат структурных исследований, сделанный на основе анализа данных МУР.

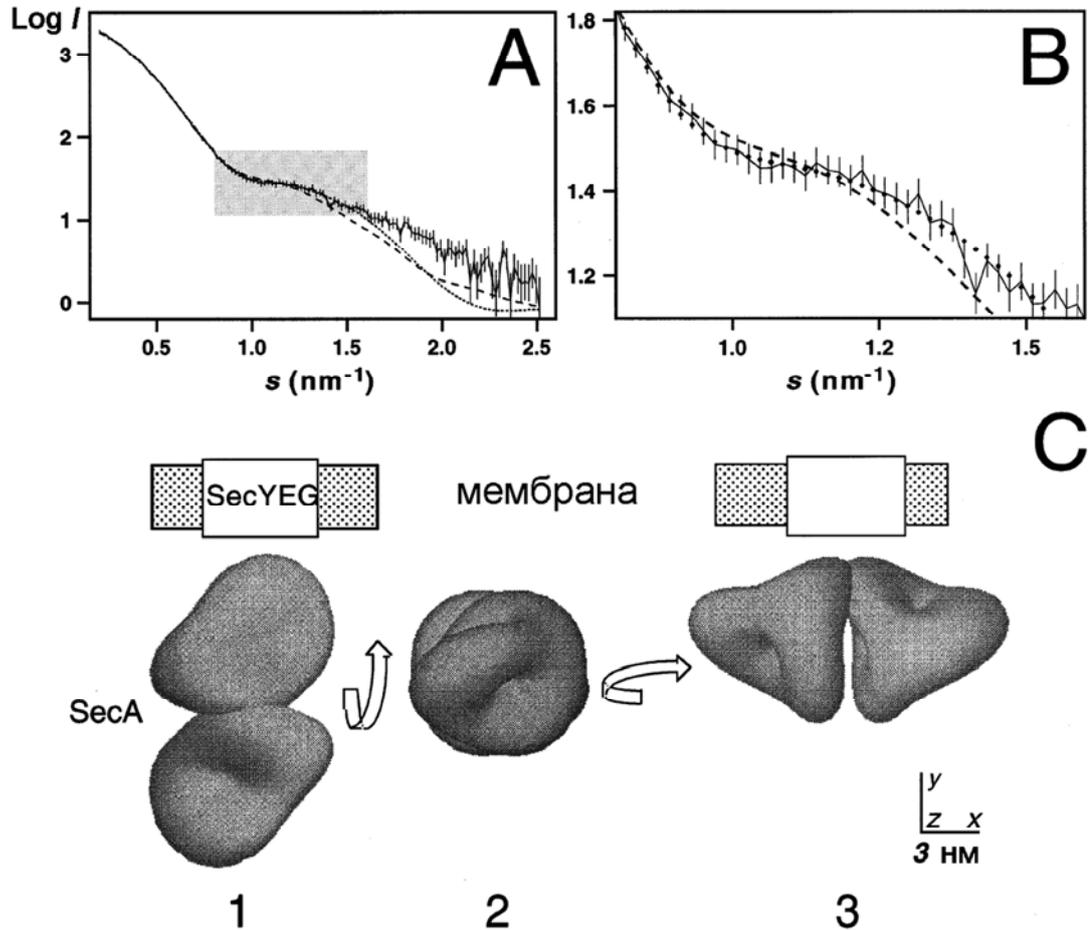


Рисунок 2.21 - Определение формы SecA в растворе. А: экспериментальная и модельные кривые МУР. Пунктир: рассеяние от модели димера, представленного одной формой при  $L = 4$ ; точки: рассеяние от модели димера, составленного из 2-х мономеров, форма которых одинакова и найдена при  $L=3$  (тело 1 внизу). Составная модель точнее передает рассеяние в средних углах, увеличенный средний участок (помечен серой областью на рисунке А) показан на рисунке В. С: Полученная модель низкого разрешения формы молекулы SecA, состоящую из двух мономеров. Модель 2 - повернута относительно модели 1 на  $90^\circ$  вокруг горизонтальной оси, и 3 - вокруг вертикальной. Ориентация SecA относительно к плоскости мембраны является гипотетической. SecA также могут быть расположены параллельно мембране, если форма 3 вращается вокруг Z на  $180^\circ$  (не показано). Для сравнения относительных размеров, схематически представлен липидный бислои мембраны (заштрихованный прямоугольник, толщина 3.4 - 4 нм [162]), охватывающей гидрофобное ядро ок. 3 нм) и SecYEG тример (белый прямоугольник).

### 2.2.2.3 Построение многокомпонентной оболочечной структуры рибосомы 70S *E.coli* в растворе по данным нейтронного малоуглового рассеяния

Методы МУР позволяют получить структурные параметры биополимеров с большой молекулярной массой и их комплексов, к которым относятся субъединицы 50S и 30S рибосомы, состоящие из РНК ядер и связанных с ними белков. Другим структурным методом исследования белковых молекул с большим молекулярным весом является электронная микроскопия, однако, как уже отмечалось в начале главы, она требует специальной подготовки образцов и не позволяет различать близкие по электронной плотности области занятые белками и РНК.

Интенсивность рассеяния от частицы и ее компонент пропорциональна квадрату контраста (разности между рассеивающей плотностью частицы и плотностью растворителя). Следовательно, если плотности компонент частицы отличаются, их относительный контраст можно менять, варьируя плотность растворителя и, тем самым, изменять их относительный вклад в суммарную интенсивность рассеяния. Для нейтронного рассеяния плотность отдельных компонент и растворителя можно менять путем замещения протонов в биополимерах на ядра дейтерия и меняя долю тяжелой воды в растворе. Набор кривых нейтронного рассеяния, соответствующих различным контрастам представляет собой более богатый информацией экспериментальный материал и позволяет селективно моделировать структуру компонент частицы. В данной работе метод вариации контраста применен к исследованию молекулы рибосомы *Escherichia coli* 70S.

Данная работа выполнялась большим коллективом сотрудников из нескольких организаций. Математическое обеспечение для интерпретации данных рассеяния было создано в Европейской лаборатории молекулярной биологии под руководством Д.И. Свергуна, которым был разработан и алгоритм анализа данных рассеяния, рассматриваемый в данном разделе. Вклад автора в

данную работу состоял в применении техники сингулярного анализа аддитивных кривых, рассмотренной в разделе 1, для фильтрации и коррекции экспериментального набора интенсивностей рассеяния, а также в проведении некоторых расчетов.

Рибосомы являются большими по массе комплексами РНК и нескольких десятков белков, организованными в две субъединицы неравного размера. Прокариотическая 70S рибосома кишечной палочки состоит из 57 макромолекулярных компонент, 54 белков и трех рРНК, при общей молекулярной массе около  $2,3 \cdot 10^6$  Da [162]. В обеих субъединицах РНК-остатки составляют около двух третей общей массы.

К моменту начала наших исследований по структуре рибосомы 70S было опубликовано большое количество работ. В качестве стартовых моделей мы использовали электронномикроскопические модели, как наиболее соответствующие данным МУР, хотя и не вполне их воспроизводящие.

Замена водорода на дейтерий позволяет контрастировать определенные области в биомолекулах в случае рассеяния нейтронов. Водород рассеивает нейтроны в основном некогерентно, создавая слабо зависящий от угла фон рассеяния, который можно вычесть из данных измерений, тогда как ядра дейтерия обладают и большим сечением рассеяния и рассеивают нейтроны когерентно. Следовательно, дейтерированная часть молекулы будет обладать значительно большей амплитудой рассеяния по сравнению с протонированной. Для настоящего исследования были приготовлены две группы выборочно дейтерированных образцов 70S (Таблица 2.2). Первая группа состояла из сборок полностью протонированных и / или дейтерированном рибосомных субъединиц, то есть из полностью дейтерированной и полностью протонированной 70S частиц (НН30+НН50) и (DD30+DD50), а также 70S состоящих из одной дейтерированной и одной протонированной субъединицы, соответственно (DD30+НН50, НН30+DD50). В обозначениях цифры означают коэффициент седиментации частицы и отвечают ее названию - 30S и 50S. Первая буква означает

протонирование (H) или дейтерирование (D) белка, вторая, по аналогии - РНК. Вторым набором содержались субчастицы с перекрестным протонированием / дейтерированием белков и РНК. Подготовка таких образцов проходила в два этапа. Сначала гибридные субчастицы синтезировали из перекрестно дейтерированных и протонированных белков и РНК, затем полную частицу 70S собирали из них и из нативных протонированных субъединиц, получив в результате DH30+HH50, HH30+DH50, и DH30+DH50. Все образцы были проверены на гомогенность (ультрацентрифугирование в градиенте сахарозы и гель-хроматография), структурную целостность и биологическую активность. Во всех случаях степень гомогенности была  $> 90\%$ . Целостность молекулы рРНК контролировали методом одномерного гель-электрофореза. Результаты измерения биологической активности образцов, т. е. активности связывания тРНК, транслокационная эффективность и пептидилтрансферазная активность были исследованы биохимически. 70S, собранные из нативных субчастиц, имели примерно такую же высокую активность, как и гибридные и не хуже 75% от активности природной 70S. Препаративная и биохимическая часть исследований была выполнена группой профессора К.Х. Нирхауса в (Max Planck Institut für Molekulare Genetik, AG Ribosomen, Ihnestr. 73, 14195 Berlin, Germany), измерения нейтронного рассеяния - группами проф. Г. Штурмана в (GKSS Research Centre, GKSS-WS, D-21502 Geesthacht, Germany) и Я.С. Педерсена в (Risø National Laboratory, DK-4000 Roskilde, Denmark). Подробно все детали эксперимента рассмотрены в [A11, A12].

Интегральные параметры целой рибосомы 70S, ее частей 30S и 50S, отдельных белков и РНК-фрагментов были оценены по инвариантам данных МУР (2.3 - 2.6). Эти данные сравнивали с результатами электронномикроскопических реконструкций 70S-рибосомы, которые были предоставлены группами Дж. Франка (Албания) и М. Ван Хила с П. Бриакombe (Берлин) [163,164]. Интегральные параметры и теоретические кривые рассеяния моделей 70S и субчастиц сравнивали с экспериментальными данными. Было найдено, что модель Франка [163] обеспечивает лучшее согласие с данными рассеяния. 50S

субъединица Ван Хила и модель Бриакombe значительно лучше соответствуют данным рассеяния, если заполнить поры в предложенных ими структурах. Этот факт послужил толчком к моделированию частицы однородными фазами - белковой и РНК.

Таблица 2.2. Состав и контрасты исследованных образцов частиц рибосомы 70S.

Частица	Растворитель	Белки	16S rRNA	Белки	23S/5S rRNA
		TP30		TP50	
HH30+HH50	0%D <sub>2</sub> O	1.82	2.27	1.79	2.27
HH30+HH50	35%D <sub>2</sub> O	0.327	1.18	0.293	1.18
HH30+HH50	50%D <sub>2</sub> O	-0.312	0.710	-0.346	0.710
HH30+HH50	75%D <sub>2</sub> O	-1.38	-0.069	-1.41	-0.069
HH30+HH50	100%D <sub>2</sub> O	-2.44	-0.848	-2.48	-0.848
HH30+DD50	0%D <sub>2</sub> O	1.82	2.27	5.57	3.82
HH30+DD50	35%D <sub>2</sub> O	0.327	1.18	4.07	2.72
HH30+DD50	50%D <sub>2</sub> O	-0.312	0.710	3.43	2.26
HH30+DD50	75%D <sub>2</sub> O	-1.38	-0.069	2.37	1.48
DD30+HH50	0%D <sub>2</sub> O	5.52	3.82	1.79	2.27
DD30+HH50	35%D <sub>2</sub> O	4.03	2.72	0.293	1.18
DD30+HH50	50%D <sub>2</sub> O	3.39	2.26	-0.346	0.710
DD30+HH50	75%D <sub>2</sub> O	2.33	1.48	-1.41	-0.069
DD30+HH50	100%D <sub>2</sub> O	1.26	0.700	-2.48	-0.848
DD30+DD50	0%D <sub>2</sub> O	5.52	3.82	5.57	3.82
DD30+DD50	35%D <sub>2</sub> O	4.03	2.72	4.07	2.72
DD30+DD50	50%D <sub>2</sub> O	3.39	2.26	3.43	2.26
DD30+DD50	75%D <sub>2</sub> O	2.33	1.48	2.37	1.48
DD30+DD50	100%D <sub>2</sub> O	1.26	0.700	1.30	0.700
DH30+HH50	0%D <sub>2</sub> O	5.52	2.27	1.79	2.27
DH30+HH50	40%D <sub>2</sub> O	3.82	1.02	0.086	1.02
DH30+HH50	60%D <sub>2</sub> O	2.96	0.398	-0.766	0.398
DH30+HH50	100%D <sub>2</sub> O	1.26	-0.848	-2.48	-0.848
HH30+DH50	0%D <sub>2</sub> O	1.82	2.27	5.57	2.27
HH30+DH50	40%D <sub>2</sub> O	0.116	1.02	3.87	1.02
HH30+DH50	60%D <sub>2</sub> O	-0.736	0.398	3.01	0.398
HH30+DH50	100%D <sub>2</sub> O	-2.44	-0.848	1.30	-0.848
Отдельные субъединицы в растворе					
HH30	0%D <sub>2</sub> O	1.82	2.27	0	0
HH30	100%D <sub>2</sub> O	-2.44	-0.848	0	0
HH50	0%D <sub>2</sub> O	0	0	1.79	2.27

HH50	100%D <sub>2</sub> O	0	0	-2.48	-0.848
DD30	0%D <sub>2</sub> O	5.52	3.82	0	0
DD50	0%D <sub>2</sub> O	0	0	5.57	3.82
Спин-зависимые данные рассеяния					
HH30+DD50	P=0	-3.24	-1.39	0.41	0.11
HH30+DD50	P=1	1.98	0.79	0.57	0.19
DD30+HH50	P=0	0.37	0.11	-3.28	-1.39
DD30+HH50	P=1	0.53	0.19	2.05	0.79
DH30+DH50	P=0	0.37	-1.39	0.41	-1.39
DH30+DH50	P=1	0.53	0.79	0.57	0.79
Данные рентгеновского МУР					
70S	0%D <sub>2</sub> O	1.16	1.82	1.16	1.82
30S	0%D <sub>2</sub> O	1.16	1.82	0	0
50S	0%D <sub>2</sub> O	0	0	1.16	1.82

Уровень дейтерирования белковых частей молекул были определены методом масс-спектрометрии и полученные значения использованы для расчета рассеивающего контраста для каждого образца индивидуально. Для РНК принимали 100% дейтерирование, что было проверено измерениями нейтронного рассеяния. В результате были определены следующие величины относительных контрастов, указанные в Таблице 2.2 и установлены правила расчета:

$$\begin{aligned}
 \overline{\Delta\rho}_{70S} &= 0.358 \overline{\Delta\rho}_{TP} + 0.642 \overline{\Delta\rho}_{RNA} = 0.369 \overline{\Delta\rho}_{30S} + 0.631 \overline{\Delta\rho}_{50S} \\
 \overline{\Delta\rho}_{30S} &= 0.412 \overline{\Delta\rho}_{TP} + 0.588 \overline{\Delta\rho}_{RNA} \\
 \overline{\Delta\rho}_{50S} &= 0.327 \overline{\Delta\rho}_{TP} + 0.673 \overline{\Delta\rho}_{RNA}
 \end{aligned}
 \tag{2.50}$$

Эти контрасты представляют собой "сухие контрасты", соответствующие сухому исключенному объему частиц. Реальные контрасты гидратированных частиц легко рассчитываются из коэффициента объемного гидратирования  $N_v$  (v/v) как  $\overline{\Delta\rho}_{wet} = \overline{\Delta\rho}_{dry} / (1+N_v)$ . Контраст частицы определяется формулой

$$\overline{\Delta\rho} = \int_V [\rho(\mathbf{r}) - \rho_s] \cdot dV = \bar{\rho} - \rho_s V, \tag{2.51}$$

$V$  - объем частицы,  $\rho_s$  - плотность растворителя,  $\rho(\mathbf{r})$  - функция плотности частицы,  $\bar{\rho}$  - ее средняя плотность.

Радиус инерции частицы как функция контраста записывают в виде

$$R_g^2 \cdot \overline{\Delta\rho} = R_c^2 + \frac{\alpha}{\Delta\rho} - \frac{\beta}{\overline{\Delta\rho}^2}, \quad (2.52)$$

где  $R_c$  - радиус инерции при бесконечном контрасте,  $\alpha > 0$ , если плотность оболочки частицы больше плотности ядра и  $< 0$  в противном случае,  $\beta \neq 0$ , если центр массы частицы смещается по мере изменения контраста [165]. Объем частицы (т.н. Породовский объем) определяли по выражению [126]

$$V_P \approx \frac{2\pi^2 \cdot I(0)}{\int_0^{s_{\max}} s^2 \cdot I(s) \cdot ds}, \quad (2.53)$$

который соответствует объему гидратированной частицы в растворе.

Вариации контраста с помощью замены растворителей, содержащих разные доли обычной и тяжелой воды составили основную часть образцов для измерений нейтронного рассеяния. Качество образцов, стабильность молекул в растворе, правильность расчета контрастов контролировали путем расчета структурных инвариантов (2.4 - 2.6). Параллельный ход кривых зависимости интенсивности рассеяния в нулевой угол для всех образцов в зависимости от процентного содержания тяжелой воды в растворителе свидетельствовал о корректности экспериментов (Рисунок 2.22). Линейный характер этих зависимостей показывал, что все частицы структурно аналогичные частицы имеют практически одинаковый инвариантный объем.

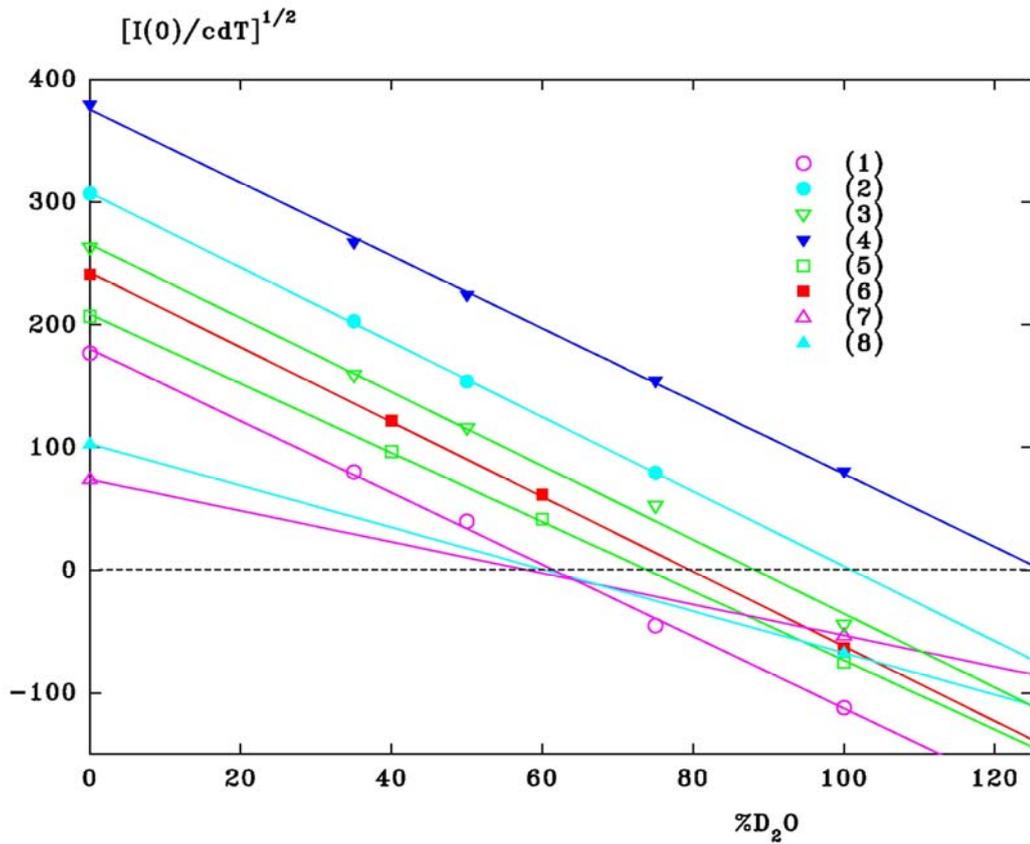


Рисунок 2.22 – Нормализованная интенсивность рассеяния в нулевой угол от гибридных образцов рибосомы и ее субчастиц в зависимости от концентрации тяжелой воды в растворителе. Регрессии (1 до 8) соответствуют последовательности образцов, приведенной в Таблице 2.3 (образцы с 1 по 8 сверху вниз). Ошибки в  $I(0)$  меньше, чем размер символов.  $d$  - толщина образца,  $T$  - пропускание.

Таблица 2.3. Основные параметры образцов рибосомы и ее субъединиц.

Наименование образца	$\overline{\Delta\rho}_{\text{calc}}$ %D <sub>2</sub> O	$\overline{\Delta\rho}_{\text{exp}}$ %D <sub>2</sub> O	$\overline{\Delta\rho}/\overline{\Delta\rho}_{\text{HH}}$ in H <sub>2</sub> O	$(I_0/I_{0\text{HH}})^{1/2}$ in H <sub>2</sub> O	Мол. вес, 10 <sup>6</sup> Da
HH30+HH50	59.9	60.5	-	-	2.07
HH30+DD50	100.8	101	1.68	1.73	2.22
DD30+HH50	84.8	87	1.42	1.49	2.20
DD30+DD50	125.8	126	2.10	2.14	2.20
DH30+HH50	73.5	74	1.26	1.17	2.03
HH30+DH50	79.0	80	1.32	1.36	2.09
HH30	58.3	58	-	-	0.84
HH50	60.8	61	-	-	1.38
DD30	125.9	-	2.16	2.30	-
DD50	125.7	-	2.07	2.06	-

Найденные по данным рассеяния структурные инварианты представлены в Таблице 2.4.

Полный набор экспериментальных данных МУР представлял собой 42 кривые рассеяния.

Таблица 2.4. Интегральные параметры рибосомы 70S и ее субчастиц, Здесь  $R_c$  - радиус инерции при бесконечном контрасте,  $V$  - исключенный объем,  $D_{\max}$  - максимальный диаметр, определяемый по форме функции парных расстояний  $p(r)$ ,  $R_g^{\text{RNA}}$  - радиус инерции РНК - области,  $R_g^{\text{TP}}$  - белковой части комплекса.

Параметр	Экспериментальное значение	Микроскопическая модель Франка	Модель Штарка
30S			
$R_c$ , nm	$7.2 \pm 0.2$	6.91	7.57
$V$ , nm <sup>3</sup>	$1400 \pm 200$	1300	970
$D_{\max}$ , nm	$24 \pm 2$	23.5	25.3
$R_g^{\text{RNA}}$ , nm	$6.6 \pm 0.2$	6.75	-
$R_g^{\text{TP}}$ , nm	$7.9 \pm 0.3$	7.21	-
50S			
$R_c$ , nm	$7.6 \pm 0.2$	7.40	8.30
$V$ , nm <sup>3</sup>	$2500 \pm 100$	2620	1980
$D_{\max}$ , nm	$25 \pm 1$	24.4	25.1
$R_g^{\text{RNA}}$ , nm	$6.5 \pm 0.1$	6.90	-
$R_g^{\text{TP}}$ , nm	$9.7 \pm 0.5$	8.12	-
70S			
$R_c$ , nm	$9.15 \pm 0.05$	8.69	9.68
$V$ , nm <sup>3</sup>	$4000 \pm 200$	3920	2950
$D_{\max}$ , nm	$27 \pm 2$	26.9	29.3
$R_g^{\text{RNA}}$ , nm	$8.2 \pm 0.2$	8.30	-
$R_g^{\text{TP}}$ , nm	$10.5 \pm 0.5$	9.24	-
$\Delta d(30-50)$ , nm	$10.5 \pm 0.4$	10.0	11.1

Модель частицы рибосомы была построена из 4-х компонент - двух РНК-областей и двух белковых фаз (попарно для субъединиц 30S и 50S). Как было рассмотрено выше, форму каждой компоненты параметризовали разложением в ряд по сферическим гармоникам (2.19, 2.20) и частичные амплитуды рассеяния представляли в виде разложения (2.21) [166]:

$$\begin{aligned}
A_{lm}(s) = (is)^l \cdot \sqrt{2/\pi} \cdot \sum_{p=0}^{p_{\max}} \frac{(-1)^p f_{lm}^{(l+2p+3)} s^{2p}}{2^p p!(l+2p+3)[(2(l+p)+1)]!!} \cdot \\
\left\{ \frac{f_{lm}^{l+2p+3}}{l+2p+3} + \sum_{j=2}^{l+2p+3} \frac{(-\Delta)^{j-1} \cdot f_{lm}^{l+2p+4-j}}{j!(l+2p+4-j)!} - \right. \\
\left. - \sqrt{4\pi} \cdot \frac{\Delta^{l+2p+3} \cdot f_{lm}^{l+2p+4}}{(l+2p+3) \cdot (l+2p+4-j)} \delta_{l0} \right\} , \quad (2.54)
\end{aligned}$$

где  $p_{\max}$  - число членов в степенном ряде и  $f_{lm}$  - мультипольные коэффициенты  $q$ -ой степени функции формы:

$$f_{lm}^{(q)} = \int_w [F(w)]^q Y_{lm}^*(w) d\omega , \quad (2.55)$$

$\delta_{l0} = 1$ , если  $l = 0$  и  $\delta_{l0} = 0$  в других случаях. Процедуры быстрого вычисления частичных амплитуд были разработаны Д.И. Свергуном и Х.Б. Штурманом в [167,168]. Для уменьшения ошибок в теоретической интенсивности центр масс фигуры должен быть как можно ближе к началу координат. Это делает ряд (2.54) значительно быстрее сходящимся. Так как координаты центров масс компонент (белковой и РНК части субъединиц) могут отличаться, это различие необходимо учитывать при вычислении рассеяния от всей частицы (усреднение ведется по всему телесному углу  $\Omega$ ):

$$I_{calc}(s) = \left\langle \left[ \Delta\rho_n \sum_{n=1}^4 \cdot e^{is\mathbf{r}_n} \cdot A_n(\mathbf{s}) \right]^2 \right\rangle_{\Omega} , \quad (2.56)$$

$A_n(\mathbf{s})$  - амплитуда рассеяния от  $n$ -ой компоненты всей частицы, смещенной своим центром масс в начало координат и нормированная к единичному контрасту,  $\Delta\rho_n$  - контраст,  $\mathbf{r}_n$  - позиция центра масс  $n$ -ой компоненты. Сдвиг вдоль оси  $Z$  на  $\Delta d$  приводит к выражению для амплитуд сдвинутой частицы

$$B_{lm}(s) = (-1)^m \sum_{p=0}^{p_{\max}} (i)^p j_p(s\Delta)(2p+1) \quad (2.57)$$

$$\sum_{k=|l-p|}^{l+p} [(2l+1)(2k+1)]^{1/2} \begin{pmatrix} l & p & k \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & k \\ -m & 0 & m \end{pmatrix} \sum_{j=-k}^k A_{lm}(s)$$

$j_p(x)$  - сферические функции Бесселя,  $\begin{pmatrix} l & p & k \\ m & q & t \end{pmatrix}$  -  $3j$  символы Вигнера [168].

Интенсивности рассеяния рассчитывают по (2.13-2.15), подставляя  $B_{lm}$  вместо  $A_{lm}$ .

Уравнение (2.56) приводит к соотношению при  $n=4$  (4 компоненты рибосомы 70S):

$$I(s) = \sum_{n=1}^4 (\Delta\rho_n)^2 I_{nn}(s) + 2 \sum_{n>k} \Delta\rho_n \Delta\rho_k I_{nk}(s), \quad (2.58)$$

где  $n$  и  $k$  - индексы компонент. Тогда  $I_{nn}(s) = \langle A_n^2(\mathbf{s}) \rangle_{\Omega}$  - интенсивность рассеяния от  $n$ -ой компоненты, и  $I_{nk}(s) = \left\langle A_n(\mathbf{s}) \cdot A_k^*(\mathbf{s}) \cdot e^{is \cdot (\mathbf{r}_n - \mathbf{r}_k)} \right\rangle_{\Omega}$  есть перекрестные члены суммы.

Таким образом, с учетом ортогональности сферических гармоник усреднение по телесному углу  $\langle \rangle_{\Omega}$  приводит к аналитическому выражению (знак приблизительного равенства означает, что суммирование ведется до  $L_{\max}$ , а не до бесконечности):

$$I(s) \approx 2\pi^2 \sum_{l=0}^{L_{\max}} \sum_{m=-l}^l \left\{ \sum_{n=1}^4 \left[ \Delta\rho_n \cdot B_{lm}^n(s) \right]^2 + \right. \quad (2.59)$$

$$\left. 2 \sum_{\forall n>k} \Delta\rho_n \cdot B_{lm}^n(s) \cdot \Delta\rho_k \cdot \left[ B_{lm}^k(s) \right]^* \right\}$$

представляющему собой сумму 10 компонент. Этот факт лег в основу применения техники сингулярного анализа матрицы интенсивностей рассеяния, рассмотренного в первой части данной диссертационной работы, с целью выяснения числа независимых компонент в наборе данных рассеяния и

коррекции экспериментальных данных путем приведения их к 10-компонентному набору.

Будем называть в дальнейшем составную часть структуры "компонентой", а слагаемое в (2.59) - "компонентом".

Основным вкладом автора в данную работу было применение техники модельно-независимого анализа многокомпонентных смесей, рассмотренной в разделе 1, к анализу и корректировке набора данных рассеяния от образцов рибосомы. 4 физических компоненты частицы соответствуют рассеянию, интенсивность которого представляет собой линейную комбинацию 10 слагаемых (2.59). Систематические ошибки в исходных данных (основной причиной которых было неточное вычитание фона рассеяния и низкочастотные составляющие случайных шумов измерений) приводят к нарушению аддитивности (2.59), поэтому по набору данных была проведена оценка числа независимых компонентов с помощью методов, рассмотренных в 1-й части. Статистический анализ левых сингулярных векторов матрицы данных (раздел 1.3.5) показал наличие 10 значимых компонентов при уровне значимости не более 0.05. Это число соответствовало условию числа слагаемых в (2.59) для 4-х фазной частицы. На основании этого набор данных был скорректирован.

Для проведения коррекции данные (42 кривые рассеяния) были представлены на одинаковой угловой сетке в виде матрицы  $D$  и получено сингулярное разложение (1.4). Далее матрица данных была восстановлена по первым 10 сингулярным векторам и получена матрица  $D^*$  (1.7). Тем самым была проведена коррекция к 10 компонентам путем минимального возмущения исходного набора (детали подхода были рассмотрены в 1-й части). Величина критерия Дарбина-Ватсона (1.14), рассчитанного по матрице остатков  $F^0$  (1.6) была равна 1.7, что свидетельствовало о наличии автокорреляции в отброшенных остатках с уровнем значимости не более 0.05. Применение критерия Вальда-Вольфовица (1.32) показало, что последовательность точек в остатках является выборкой из последовательности независимых случайных чисел с уровнем

значимости 0.01. Критерий Вальда-Вольфовица, будучи ранговым непараметрическим критерием, практически нечувствителен к малым смещениям в последовательностях. Таким образом, незначительная систематическая составляющая в матрице остатков вероятнее всего имела вид смещения, которое было тем самым скорректировано в восстановленной матрице. Пример контуров первых 15 сингулярных векторов для набора данных рассеяния низкого разрешения (до  $s = 0.6 \text{ нм}^{-1}$ ) приведен на Рисунке 2.23.

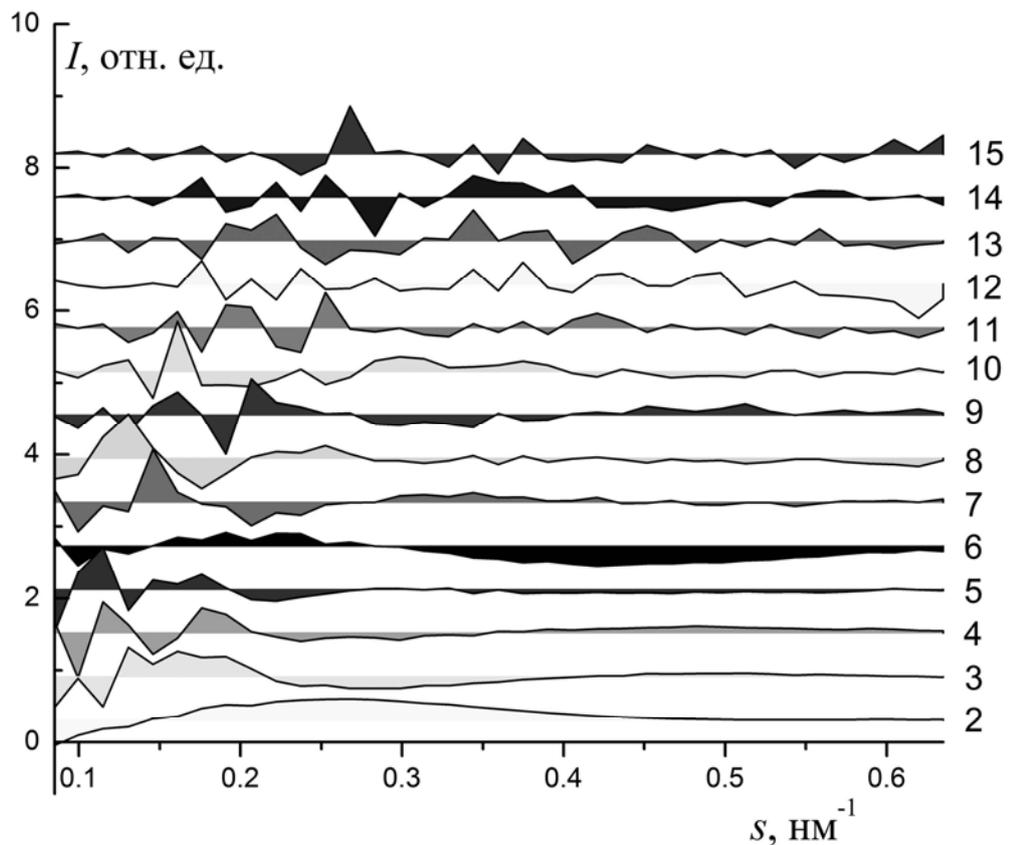


Рисунок 2.23 – Первые 15 сингулярных векторов матрицы данных нейтронного рассеяния, представленные в виде контуров относительно вектора рассеяния. Кривые смещены по вертикали для большей наглядности. Первый сингулярный вектор не показан: он представляет собой усредненный контур кривых рассеяния.

Из рисунка видно, что систематическое поведение контуров заметно до 10 вектора включительно.

Данный пример демонстрирует предложенный автором методологический прием анализа остатков на наличие систематических отклонений путем

комплексного применения разных критериев оценки автокорреляции в элементах сингулярных векторов и матрицы остатков.

Схема модели частицы рибосомы показана на Рисунке 2.24.

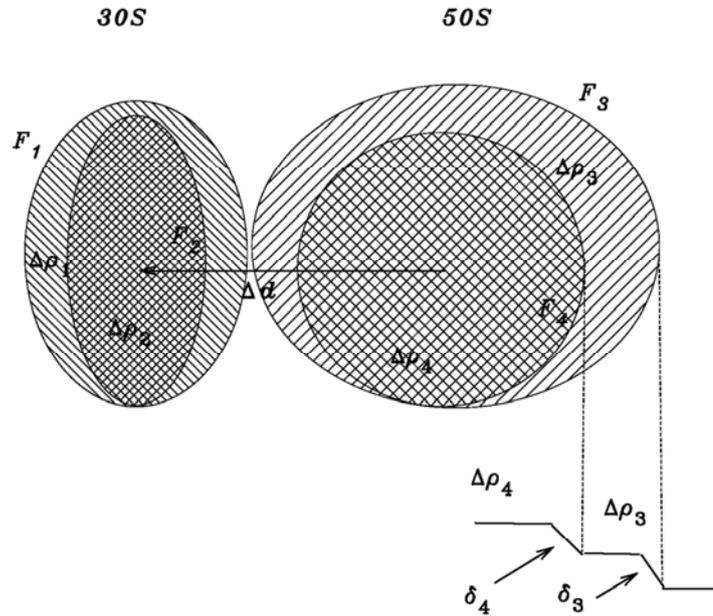


Рисунок 2.24 – Схематическое изображение четырехфазной модели рибосомы.  $F_1$  и  $F_3$ : функции оболочек 30S и 50S субъединиц, соответственно,  $F_2$  и  $F_4$ : оболочки соответствующих фрагментов РНК. Переходные слои:  $\delta_1$  и  $\delta_2$  обеспечивают плавный интерфейс частица-растворитель, слои  $\delta_3$  и  $\delta_4$  служат для описания проникновения белковой фазы в плавную оболочку РНК. позволяют возможно взаимопроникновение белков и рРНК. Две фазы 50S субъединицы (23S/5S рРНК и TP50) сориентированы в начало координат.

Разделение областей фаз белков и РНК было сделано на основании сравнительного анализа литературных данных [163]. В ходе работы такое строение было подтверждено расчетами радиусов инерции по данным малоуглового нейтронного рассеяния от образцов, в которых контраст белковых или РНК частей в среднем был равен 0 (Таблица 2.3).

Вывод, сделанный в разделе 2.2.2.1 о том, что число параметров оболочечной модели формы не должно превышать полуторакратное число шенноновских каналов ( $L_{\max} \leq 1.5N_S$ ) в данных рассеяния, привел к выбору  $L_{\max} = 7$  в формуле (2.59), что соответствует полному числу параметров модели

рибосомы  $N_L = 4[(L+1)^2 + 1] = 260$  (с учетом параметров толщин переходных слоев оболочек) при  $N_S = 170$  во всем наборе кривых рассеяния. Формальное пространственное разрешение модели (2.37) при  $L_{\max} = 7$  составило 3.5 нм.

Перед началом обработки экспериментальных данных была проведена серия численных экспериментов по восстановлению форм оболочек искусственных моделей 4-х фазной частицы 70S, построенных на основании электронномикроскопических моделей. От моделей рассчитывали интенсивности рассеяния, соответствующие образцам с реальными комбинациями дейтерирования компонентов и концентрациями тяжелой воды в растворителе. Восстановления проводили при  $L_{\max} = 5 - 8$ . Было найдено, что  $L_{\max} = 8$  является избыточным, так как не приводит к заметному улучшению качества восстановления исходных данных рассеяния. Разрешение  $L_{\max} = 5$  не позволило описать моделью такого низкого разрешения данные рассеяния с удовлетворительной точностью. Поиск формы компонентов частицы проводили с помощью программы MONSTER, разработанной Д.И. Свергуном.

В качестве начального приближения для внешних форм субъединиц были взяты оболочки электронномикроскопической модели [163], а rРНК области были первоначально аппроксимированы сферами. Далее построенная модель была уточнена методом нелинейных наименьших квадратов (2.31-2.36 с 2.58-2.59) по алгоритму, аналогичному описанному в разделе 2.2.2. Найденные формы внешних оболочек субъединиц 30S и 50S отличались примерно на 10% от начального приближения (критерием служил R-фактор (2.48)), а форма окончательной модели 70S лежала между оболочками моделей [163,164].

Эта модель была оптимизирована с помощью программы MONSTER в несколько этапов минимизации функционала (2.39). На каждом из циклов минимизации осуществляли поиск параметров только двух функций оболочек: либо субъединиц (белковых F1 и F3), либо rРНК (F2 и F4). Это было связано со слишком плохим (большим) числом обусловленности задачи при всех свободных переменных модели, которое превышало  $10^7$  (уравнение 1.52), что приводило к

неустойчивости решений. Более того, на начальных этапах поиска с целью улучшения обусловленности использовали меньшее пространственное разрешение при  $L_{max} = 5$ . Во время нескольких первых циклов использовали только данные рассеяния низкого разрешения до  $s = 0.64 \text{ nm}^{-1}$ , измеренные при расстоянии образец-детектор 4 м. Для уменьшения возможного разброса решений к целевой функции был добавлен дополнительный штрафной член (2.48) за отклонение решения от стартового приближения. На более поздних этапах вес этого штрафа постепенно снижался до нуля, к набору данных рассеяния добавляли рентгеновские и нейтронные данные рассеяния высокого разрешения, снятые до  $s = 2.07 \text{ nm}^{-1}$  (измеренные при расстоянии образец-детектор 1 м) и разрешение модели увеличивали до  $L_{max} = 5$ .

Все полученные решения были довольно похожи: средний квадрат отклонения между соответствующими оболочками (2.48) не превышал 5%. На Рисунке 2.25 показаны набор скорректированных экспериментальных данных рассеяния с соответствующими теоретическими кривыми от найденной модели и сама оболочечная модель. Окончательная модель частицы показана на Рисунке 2.26. Соответствующее значение было  $\chi = 1.7$  при найденной толщине переходных слоев между белковыми оболочками 0.5 и 0.3 нм для 30S и 50S субчастиц, соответственно, и 0.1 и 0.0 нм для их РНК ядер.

Полученная модель, конечно не может передать детали строения, однако, свидетельствует о том, что белковые части окружают компактные РНК-области. Формы областей показанные на рисунке, довольно близки к электронномикроскопическим моделям, а сама структура лучше описывает экспериментальные данные рассеяния: значение  $\chi$  для найденной модели в 1.6 раза меньше, чем для ЭМ моделей. *Большее* расхождение между данными рассеяния и исходными ЭМ моделями видно на Рисунке 2.27.

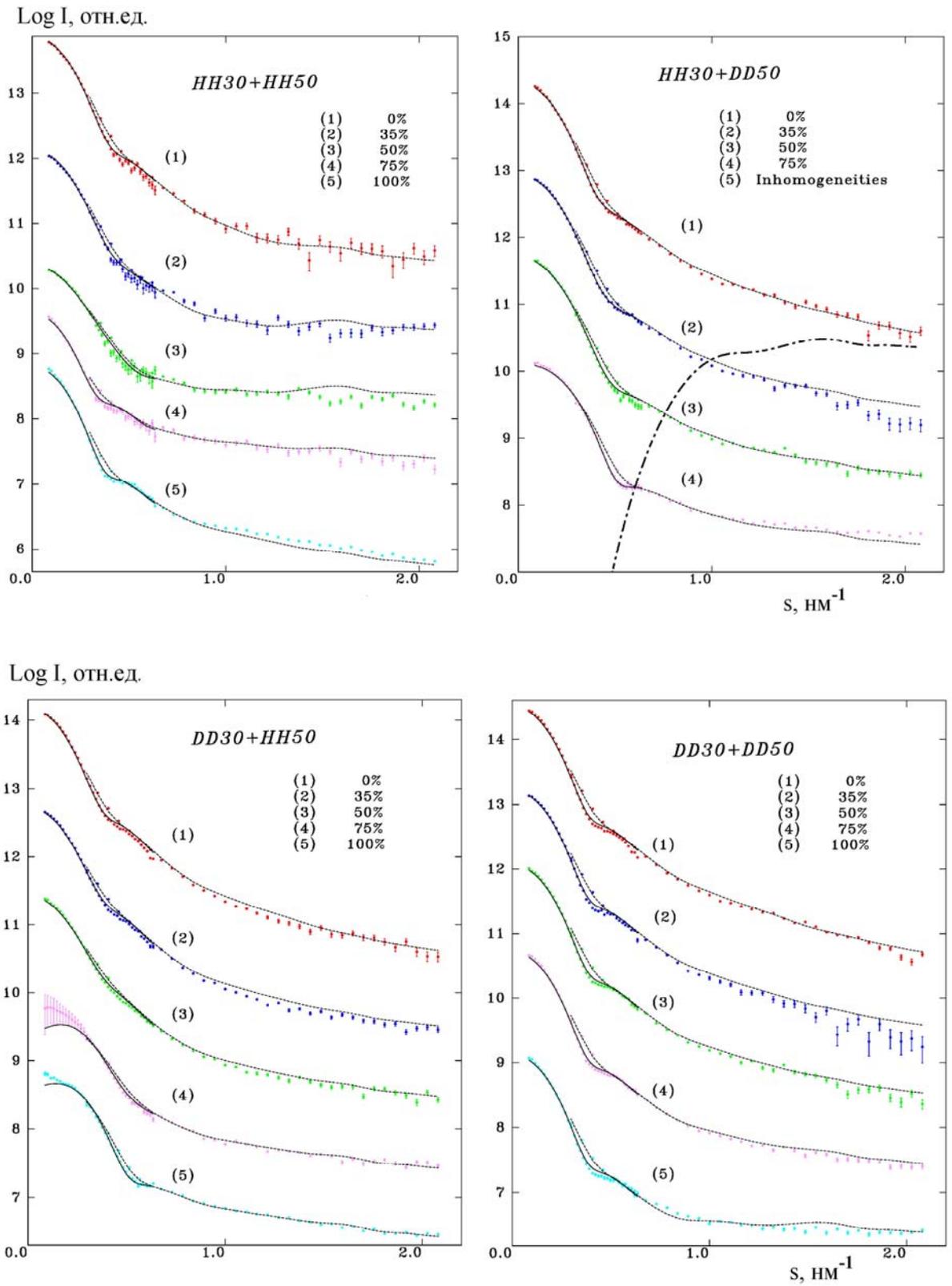


Рисунок 2.25 – См. продолжение.

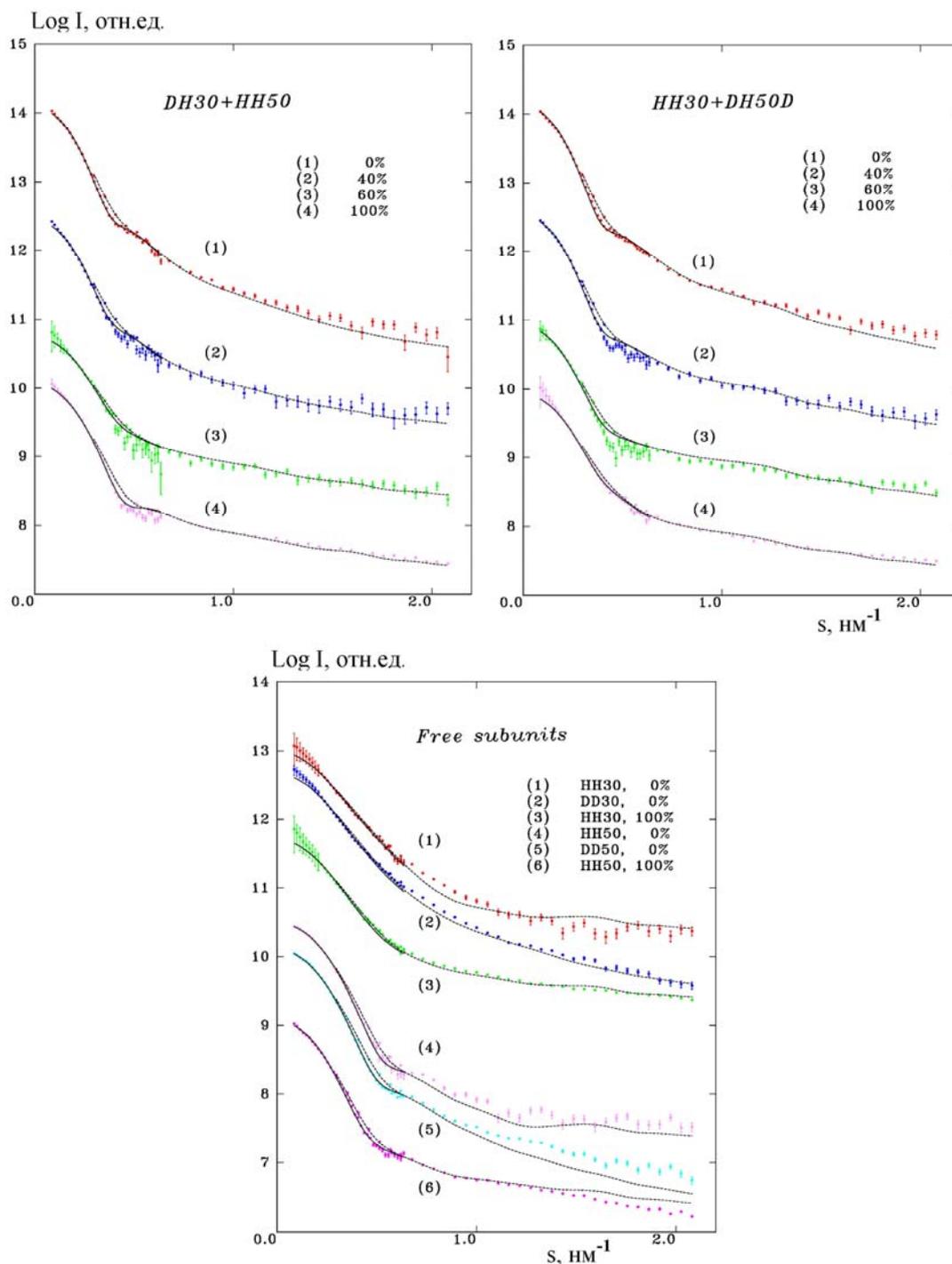


Рисунок 2.25 – Данные нейтронного рассеяния от образцов растворов рибосомы 70S. Точки: экспериментальная интенсивность. Пунктир: рассеяние от найденной по всему угловому диапазону данных модели. Сплошная линия до  $s = 0.06 \text{ nm}^{-1}$ : теоретическое рассеяние от модели, найденной по набору данных, снятых при расстоянии образец - детектор 4 м. Линия точки-тире представляет модельную интенсивность рассеяния от гипотетических внутренних флуктуаций плотности в структуре оболочечной модели. Данные рассеяния смещены по вертикали для большей наглядности. Обозначения образцов объяснены в тексте и соответствуют Таблице 2.3. Процентами показаны концентрации тяжелой воды в растворителе.

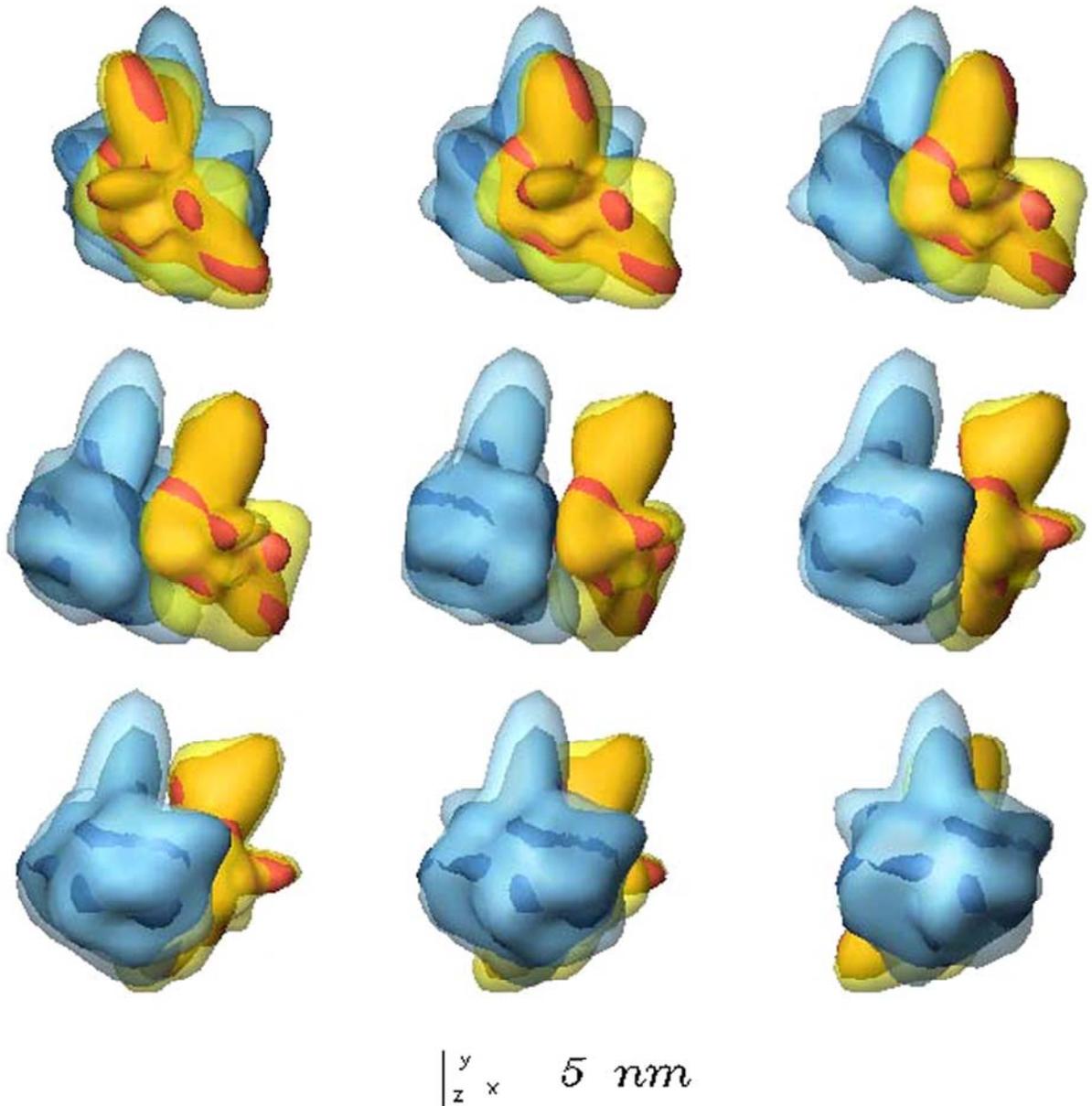


Рисунок 2.26 - Оптимизированная по данным нейтронного рассеяния оболочечная модель частицы рибосомы *E. coli* 70S. Желтым цветом показаны области 30S субъединицы, синим - 50S. Полупрозрачные участки соответствуют областям, занятым белками, расположенными вокруг РНК-ядер, области которых показаны более темными цветами. Левый столбец: частица повернута на  $90^\circ$  вокруг координатных осей. Второй и третий столбцы: структуры 1-го столбца, повернутые на  $22.5^\circ$  и  $50^\circ$  вокруг вертикальной оси, соответственно.

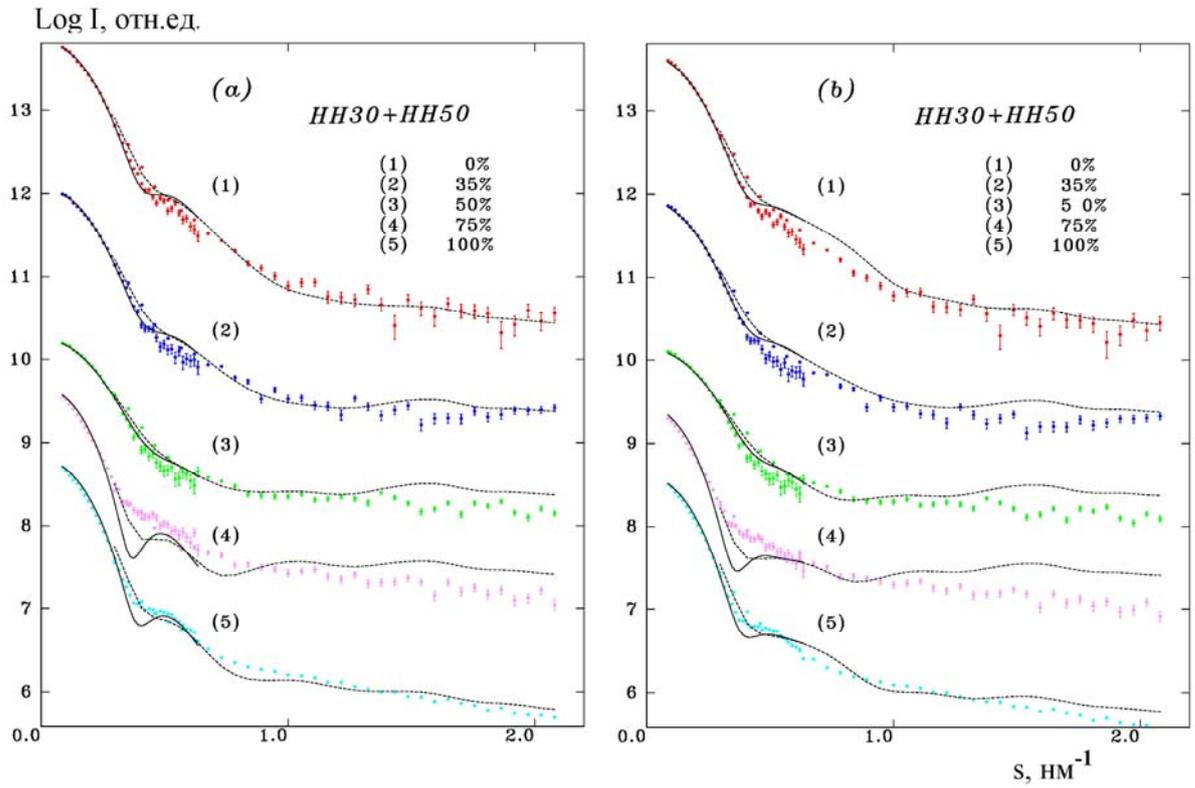


Рисунок 2.27 – Сравнение данных нейтронного рассеяния от образцов растворов рибосомы 70S с теоретическим рассеянием от ЭМ моделей. Обозначения соответствуют Рисунку 2.25.

### 2.2.3 Определение формы частиц методом шариковых моделей

Рассмотренный в разделе 2.1 метод конечных объемных элементов (или "метод шариковых моделей", если объемными элементами служат сферы малого, по сравнению с размером частицы, диаметра) позволяет описывать строение частиц с внутренними полостями, что недоступно методу оболочек, рассмотренному выше. Кроме того, метод позволяет описывать структуры с любым отношением длина / диаметр без каких-либо специальных ограничений. Степень достижимого пространственного разрешения моделей определяется угловым диапазоном данных рассеяния, однако такие оценки могут рассматриваться только как формальные. Однако, остается рассмотренная ранее проблема однозначности и устойчивости решений. Несмотря на ее серьезность, в литературе очень мало работ, посвященных систематическому анализу надежности определения формы по данным рассеяния.

### 2.2.3.1 Алгоритмические принципы поиска шариковых моделей

В данном разделе приведены результаты численных экспериментов по определению формы модельных однородных частиц различной формы с помощью программы DAMMIN, разработанной Д.И. Свергнуном [140], а также ее варианта, модифицированного автором. Алгоритм DAMMIN, как было коротко рассмотрено в разделе 2.1, основан на представлении структуры частиц с помощью набора небольших сфер с одинаковой (единичной) плотностью, и поиске их взаимного расположения в пространстве методом случайного поиска. Абсолютная величина плотности при данном подходе не имеет значения. В таком случае интенсивность малоуглового рассеяния может быть вычислена по формуле

$$I(s) = 2\pi^2 \sum_{l=0}^{L_{\max}} \sum_{m=-l}^l [A_{lm}(s)]^2 \quad (2.60)$$

где

$$A_{lm}(s) = i^l \cdot \sqrt{2/\pi} \cdot f(s) \cdot \sum_{j=1}^{N_a} j_l(sr_j) \cdot Y_{lm}^*(\omega_j) \quad (2.61)$$

есть амплитуда рассеяния структурой, состоящей из  $N_a$  сфер единичной плотности,  $r_j$ ,  $\omega_j$  - полярные координаты сферических тел, заполняющих форму частицы,  $f(s)$  - амплитуда рассеяния одной сферой,  $j_l(x)$  - сферические функции Бесселя,  $Y_{lm}(\omega)$  - сферические гармоники. Максимальный индекс гармоники  $L_{\max}$  задают не менее  $1.5 N_s$ , чтобы ошибки вычисления интенсивности из-за обрыва ряда не превосходили дисперсию экспериментальных шумов.

В отличие от метода оболочек, задача минимизации невязки при поиске шариковой модели дискретна по своей природе, имеет очень большое число параметров, равное числу координат шариков, и поэтому алгоритмы непрерывного поиска здесь неприменимы.

Для поиска пространственного расположения объемных элементов в программе DAMMIN применен метод глобальной минимизации, известный под названием "моделирование отжига", или алгоритм Метрополиса. Согласно

данному методу, программа начинает поиск с некоторой структурной модели, состоящей из шариков (обычно  $N_a = 5-10$  тысяч), случайным образом расположенных в пространстве в узлах плотнейшей упаковки. Эти узлы расположены внутри ограниченной области поиска, которая имеет диаметр, равный максимальному размеру частицы. Величину максимального размера области предварительно определяют из экспериментальных данных по функции  $p(r)$  с помощью программы косвенного преобразования Фурье GNOM [169,170], как описано в предыдущих разделах. Каждый пространственный узел может содержать сферу с плотностью растворителя (0) или частицы (1). Начало координат выбирают в центре области.

Структурная модель частицы представляет, тем самым, набор координат и конфигурационного массива  $\mathbf{V}$  длиной  $N_a$ , элементы которого ассоциированы с координатами расположения виртуальных атомов и содержат последовательные индексы узлов, индексы соответствующих фаз (0 или 1, в зависимости от того, описывает ли данный узел область, занятую растворителем или частицей), индексы разрешенных фаз, количество фаз, которые может принимать каждый узел.

В процессе поиска программа минимизации инвертирует заселенность случайным образом выбранного узла, то есть изменяет ее с фазы частицы (1) на фазу растворителя (0) или наоборот. После каждой модификации программа рассчитывает от текущей структуры кривую интенсивности малоуглового рассеяния и вычисляет взвешенное суммарное квадратичное отклонение от экспериментальной кривой рассеяния по  $N$  точкам (обычно  $N$  выбирают не меньше удвоенного числа шенноновских каналов  $N_S$  (2.28)), аналогично формуле (2.30):

$$R^2 = \left\{ \frac{\sum_{i=1}^N [(I_{\text{exp}}(s_i) - \xi \cdot I_{\text{mod}}(s_i)) \cdot W(s_i)]^2}{\sum_{i=1}^N [I_{\text{exp}}^2(s_i) \cdot W^2(s_i)]} \right\}, \quad (2.62)$$

где  $\xi = \frac{(I_{\text{exp}} \cdot I_{\text{mod}})}{\|I_{\text{exp}}\|^2}$  - МНК множитель, совмещающий кривые рассеяния,

умноженные на  $W(s)$  - весовую функцию, которую назначают в виде

$$W(s) = \begin{cases} s^n & \text{если } s > s \Big|_{\max[I_{\text{exp}}(s) \cdot s^n]} \\ 0.5 \cdot \left\{ s^n + \frac{\max[I_{\text{exp}}(s) \cdot s^n]}{I_{\text{exp}}(s)} \right\} & \text{если } s \leq s \Big|_{\max[I_{\text{exp}}(s) \cdot s^n]} \end{cases} \quad (2.63)$$

Умножение на такую весовую функцию позволяет управлять динамическим диапазоном данных: при  $n = 0$  расчет невязки (2.62) происходит на исходной шкале данных (что имеет смысл при спаде интенсивности не более 1 порядка), с увеличением  $n$  кривая интенсивности трансформируется в контур, проходящий через максимальное значение на некотором  $s = s \Big|_{\max[I_{\text{exp}}(s) \cdot s^n]}$  и спадающий до некоторого значения, тем большего, чем больше  $n$ . Тем самым достигается возможность несколько ослабить вклад в суммарную невязку начального участка, который наиболее подвержен искажающему влиянию рассеяния от агрегатов частиц и одновременно увеличить вклад малоинтенсивного участка при больших  $s$ . Нижняя строка в (2.63) служит для того, чтобы ослабление начального участка было не слишком большим. На практике  $n$  рекомендуется выбирать таким

образом, чтобы величина  $\frac{\max[I_{\text{exp}}(s) \cdot s^n]}{\min[I_{\text{exp}}(s) \cdot s^n]}$  была в диапазоне не более 10 - 50. При

таким взвешивании обычное в МНК подходе деление невязок на дисперсию экспериментальных шумов не применяется.

Здесь следует сделать одно **важное замечание**. Использование совмещающего множителя  $\xi$  перед расчетом невязки применяется практически во всех рассматриваемых в данной работе программах интерпретации данных. Специальное исследование, проведенное автором, показало, что в случаях

подгонки формы кривых (а не значений самих ординат) без применения  $\xi$ , целевые функции приобретают значительно более сложную форму поверхности отклика, что, как правило, приводит к появлению большого количества локальных минимумов и ухудшению математической обусловленности задачи, приводящих к замедлению скорости поиска на многие порядки и к невозможности достижения приемлемого решения. Таким образом, множитель  $\xi$  обеспечивает не только независимость от масштаба данных, но и саму возможность получения решения. Для исследования его влияния на поверхность отклика целевой функции в задачах непрерывной минимизации были проведены расчеты одномерных сечений целевых функций по заданным комбинациям параметров моделей (такая возможность изначально была заложена в пакете минимизации OPTIS) в различных промежуточных точках поиска решения как задач разложения спектров смесей, так и в задачах поиска оболочечных моделей частиц по данным рассеяния и поиска распределений по размерам наночастиц. Множитель  $\xi$  ослабляет корреляцию между параметрами моделей, что было показано численными экспериментами.

По аналогии с задачами непрерывной минимизации, совмещающий множитель применен и в программе DAMMIN.

Минимизируемая целевая функция задачи имеет вид штрафной функции (2.30), аналогичной оболочечной модели и зависящая от элементов конфигурационного массива  $\mathbf{V}$ :

$$\Phi(\mathbf{V}) = R^2 + w_D \cdot P_D + w_L \cdot P_L + w_C \cdot P_C + w_G \cdot P_G. \quad (2.64)$$

Для адекватного описания формы частицы и исключения влияния на кривую рассеяния упаковки шариков, их диаметр должен быть достаточно мал, на практике не менее 2 - 3 тысяч в области поиска для глобулярных частиц и до 20 - 30 тысяч для анизометричных тел. При этом размер шариков оказывается в 5 - 20 раз меньше, чем пространственное разрешение структурной модели и для описания однородных областей структуры они должны находиться в плотноупакованном состоянии. Для обеспечения этого в целевую функцию

добавляют штрафные члены (уменьшающиеся по мере поиска в той же степени, что и температура), которые отражают требования неразрывности структуры  $w_D P_D$  и отсутствия отдельно расположенных шаров  $w_L P_L$  (штраф за "рыхлость"). Штраф  $P_D$  вычисляют как отношение общего числа шаров структуры к числу шаров, составляющих домен, образованный непосредственно контактирующими атомами.  $P_L$  вычисляют по формуле  $P_L = 1 - \langle 1 - \exp(-N_e) + \exp(-12) \rangle_N$ , где 12 - максимальное число контактов шарика с ближайшими соседями,  $N_e$  - реальное число контактов у данного шарика в структуре,  $\langle \rangle_N$  означает усреднение числа контактов, приведенное к одному шарика по всей структуре. Для весовых коэффициентов  $w_D$  и  $w_L$  рекомендованы значения 0,01 - 0,001 из соображения примерного равенства вкладов от  $R^2$  и штрафов в точке минимума функционала (2.64).

В процессе поиска структуры основной (максимальный по объему) связный домен, состоящий из контактирующих шариков, может сформироваться вблизи границы области поиска и, тем самым, приобрести форму, ограниченную оболочкой области. Для предотвращения этого в программе DAMMIN первоначально предусмотрен штраф за уход центра тела из начала координат в виде  $P_C = R_g / R_{g,GNOM}$ , где  $R_g$  - радиус инерции текущей модели, рассчитанный по всем формирующим ее  $N_p$  шарикам,  $R_{g,GNOM}$  - оценка (2.7). Модельные расчеты показали неэффективность такого подхода, так как, во-первых, многочисленные оторванные от основного домена шарики вносят свой вклад в радиус инерции и могут "уравновешивать" смещенный домен и, во-вторых, тенденция штрафа уменьшать радиус инерции противоречит смыслу радиуса инерции как структурного инварианта, который должен быть определенной величины, определяемой ходом кривой рассеяния. Поэтому в новой версии программы, которую использовал автор, штраф  $P_C$  назначают за уход центра тяжести частицы из центра области поиска как отношение максимального радиуса области поиска к расстоянию от начала координат центра тяжести модели, состоящей только из контактирующих шариков с числом контактов не менее 2. Весовой множитель  $w_C$

назначают в пределах 0.1 - 1.0. Модифицированный штраф  $P_G$  за радиус инерции  $R_G$  представляет собой отношение разности радиуса инерции такого связного домена и радиуса инерции, предварительно найденного с помощью программы GNOM [169] по экспериментальным данным рассеяния, к диаметру области поиска  $D_{max}$ . Соответствующий вес  $w_G$  также можно назначать в пределах 0.1 - 1.0.

На Рисунке 2.28 показан принцип построения шариковой модели.

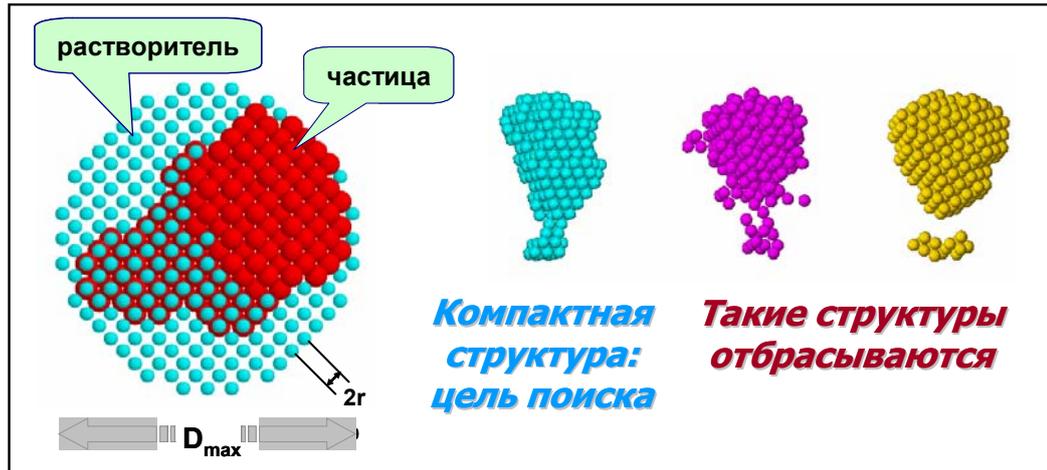


Рисунок 2.28 – Наглядное представление шариковой модели в программе DAMMIN.

Если значение целевой функции (2.64) после варьирования заселенности узла в области поиска оказывается меньше предыдущего значения ( $\Delta = \Phi_k - \Phi_{k-1} < 0$ ,  $k$  - номер шага), то структура запоминается и следующая вариация будет проводиться относительно нее. Процедура моделирования отжига состоит в том, что на каждом шаге в качестве следующей структуры с некоторой вероятностью программа может принимать ту, которая имеет худшее значение  $\Phi$ . Вероятность задается параметром "температуры"  $T$ : если  $\Delta > 0$ , структура принимается с вероятностью  $\exp(-\Delta/T)$ . Со временем, после каждых  $K_{max}$  шагов (или  $10N$  шагов с  $\Delta < 0$ ) параметр  $T$  уменьшается ( $T_{k+1} = T_{fact} \cdot T_k$ ), что позволяет программе все реже запоминать неудачные структуры. Коэффициент  $T_{fact}$  может назначаться в пределах  $T_{fact} = 0.90 - 0.95$ . Необходимость процедуры отжига связана с тем, что целевая функция  $\Phi$  имеет много локальных минимумов, выход из которых обеспечивается ненулевой вероятностью принятия худшего решения. С ростом

числа испытаний эта вероятность стремится к нулю и в конце концов процедура поиска становится идентичной методу Монте-Карло, т.е. на каждом шаге принимаются только удачные попытки. Такой алгоритм позволяет программе на начальных этапах поиска при больших "температурах"  $T$  сформировать конфигурацию, которая соответствует области существования устойчивых решений, разумеется, с некоторой вероятностью.

Для управления скоростью поиска назначают максимальное число испытаний  $N_T$  для каждой температуры, по достижению которого температура снижается и цикл повторяется. В начале поиска, при высоких значениях  $T$ , принятие с высокой вероятностью плохих решений приводит к тому, что конфигурация шариков в пространстве остается в целом случайной и для ускорения сходимости вводят второе максимальное число - число принятых решений  $N_{succ}$  при данной температуре,  $N_{succ} = (0.1 - 0.3) N_T$  по достижению которого программа также переходит к следующей температуре. Тем самым в начале поиска экономится число испытаний. По мере снижения температуры относительное число принятых решений уменьшается и программа всегда делает  $N_T$  шагов при числе принятых решений  $N_{acc} < N_{succ}$ . Оптимальный на практике выбор  $N_T = (1 - 5) K$ , где  $K$  - число узлов в области поиска.

Слишком большая начальная величина  $T$  приводит к большому числу итераций в начале поиска, при которых структура не формируется, представляя собой случайным образом расположенный набор шариков. Слишком малая величина  $T$  приводит к тому, что программа, относительно редко принимая худшие решения в качестве рабочей модели, попадает в область локального минимума, вероятность выхода из которого достаточно мала.

Поэтому автором был разработан и применен алгоритм оценки оптимального начального значения  $T$ . Согласно разработанной схеме, программа делает 1000 испытаний и определяет дисперсию значений целевой функции (2.64). Стартовое значение  $T$  назначается равным утроенной величине дисперсии. Это обеспечивает не более 1-3 "бесполезных" температурных циклов в начале

поиска, тогда как величина начальной температуры, в зависимости от задачи, может иметь значения от  $10^{-3}$  до  $10^{-6}$ . Такой разброс говорит о неэффективности использования константного значения стартовой температуры: ее завышение, по сравнению с найденным оптимальным значением, например, всего в 2 раза, приводит к 10-15 начальным циклам поиска, во время которых члены целевой функции практически не изменяются и структура остается набором шариков, случайным образом расположенных в пространстве.

Автор в модифицированном варианте программы применил следующий шаг повышения эффективности поиска, который состоит в том, что после достижения  $N_{acc} < 0.3 N_{succ}$ , или когда относительное число изолированных шариков структуры становится менее 3%, программа начинает пропускать узлы растворителя, не контактирующие с каким-либо узлом, занятым шариком частицы. При этом варьируются только узлы, расположенные вблизи частицы и узлы самой частицы. Такая модификация позволила существенно повысить эффективность поиска: относительное число успешных испытаний выросло с 6 - 9% до 16 - 20% по отношению к общему числу испытаний и при этом варьируемость формы в процессе поиска даже несколько выросла. Возросшая варьируемость полезна, так как программа имеет возможность проверить большее число структур при том же числе модификаций области поиска. При этом разброс окончательных решений уменьшился, вплоть до 4-6% против 5-15% в исходном варианте программы для глобулярных частиц.

Вместе с изменением схемы поиска, была применена другая схема штрафования за отклонение центра тела от начала координат, речь о которой пойдет в следующем разделе.

По мере снижения температуры весовые коэффициенты при штрафных членах также уменьшаются с тем же самым коэффициентом  $T_{fact}$ , что и температура. В самом конце поиска при малых  $T_{fact}$  программа принимает только лучшие решения и алгоритм становится идентичным методу Монте-Карло.

Оценка качества финального решения проводится по обычному критерию хи-квадрат по теоретической кривой рассеяния, рассчитанной на экспериментальной сетке углов из  $M$  точек:

$$\chi^2 = \frac{1}{M-1} \sum_{j=1}^M \left[ (I_{\text{exp}}(s_j) - I(s_j)) / \sigma(s_j) \right]^2, \quad (2.65)$$

где  $\sigma(s_j)$  - оценка экспериментальной ошибки в  $j$ -ой точке.

Основным преимуществом данного подхода является отсутствие ограничений на сложность строения частицы. Частицы могут быть представлены однофазной моделью, состоящей из сфер с одинаковой (единичной) плотностью, или многофазной моделью (до 4-х фаз в программе DAMMIN) из шариков разной плотности. При поиске модели в программе можно задавать оси и плоскости симметрии в качестве дополнительных требований к решению, что существенно сужает разброс решений.

Для оценки устойчивости решений и нахождения наиболее вероятной структуры группой сотрудников EMBL с участием автора был разработан ряд программ. Программа SUPCOMB [171] позволяет проводить автоматическое совмещение объектов в прямом пространстве, минимизируя степень различия между двумя моделями, представленными наборами точек (центрами шариков или узлами сетки, описывающей форму частиц). Критерием различия является нормализованное пространственное отклонение (**NSD**), рассчитываемое следующим образом: от каждой точки в первой модели определяются расстояния до каждой точки во второй модели и находятся их минимальные значения. Затем та же процедура выполняется для каждой точки второй модели относительно первой и все полученные величины суммируются и нормируются на средние расстояния между соседними точками в каждой модели. Для похожих структур NSD принимает значения меньше единицы,  $\text{NSD} > 1.0$  соответствует существенно различным структурам. Программа DAMAVER [A28] использует SUPCOMB для оптимального перекрестного совмещения нескольких (обычно не менее 10)

независимо рассчитанных по одному и тому же набору данных моделей, после чего проводит их наложение, подсчитывая число попаданий структур в соответствующие ячейки области поиска. Все полученные заселенные ячейки образуют некую структуру, которая представляет область разброса решений (total spread region, **TSR**), из которой выделяется часть, представленная ячейками с наибольшим числом попаданий (наиболее вероятный объем, most populated volume, **MPV**), и имеющую объем, соответствующий инварианту объема (2.5). Полученная структура, которую можно назвать общей частью решений, может быть уточнена в процессе нового поиска, осуществляемого внутри области разброса, при невысокой стартовой температуре.

### 2.2.3.2 Исследование устойчивости и однозначности определения строения частиц методом шариковых моделей

Однозначность определения структуры может быть выражена в терминах дисперсии решений для конкретной задачи. Из-за отсутствия линейной связи между параметрами структуры (координатами шариков) и входными данными рассеяния выражение для дисперсии решений аналитически получить невозможно. Тем не менее, можно получить ответ на вопрос о степени однозначности (или устойчивости), численно проводя анализ серии решений, которые получены методом случайного поиска и, вследствие этого, независимы. Конечно, такой ответ всегда является неполным, но это единственная возможность исследовать проблему неопределенности в данном подходе.

Для исследования пределов возможности метода шариковых моделей были рассчитаны теоретические кривые рассеяния для различных геометрических тел по аналитическим формулам, приведенным в [126]. Влияние случайных шумов в данной серии численных экспериментов не изучали, но такое исследование рекомендуется при решении реальных задач на стадии предварительного моделирования. Рассчитанные кривые рассеяния использовали в качестве экспериментальных для восстановления по ним форм соответствующих частиц без привлечения дополнительной информации.

На Рисунках 2.29 - 2.33 показаны кривые рассеяния от модельных тел, сами тела и результаты восстановления [A28].

Восстановление форм показало, что глобулярные тела с отношением размеров не более 1:3 восстанавливаются достаточно уверенно (Рисунок 2.29), NSD факторы в прямом пространстве обычно не превышают 0.5 - 0.6. Увеличение анизотрии (т.е. отношения максимального размера частицы к минимальному) приводит к снижению устойчивости решения (Рисунок 2.32) с NSD-факторами 0.7 - 0.9. Можно заметить, что в случае тонких дисков с пропорцией диаметр/высота = 5:1 отклонение расчетной модели от истинного

тела может быть иногда значительно, в данном примере NSD-факторы были 1.0 - 1.3.

Если размеры внутренних деталей структуры относительно невелики, как в примере 13, то артефактом будет в этом случае смещение неоднородности относительно центра, что может приводить, как в данном примере, к закрыванию сквозных каналов в структуре.

В случае анизометрии 10:1 в случае плоских тел структуру уже восстановить не удастся (NSD-фактор 1.2 - 1.5 для тела 12 на Рисунке 2.32), если на решение не накладывать никаких ограничений (например, ось или плоскость симметрии). В случае поиска такого тонкого диска с введением осей симметрии выше 3-ей NSD становится меньше 0.5 вплоть до 0.1 при оси 6-го порядка - точность восстановления очень высока. Впоследствии этот факт был использован при определении формы макромолекул иммуноглобулина М и ревматоидного фактора человека.

Сильно вытянутые частицы, несмотря на их значительную анизометрию, можно восстановить с точностью до изогнутости (тело 10 на Рисунке 2.32), но только с помощью шариковых моделей.

Оболочечные модели не могут обеспечить разумные решения в случае анизометрии большей, чем 2:5, как видно из Рисунка 2.32.

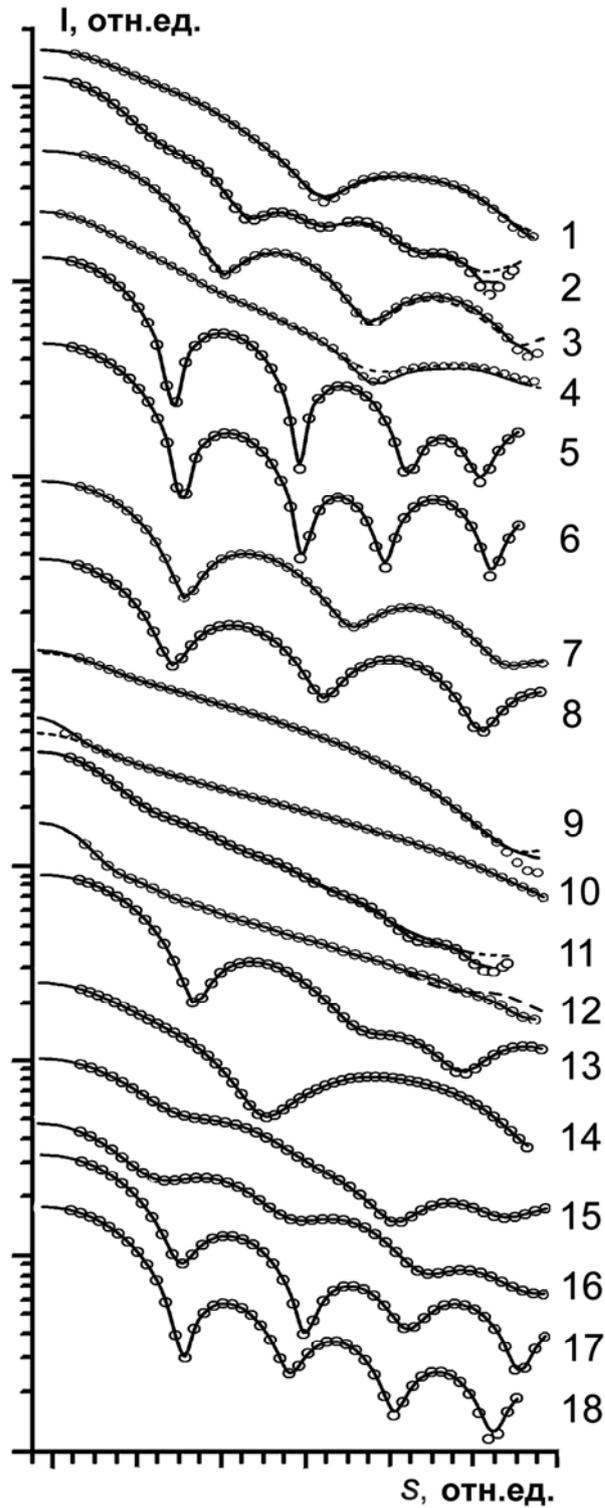


Рисунок 2.29 - Формы кривых интенсивностей рассеяния от исходных модельных тел (кружки), кривые от найденных оболочечных моделей (программа SASHA , пунктирные линии) и от шариковых решений DAMMIN (сплошные линии). Числа справа соответствуют номерам тел, представленных на Рисунках 2.30 - 2.33. Абсциссы и ординаты отдельных групп кривых умножены на соответствующие множители и смещены по вертикали для лучшей визуализации, поэтому числа не проставлены.

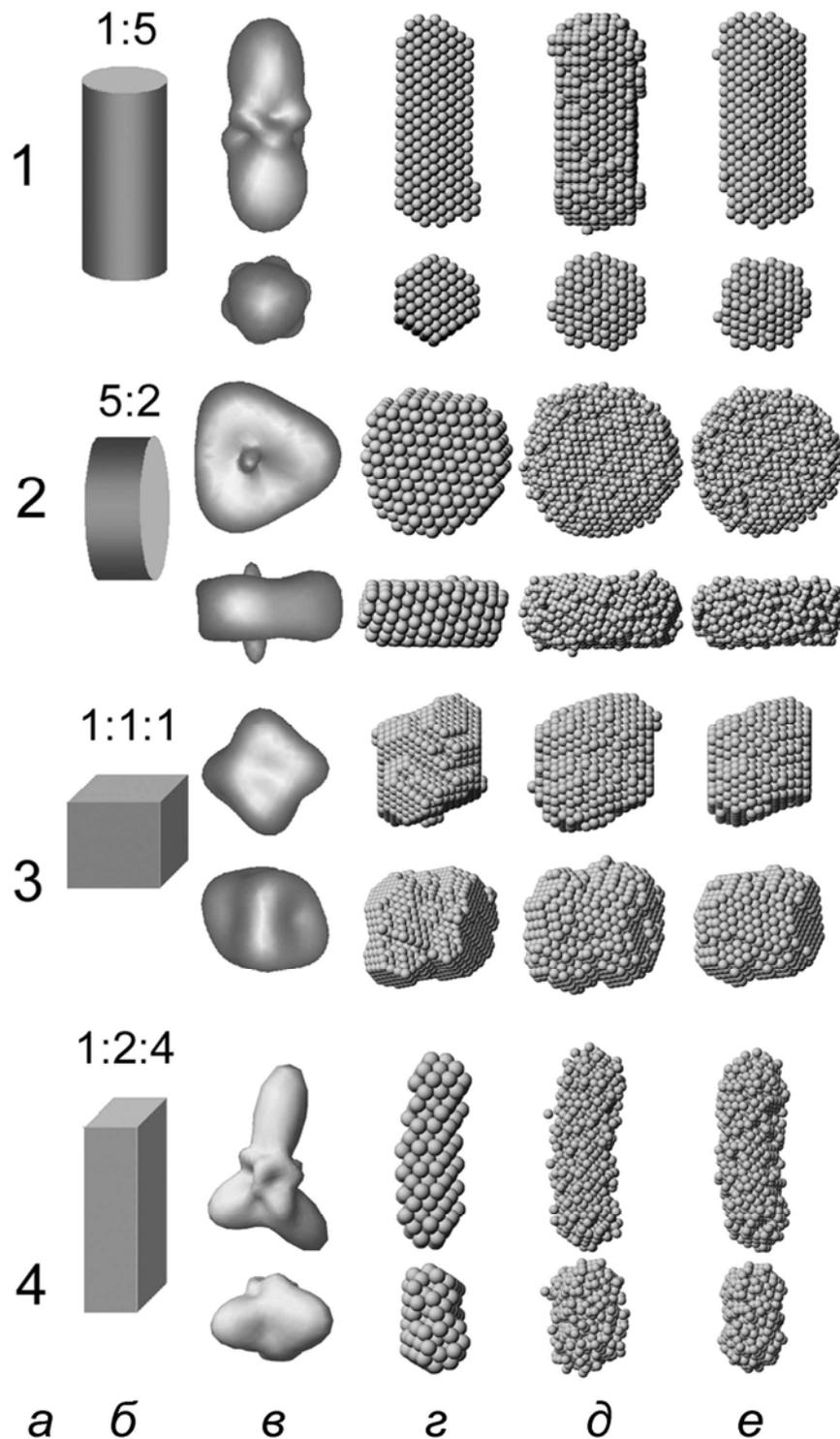


Рисунок 2.30 - Определение формы глобулярных частиц. Столбец *a*: номер структуры; *б*: исходное геометрическое тело (числами показаны отношения основных размеров – диаметр : высота для цилиндров, отношения сторон для призм, отношения диаметров и высоты для полых тел); *в*: оболочечная модель, полученная программой SASHA; *г*: типичная шариковая модель, находящаяся программой DAMMIN; *д*: область перекрытия решений; *е*: общая область решений. Структуры показаны в двух перпендикулярных ориентациях.

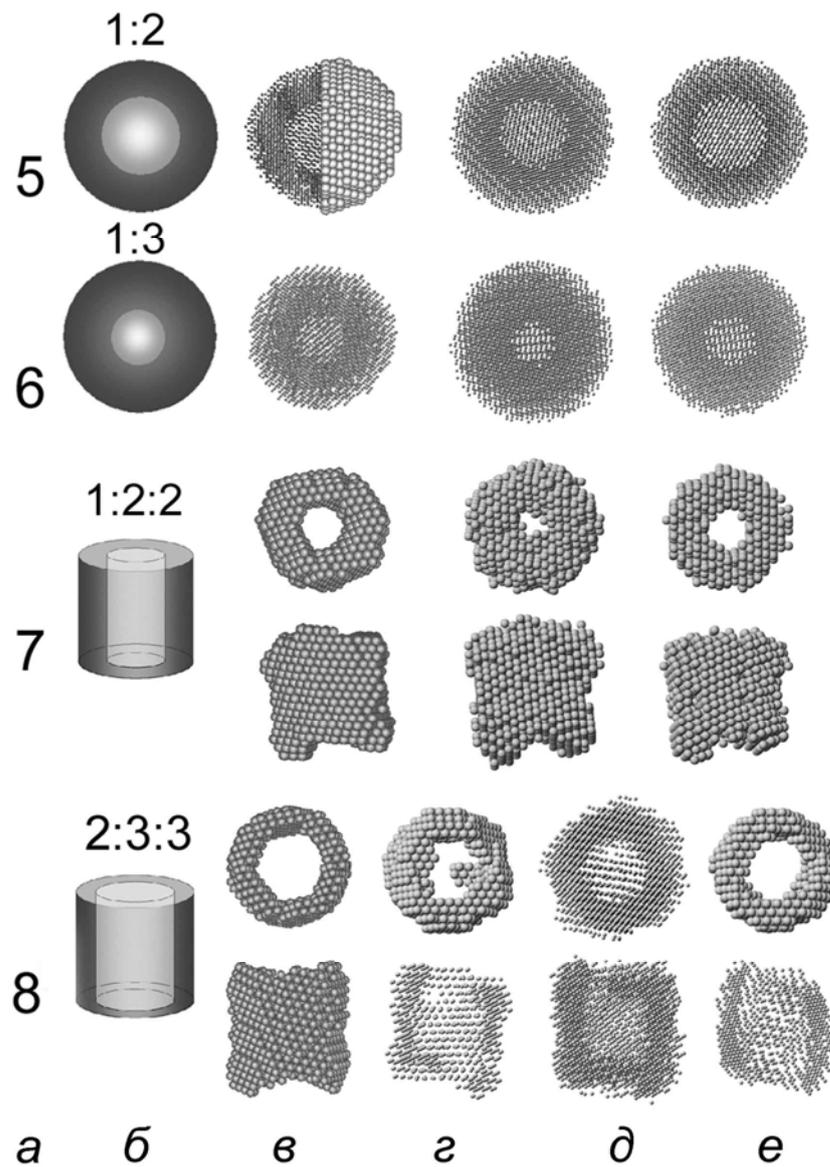


Рисунок 2.31 - Определение формы глобулярных частиц с полостью внутри. Восстановление оболочечных моделей не проводилось из-за наличия полостей.

Для нижней структуры в столбце *г* показан разрез модели. Столбцы *д, е* - аналогично Рисунку 2.29.

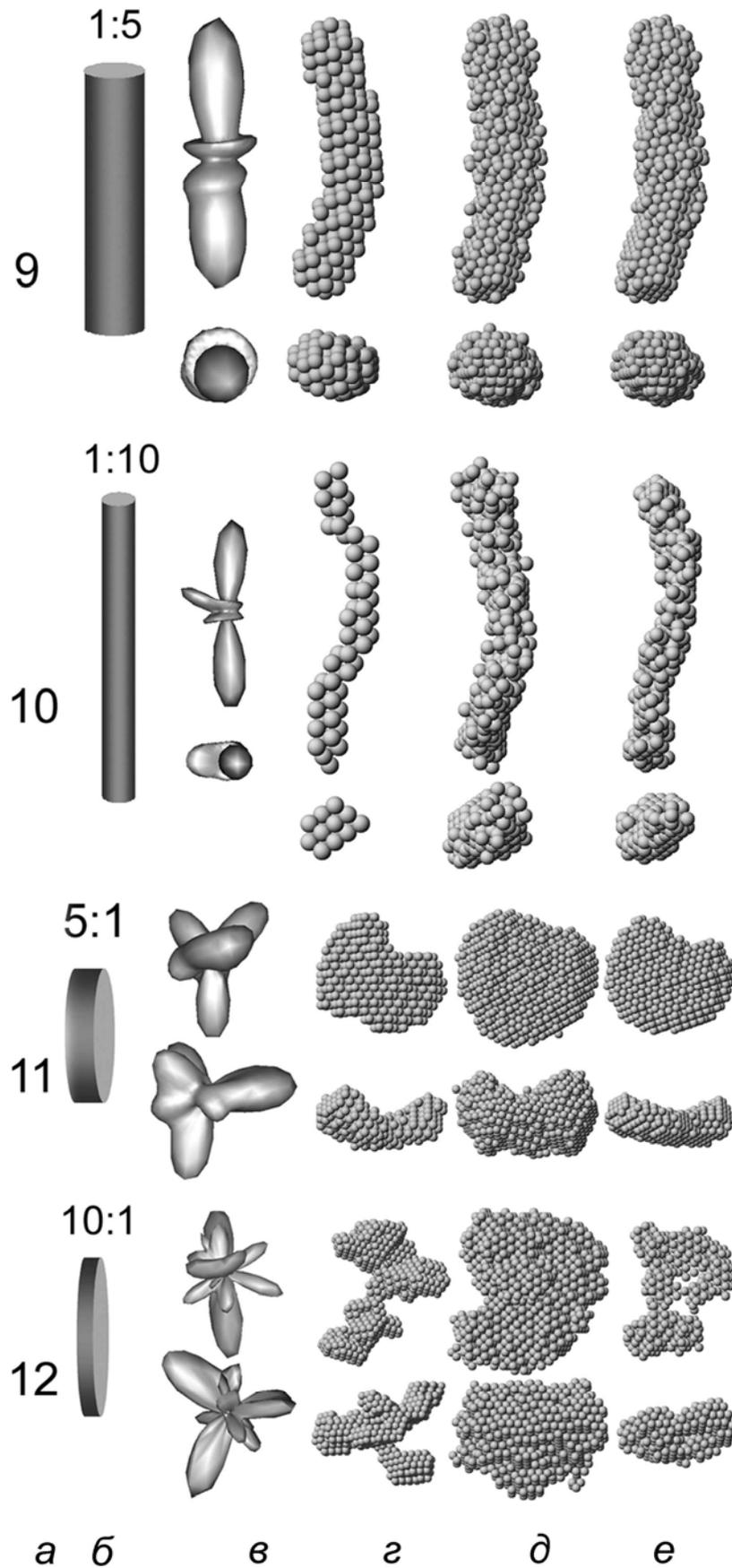


Рисунок 2.32 - Определение формы частиц со значительной анизотрией.  
Обозначения аналогичны Рисунку 2.29.

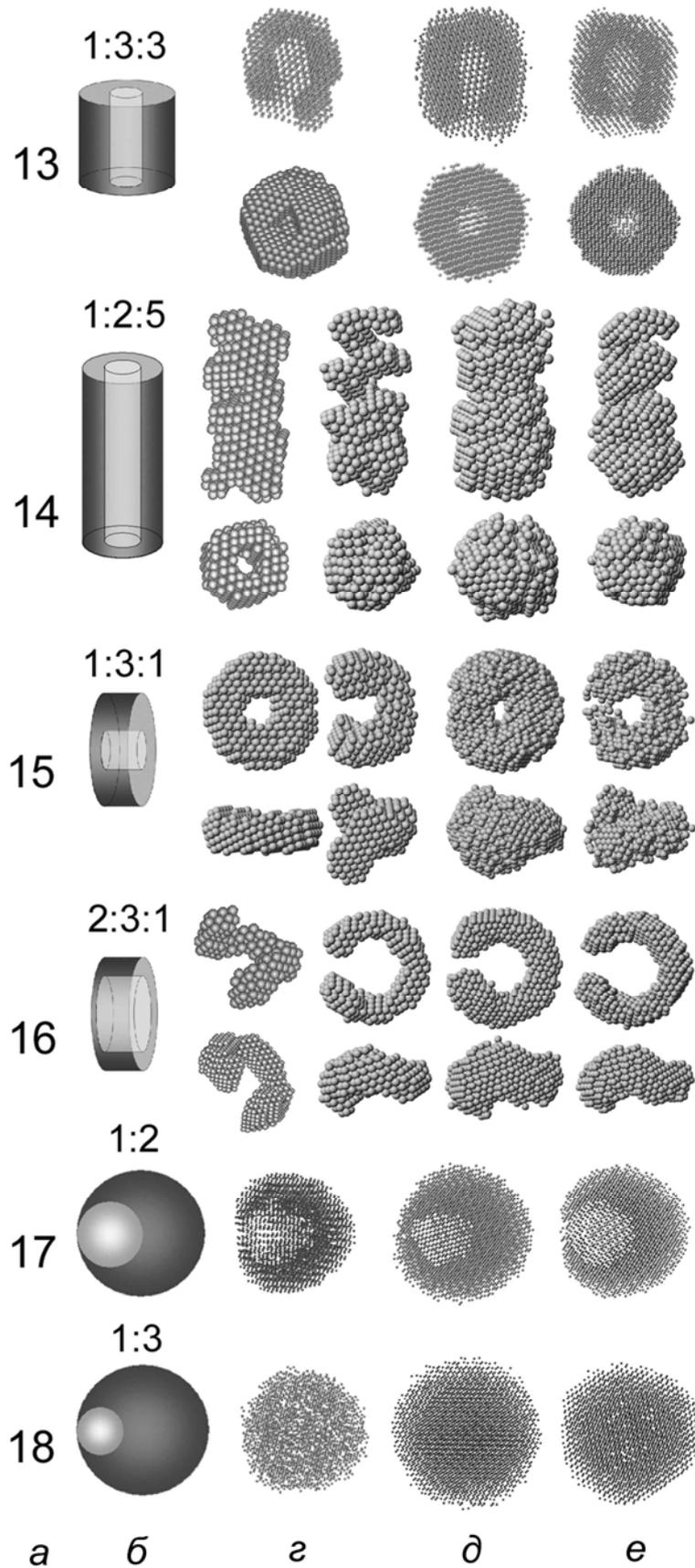


Рисунок 2.33 - Определение формы частиц со значительной анизотрией и внутренней полостью. Обозначения аналогичны Рисунку 2.30. Для структур 14 - 16 в столбце *г* показаны по два типичных решения.

В случае полых тел, если отверстие находится в центре, восстановление практически точное (тела 5 и 6 на Рисунке 2.31), что согласуется с теоремой единственности определения структуры для сферически-симметричных частиц. Ацентричные пустоты уже не удастся восстановить: программа поиска находит рыхлые структуры, в которых полость часто разбивается на несколько небольших, хаотично расположенных в пространстве (пример 18 на Рисунке 2.33). Тонкие кольцевые структуры часто разрываются и решение напоминает "шайбу Гровера" (примеры 15 и 16). Полые цилиндры в случае значительной анизотропии восстанавливаются в виде тела, часть которого представляет собой спиральную структуру (пример 14).

Основные выводы из модельных расчетов следующие [A28]:

**(I)** Плоские частицы, как правило, сложнее восстанавливать, чем удлиненные. Степень допустимой анизотропии для надежной реконструкции формы не должна превышать 1:10 и 5:2 для вытянутых и плоских частиц, соответственно. Перед проведением обработки реальных данных желательно оценивать анизотропию частиц, например, находя параметры наиболее подходящего трехосного эллипсоида - данная опция имеется, в частности, в программе поиска формы оболочек SASHA.

**(II)** Для глобулярных частиц внутренние полости могут быть восстановлены, если их размер составляет  $1/3 - 2/3$  от внешнего диаметра. В случае плоских частиц с отверстиями общая форма может быть восстановлена, но могут появляться артефакты (например, кольцо может быть разорвано). Длинные гладкие полые цилиндрические структуры могут давать в решении винтовые участки. В сферических частицах концентрические пустоты восстанавливаются значительно лучше, чем ацентрические.

**(III)** Среднее NSD по набору решений дает полезную на практике оценку надежности решения. Значения NSD более 0.9 указывают на неустойчивость восстановления и в таких случаях для надежного определения формы требуется привлечение дополнительной информации.

(IV) Усреднение решений позволяет в большинстве случаев и оценить качество реконструкции формы, и провести ее уточнение.

Расчеты шариковых структур программами DALAI\_GA (Чакон и др., 1998;.. Чакон и др., 2000) и SAXS3D (Walther, Cohen & Doniach, 2000) показали результаты, близкие к DAMMIN, хотя и несколько менее стабильные, что может быть связано с неоптимальным назначением параметров алгоритмов, выбор величин которых неочевиден. Большинство модельных расчетов в данном разделе были сделаны без наложения ограничений на симметрию возможных моделей, и без привлечения информации об анизотропии частиц. Хотя SASHA и DAMMIN позволяют вводить ограничения как на симметрию, так и на степень анизотропии, эти параметры не использовали, чтобы сделать результаты моделирования как можно более общими. Сделанные выводы относятся к определению формы частиц без привлечения дополнительной информации.

В реальных системах, таких, как сколлапсированные полимерные гели иногда образуются винтовые и спиральные структуры [A29]. Вывод о пользе нахождения общей части решения с помощью SUPCOMB можно видеть из Рисунков 2.34 - 2.38.

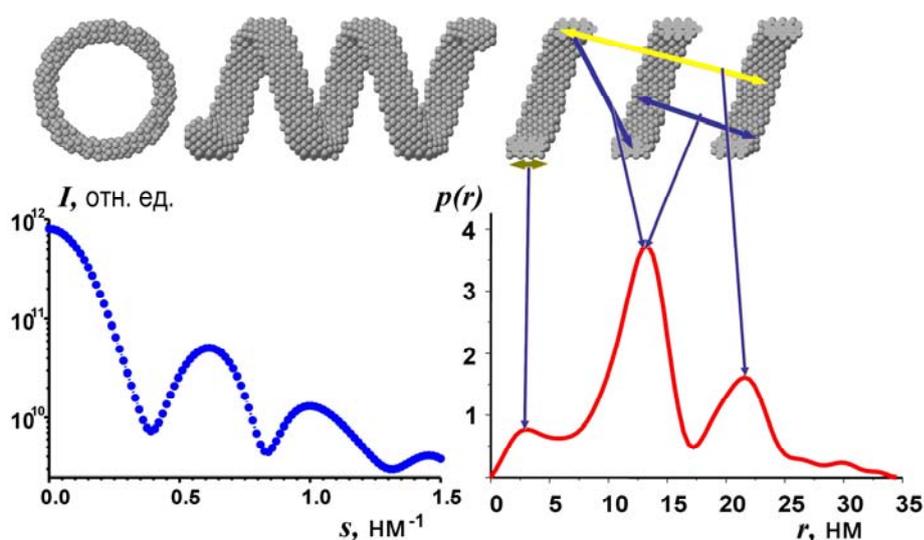


Рисунок 2.34 - Исходная модель 3-х витковой спирали. На графике внизу слева представлена теоретическая кривая интенсивности рассеяния, справа - функция парных расстояний (в относительных единицах) с указанием типичных размеров элементов структуры.

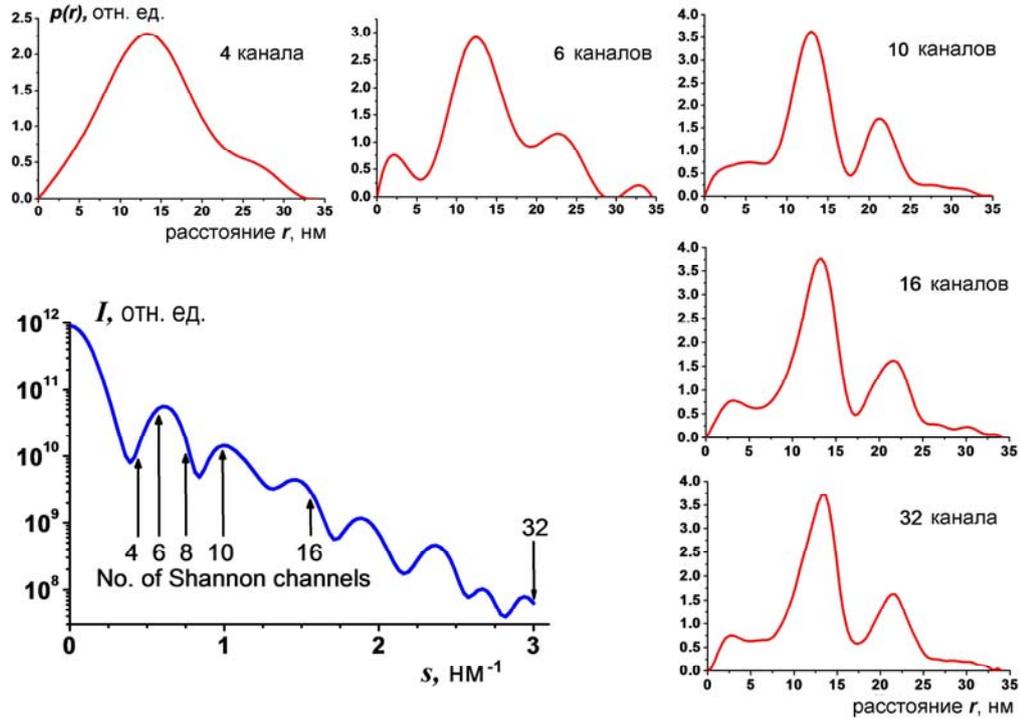


Рисунок 2.35 - Кривая рассеяния от 3-х витковой спирали с пометкой максимальных значений  $s$ , использованных при восстановлении формы (график внизу слева). Соответствующие функции парных расстояний показаны вверху и справа. Видно, как по мере увеличения углового диапазона данных рассеяния в функции парных расстояний проявляются типичные расстояния от внутренней структуры.

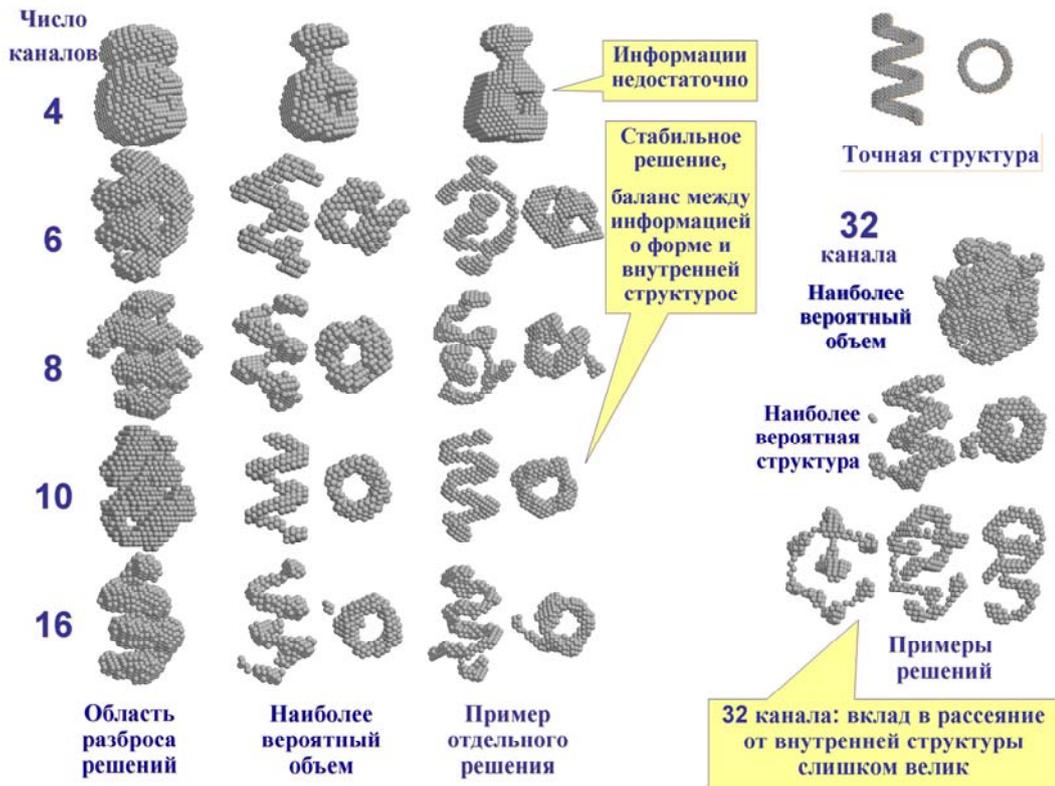


Рисунок 2.36 - Результаты восстановления формы 3-х витковой спирали по данным Рисунка 2.35. Пояснения даны на рисунке.

Из Рисунка 2.36 видно, что при восстановлении данной модели увеличение углового диапазона приводит сначала к улучшению качества восстановления, достигая максимума при числе шенноновских каналов 10, затем - к потере стабильности решения, несмотря на то, что, согласно Рисунку 2.35, кривая функции парных расстояний при  $s > 10$  каналов практически не меняется. Это послужило поводом к созданию варианта программы поиска формы путем минимизации отклонения между экспериментальной и теоретической (рассчитанной от модели) функциями парных расстояний. Однако расчеты показали, что устойчивость решений в данном случае как правило оказывается хуже, чем при минимизации невязки в обратном пространстве.

В случае структуры с большим количеством деталей, как на Рисунках 2.37 и 2.38, восстановление менее устойчиво, хотя рассмотренная основная тенденция зависимости от углового диапазона также сохраняется.

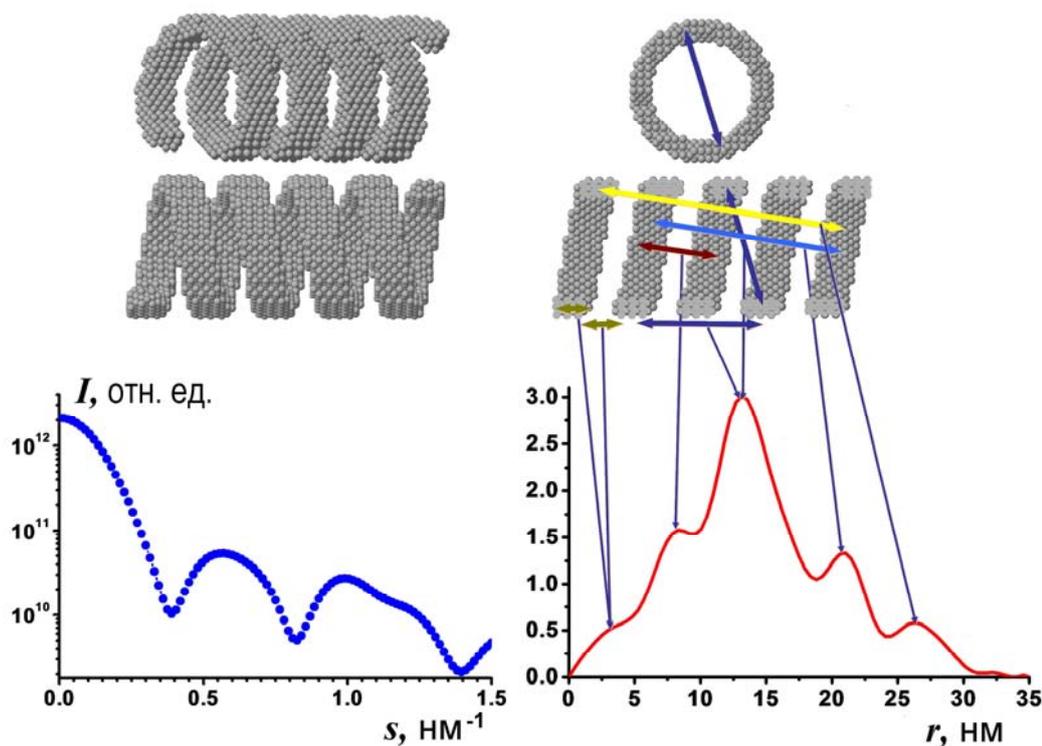


Рисунок 2.37 - Исходная модель 5-витковой спирали. На графике внизу слева представлена теоретическая кривая интенсивности рассеяния, справа - функция парных расстояний (в относительных единицах) с указанием типичных размеров элементов структуры.

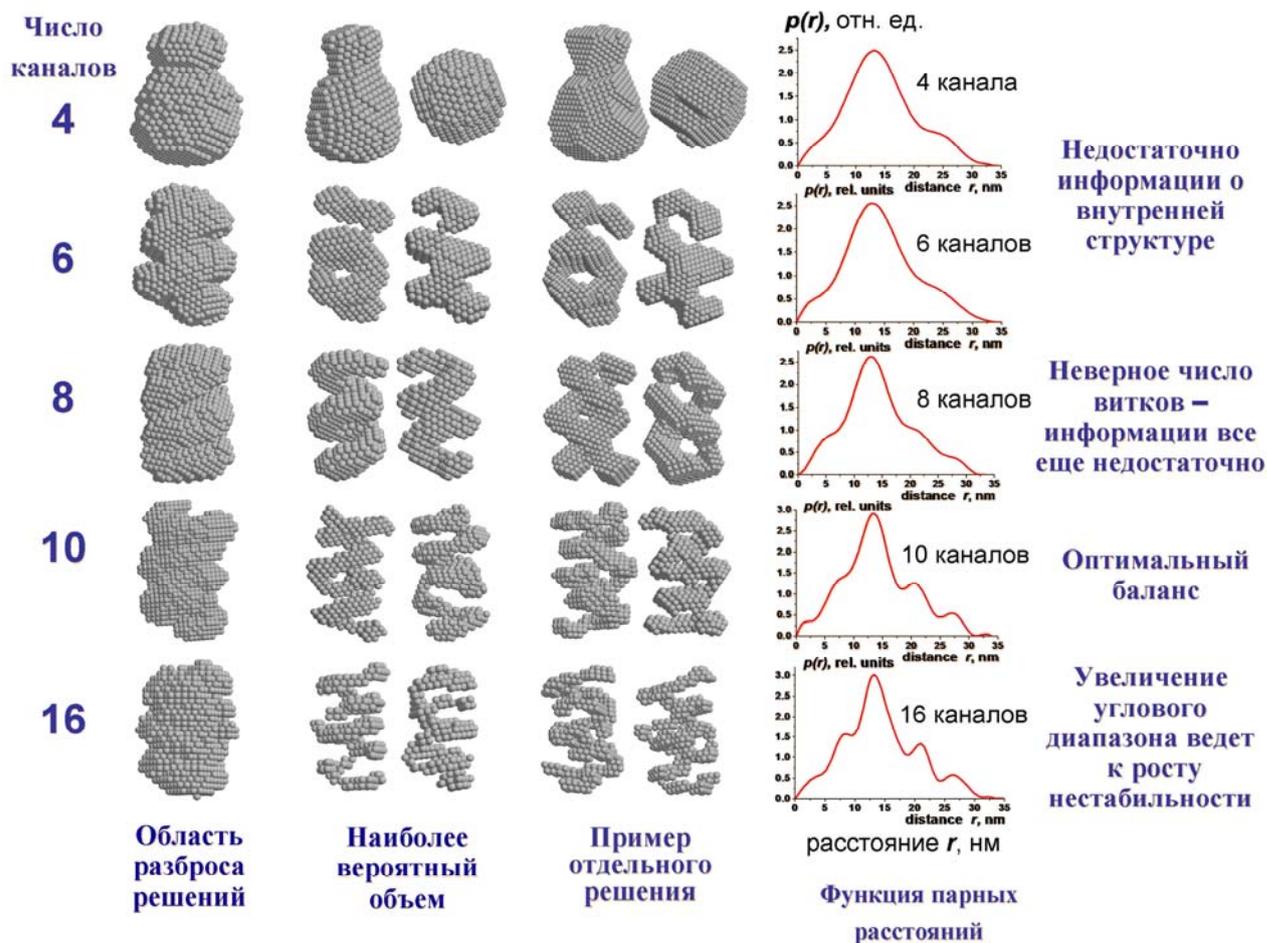


Рисунок 2.38 - Результаты восстановления формы 3-х витковой спирали по данным Рисунка 2.37. Пояснения даны на рисунке.

Очередной вывод из проведенного моделирования состоит в том, что - **(V)** при восстановлении формы частиц существует некоторый оптимальный угловой диапазон данных рассеяния. Его величину для всех возможных случаев указать невозможно, но его можно определить путем моделирования, если существует априорная информация о возможной структуре.

Важный вопрос, ответ на который часто оказывается затруднительным, заключается в выборе **размера шариков**, пространственная упаковка которых может повлиять на качество восстановления. На Рисунке 2.39 представлены результаты поиска формы по кривой интенсивности, рассчитанной от структуры 1, построенной из шариков значительно меньшего размера, чем в восстанавливаемых формах. Из Рисунка 2.40 видно, что в этом случае невозможно добиться малой величины невязки на углах, соответствующих  $s > 2$

$\text{нм}^{-1}$  (число шенноновских каналов  $N_S = 8$ ). При восстановлении со старым вариантом штрафа за смещение тела из центра и  $W(s) = I(s) s^2$  (уравнение 2.62) находимые решения оказываются дальше от истинной формы ( $\text{NSD} > 0.9$ , Рисунок 2.39(2)), чем с новым вариантом штрафов за смещение центра и радиус инерции ( $\text{NSD} < 0.7$ , Рисунок 2.39(3)). При этом в обоих случаях качество подгонки данных рассеяния оказывается неудовлетворительным по всему диапазону (Рисунок 2.40, график слева, R-фактор = 60 - 150). Применение весовой функции (2.63) с  $n = 1$  позволяет в данном случае ослабить вклад больших углов в невязку и восстановить форму однородной частицы с приемлемым качеством ( $\text{NSD} < 0.6$ , Рисунок 2.39(4)) фактически только по начальному участку (Рисунок 2.40, график справа). На практике, размер шариков можно выбирать таким, чтобы число узлов в области поиска было  $N_p \sim 1000 N_S$ , где  $N_S$  - число шенноновских каналов (2.28).

Проведенные исследования показывают, что по одномерным данным малоуглового рассеяния можно восстанавливать форму низкого разрешения, специфические детали которой могут варьироваться от решения к решению.

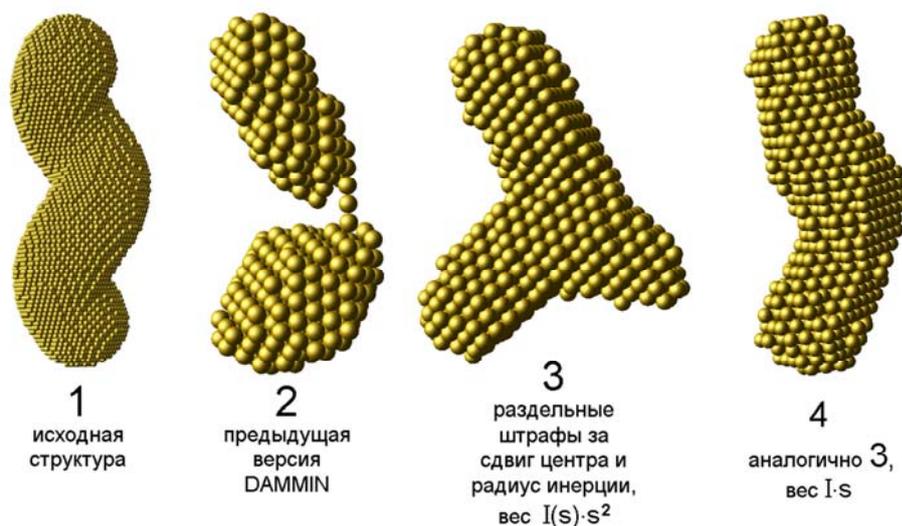


Рисунок 2.39 - Результаты восстановления формы по данным Рисунок 2.40 с разными весовыми функциями для интенсивности и набором штрафов в целевой функции (решения 3 и 4 получены модифицированной программой DAMMIN).

Пояснения даны на рисунке и в тексте.

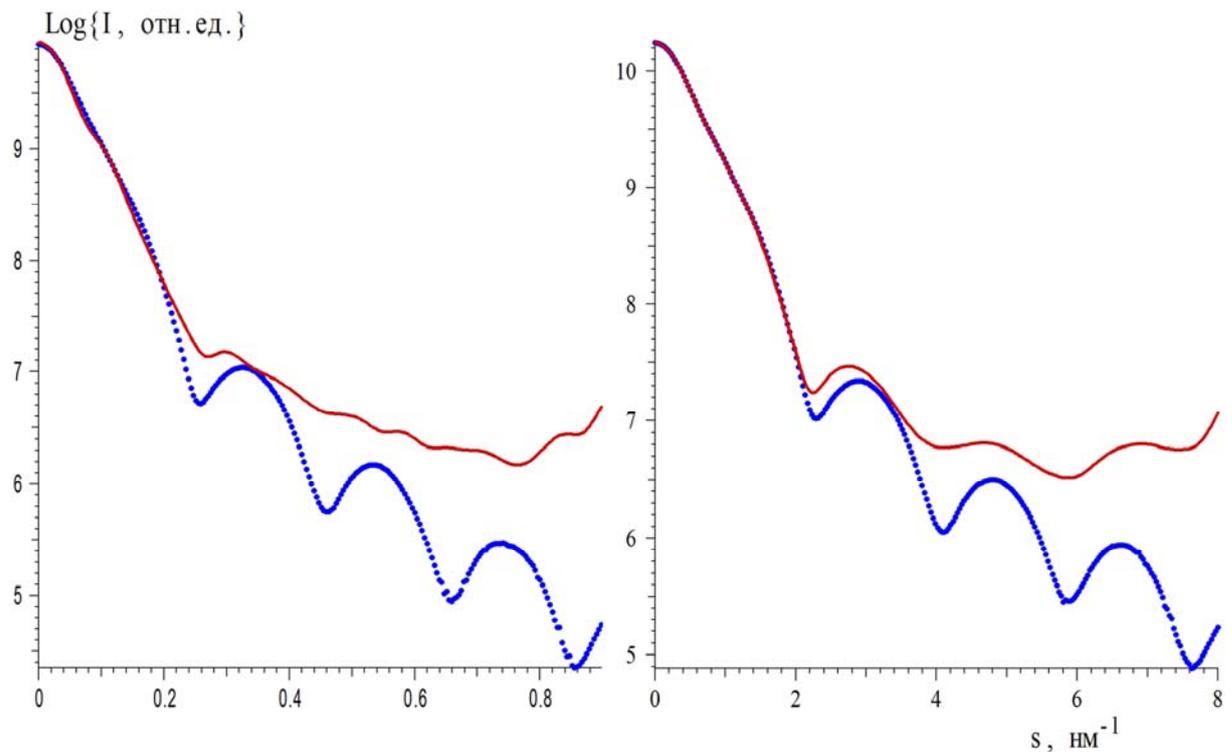


Рисунок 2.40 - Результаты восстановления формы Рисунок 2.39 (1). Точки: рассеяние от структуры 1. Сплошная линия слева - типичное теоретическое рассеяние от найденных моделей 2 и 3 (Рисунок 2.39) и справа - от моделей вида 4. Для модели 4 видно лучшее совпадение на начальном участке рассеяния до  $s = 2$  нм. Расхождение между "экспериментом" и моделью обусловлено рассеянием на неоднородностях упаковки шариков (фон "атомного" рассеяния).

Увеличение детализации формы частиц при близких значениях характеристических размеров деталей и расстояний между ними приводит к неоднозначности решений, как показано на Рисунке 2.41.

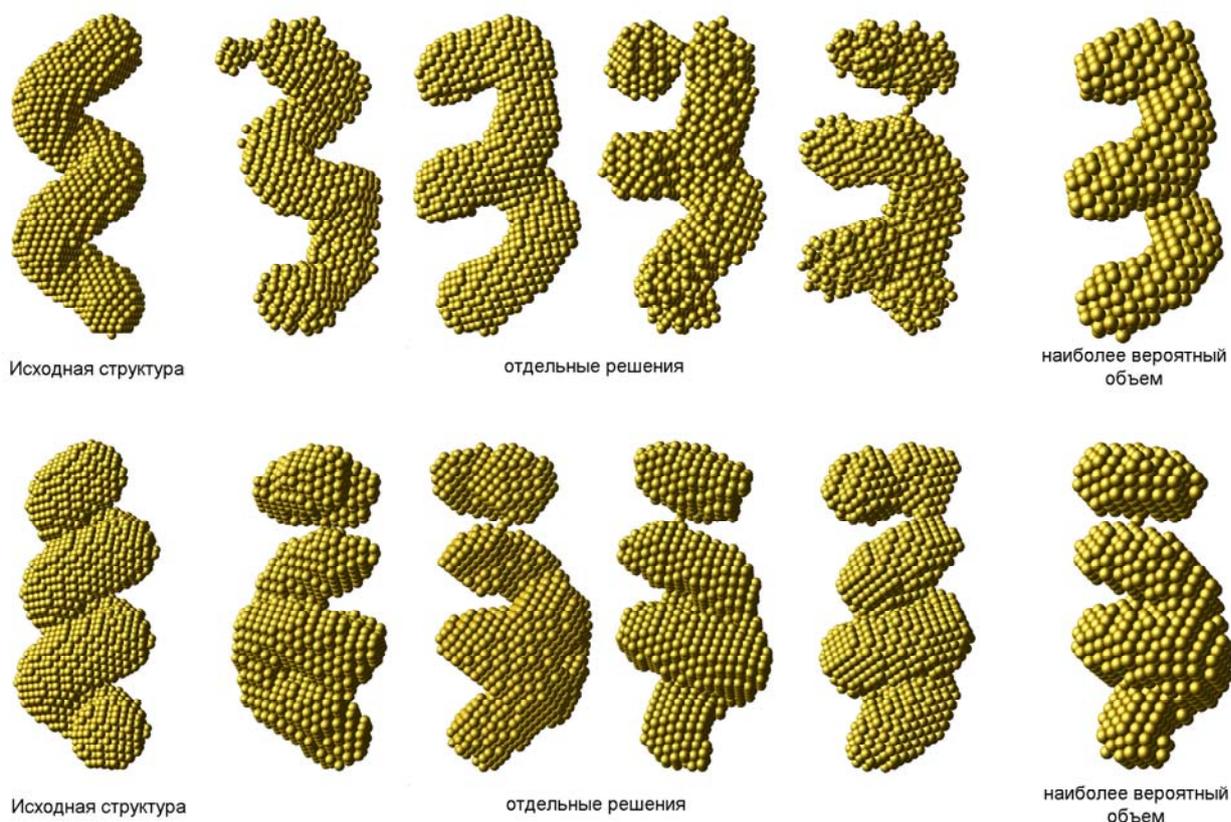
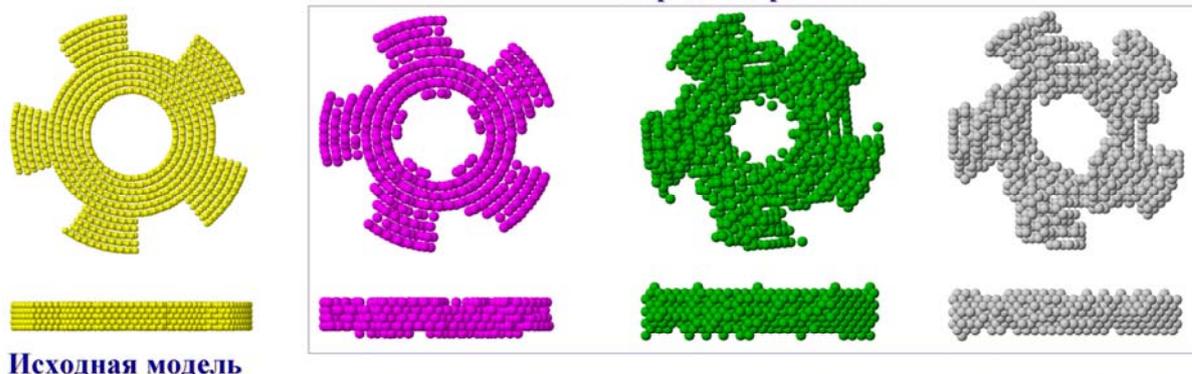


Рисунок 2.41 - Структуры восстанавливаются неоднозначно, решения передают только степень анизотрии и некоторые характерные размеры деталей (толщину нити спирали и число витков в данном случае).

Число шенноновских каналов  $N_S = 14$ .

Как было показано, восстановление плоских структур и структур с большим количеством деталей неустойчиво, если не привлекать дополнительных ограничений на форму. Эффективность использования информации о наличии осей симметрии можно продемонстрировать на восстановлении формы, показанной на Рисунке 2.42. При восстановлении использовали кривую рассеяния от исходной фигуры, рассчитанную на 16 шенноновских каналах (пример типичного диапазона при измерении рассеяния от молекул иммуноглобулина М). Опыт моделирования структур подобного рода в дальнейшем был использован при исследовании морфологии макромолекул иммуноглобулина М и ревматоидного фактора человека.

Типичные решения при использовании оси симметрии 5-го  
порядка в решении



Пример решения без использования дополнительной информации

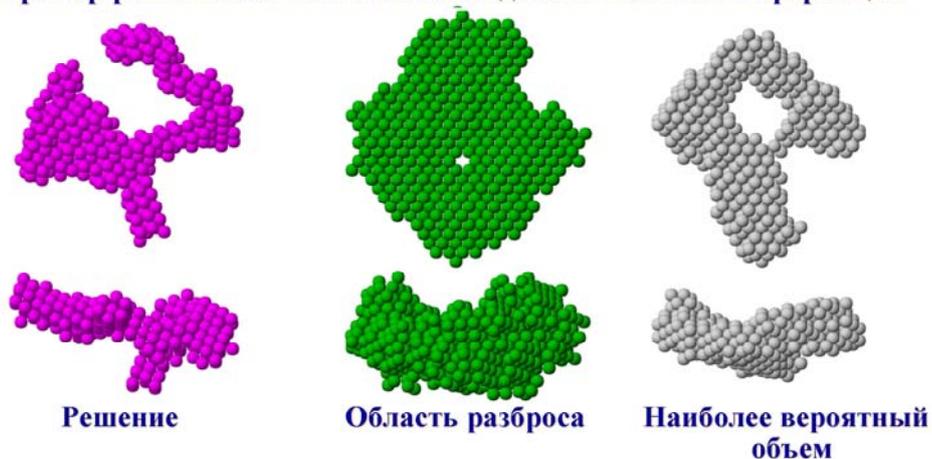


Рисунок 2.42 - Восстановление формы плоской частицы с осью симметрии 5-го порядка. Формы частиц показаны в 2-х ориентациях.

В данном разделе были представлены некоторые результаты численных экспериментов по определению формы частиц из данных малоуглового рассеяния от разбавленных изотропных монодисперсных систем и сделаны основные выводы, которые были учтены при решении реальных задач по определению формы наночастиц и макромолекул. Весь спектр возможных случаев охватить в одном исследовании невозможно и не нужно, так как -

**(VI) - проведение предварительного моделирования может помочь ответить на вопрос о надежности полученных решений в конкретном исследовании.**

Развитые с участием автора программы SASHA (оболочечная модель) и DAMMIN (шариковая модель) широко используются во всем мире и с их помощью были определены формы молекул многих белков в растворе, в том

числе и перечисленные в работах [A16,A18,A20,A21,A59,A] кластеров полимеров [A29] и дендритных полимеров [A36]. В разделе 2.2.4 при исследовании морфологии частицы 30S рибосомы *Thermus Thermophilus* применена также коррекция данных рассеяния методом, рассмотренным в разделе 2.2.2.3, а практические аспекты восстановления формы методом шариковых моделей обсуждаются в разделе 2.2.5.

#### **2.2.4 Исследование структуры белковой фазы частицы 30S *Thermus Thermophilus* в растворе методом малоуглового рассеяния с вариацией контраста**

Методом моделирования шариковыми структурами было исследовано строение субъединицы 30S рибосомы *Thermus Thermophilus* в растворе по данным рентгеновского и нейтронного малоуглового рассеяния, полученных методом вариации контраста [A26]. Измерение протонированных и дейтерированных образцов при нескольких относительных концентрациях D<sub>2</sub>O в растворе, т.е. при различной относительной рассеивающей плотности молекул и их составных частей позволило увеличить информативность экспериментальных данных и определить пространственное размещение белковых компонентов в 30S. Данная работа близка по своему характеру исследованию 70S рибосомы *E-coli*, рассмотренному в разделе 2.2.2.3, за исключением того, что теоретические кривые рассеяния рассчитывались от структурных моделей, состоящих из плотноупакованных сфер (шариков) диаметром 0,5 нм и плотности, соответствующей плотностям белка и РНК. Полученные результаты позволили проверить соответствие кристаллографических данных о структуре 30S ее структуре в растворе.

Данная работа, как и исследование 70S *E.coli*, выполнялась большим коллективом сотрудников из нескольких организаций. Математическое обеспечение для интерпретации данных рассеяния также было создано в Европейской лаборатории молекулярной биологии группой Д.И. Свергуна, разработавшим алгоритм и базовые программы анализа данных рассеяния. Участие автора в данной работе так же, как и в работе, рассмотренной в разделе 2.2.2.3, состояло в применении техники анализа спектров смесей, коррекции набора данных рассеяния и проведении части расчетов структурных моделей.

Набор данных представлял собой 23 кривые рассеяния, измеренные для 11 протонированных и одного дейтерированного образца рибосомы при различных

содержаниях тяжелой воды в растворителе. Устойчивость решения - пространственные координаты виртуальных атомов обеих фаз молекулы - оценивали путем совмещения 24 независимо полученных оптимальных структур и нахождения выборочных распределений занятости пространственных узлов атомами структуры.

#### Выделение 30S рибосомных субъединиц *Thermus thermophilus*.

70S рибосомы *Thermus thermophilus* выделяли по слегка модифицированной методике [172]. Фракции, соответствующие 30S рибосомной субъединице, собирали, концентрировали и диализовали против буферов, содержащих необходимый для экспериментов по малоугловому рассеянию нейтронов процент тяжелой воды. Качество препаратов оценивали седиментационным контролем, а также электрофоретическим анализом РНК и белков, входящих в состав субъединицы.

#### Измерение малоуглового рассеяния.

Измерения нейтронного рассеяния были проведены в лаборатории Risø (Роскилд, Дания), на установке в ILL (Гренобль, Франция), и на реакторе НМІ. Угловой диапазон измерений  $0.07 < s < 2.0 \text{ nm}^{-1}$  достигался путем использования измерений при двух разных расстояниях образец-детектор с последующим объединением обеих кривых рассеяния в одну. Рассеяние рентгеновских лучей измеряли на синхротронной установке X33 EMBL на синхротроне DORIS III, DESY (Гамбург, Германия) в диапазоне  $0.11 < s < 1.7 \text{ nm}^{-1}$ . Список образцов приведен в Таблице 2.5.

Таблица 2.5. Список измеренных образцов рибосомы 30S *Thermus Thermophilus*. Контрасты приведены в единицах  $10^{10} \text{ cm}^{-2}$ .

No	%D <sub>2</sub> O растворе	контраст белка	контраст РНК
протонированная 30S, рассеяние нейтронов			
1	0	2.309	4.245
2	20	1,254	3.067
3	29	0.756	2.581
4	32	0.590	2.373
5	35	0.378	2.207
6	40	0.142	1.907
7	72	-1.622	0.046
8	76	-1.845	-0.189
9	80	-1.819	-0.129
10	98	-3.199	-1.586
дейтерированная 30S, рассеяние нейтронов			
11	0	7.121	7.143
протонированная 30S, рентгеновское рассеяние			
12	0	2.26	4.23

### Моделирование структуры виртуальными атомами.

Молекула рибосомы, как было рассмотрено в разделе 2.2.2.3, состоит из двух компонентов, соответствующих белкам и РНК. Структуру этих компонентов моделировали путем размещения малых сфер, или виртуальных атомов, с радиусом 0,25 нм в узлах плотноупакованной пространственной решетки, в пределах допустимой области поиска. Область поиска представляла собой пространственную оболочку 30S частицы, определенную из электронномикроскопической модели рибосомы 70S *E. coli* (см. раздел 2.2.2.3). Такое ограничение области возможного пространственного расположения виртуальных атомов необходимо для уменьшения степени неопределенности решения при моделировании. Модель молекулы, аналогично алгоритму DAMMIN, рассмотренному ранее, представляла собой конфигурационный вектор

$X$ , элементы которого ассоциированы с определенными узлами плотноупакованной пространственной сетки (т.е. координатами расположения виртуальных атомов) и могут принимать значения 0, 1 или 2, в зависимости от того, описывает ли данный атом область, занятую молекулами растворителя, белка или РНК. Форма области представляла собой увеличенный объем электронномикроскопической модели Франка [163] (см. раздел 2.2.2.3). Такое количество атомов является оптимальным: оно достаточно велико, чтобы модель описывала детали структуры до доступного в малоугловых экспериментах разрешения 3,0-3,5 нм и достаточно мало для того, чтобы численные процедуры определения структуры завершали свою работу в разумное время.

Правильность определения констант, вычитаемых из экспериментальных кривых и множителей для абсолютизации интенсивностей проверяли путем расчета трех базисных функций, соответствующих трем слагаемым в (2.59) при числе фаз в частице, равным двум. Как описано в разделе 2.2.2.3, проводили коррекцию данных к трем компонентам с помощью сингулярного анализа.

По набору экспериментальных данных было проведено 24 независимых поиска структуры. Типичное совпадение экспериментальных и теоретических кривых рассеяния показано на Рисунке 2.43 для нескольких образцов.

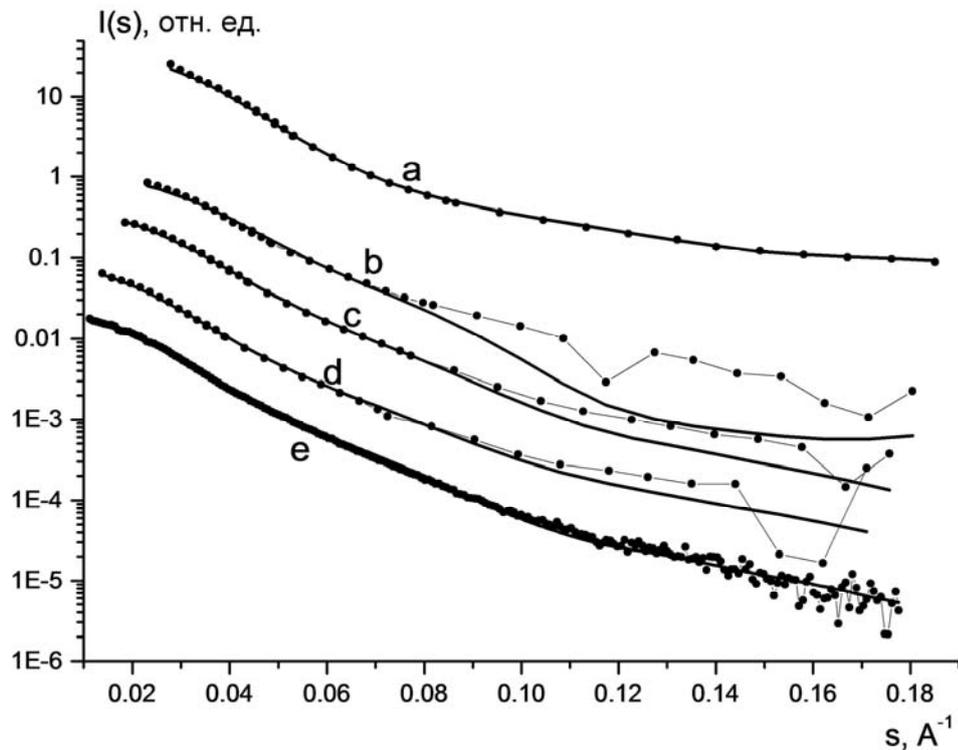


Рисунок 2.43 - Сравнение экспериментальных и теоретических кривых, показанное для пяти образцов растворов рибосомы из набора данных. Точками показаны экспериментальные интенсивности, прямые линии - теоретические кривые. Кривые a-d: нейтронное рассеяние от протонированных частиц рибосомы при доле тяжелой воды в растворителе 0, 32, 72 и 100%, соответственно. (e) - данные рентгеновского рассеяния. Графики смещены по вертикали для лучшей визуализации. Большое расхождение для некоторых данных объясняется малой величиной контраста.

Полученные результаты показали, что структура области, занимаемой РНК получается рыхлой (Рисунок 2.44), в то время как область пространства, занятая белками, формируется в виде нескольких кластеров, которые можно ассоциировать с белковыми молекулами, входящими в состав рибосомы. На первый взгляд, высокая детализация найденных моделей противоречит сделанным в предыдущем разделе выводам о невозможности восстановления деталей высокого разрешения, однако в данном случае проводилось уточнение детализированной электронномикроскопической модели, которая была использована в качестве дополнительной информации.

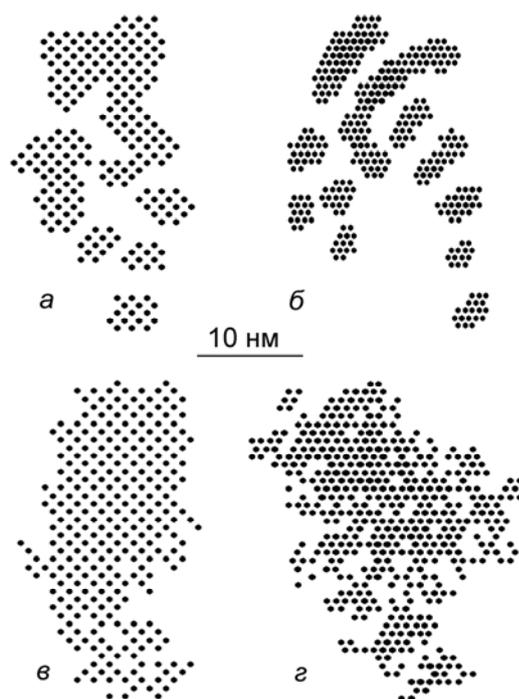


Рисунок 2.44 - Одно из полученных решений, полученное методом моделирования отжига: структура белковой фазы (а,б) и области РНК (в,г) показанные в двух ориентациях.

Для проверки того, насколько устойчива в пространстве такая кластеризация, был произведен подсчет попаданий атомов каждого сорта в узлы пространственной сетки. Область поиска решения была выбрана достаточно анизометричной и поэтому необходимость в предварительном совмещении ориентации структур не возникла. Так как решение для белковой фазы не должно представлять собой связанное тело в силу того, что рибосомные белки располагаются в структуре частицы независимо, штраф за разорванность структуры для белковой части не использовали. Частотная гистограмма попаданий атомов белковой фазы в узлы пространственной сетки от решения к решению показана на Рисунке 2.45. Там же для сравнения показана гистограмма попаданий для такого же набора случайных структур, состоящих из 500 виртуальных атомов, расположенных в такой же пространственной области из 2600 плотноупакованных узлов. Видно, что для полученного набора решений гистограмма имеет четко выраженный максимум при 8 попаданиях из 24, что

свидетельствует о систематическом восстановлении пространственной структуры, в то время как случайные структуры не дают распределение с максимумом. Вычислить соответствующие вероятности в данном случае невозможно, так как невозможно учесть вероятностный вклад в расположение атомов штрафов за наличие отдельно стоящих атомов. Однако, можно предположить, что наличие 53 узлов в гистограмме (последняя точка), которые оказались занятыми в 24 случаях из 24 решений чрезвычайно маловероятное событие, если бы узлы занимались случайным образом. Типичное значение параметра штрафа за наличие оторванных атомов для белковой фазы в решениях было около 0,03, что соответствует компактному расположению атомов, тогда как для РНК-структур типичным было значение 0.08, т.е. структурные элементы располагались более диффузно. В качестве решения предлагается использовать объединенную модель, описываемую только теми узлами, в которых атомы из 24 решений попадали чаще всего. Критерием отбрасывания редко заселенных узлов является ожидаемый объем белковой фазы, известный из химического состава. Соответствующая структура показана на Рисунке 2.46. Несмотря на то, что кривая рассеяния от такой структуры хуже приближает экспериментальные данные (так как она получена не в результате процедуры минимизации невязки), такая модель отражает устойчивую при поиске часть структуры белковой фазы. Сравнение с известной кристаллографической моделью рибосомы (Schluenzen, F. Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution / F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschisend, A. Yonath // Cell. – 2000. – V. 102. – No. 5. – P. 615–623.) показывает значительное несоответствие в пространственном расположении белков (Рисунок 2.46), однако, малоугловая модель получена от образцов, в которых молекулы находятся в состоянии, близком к естественному и полученные данные могут использоваться при уточнении кристаллографической модели.



Рисунок 2.45 - Частотные гистограммы занятости узлов пространственной сетки в области поиска виртуальными атомами белковой фазы накопленные для 24 независимых решений (кривая а). б - аналогичное распределение полученное для 24 случайных структур, расположенных в той же области.

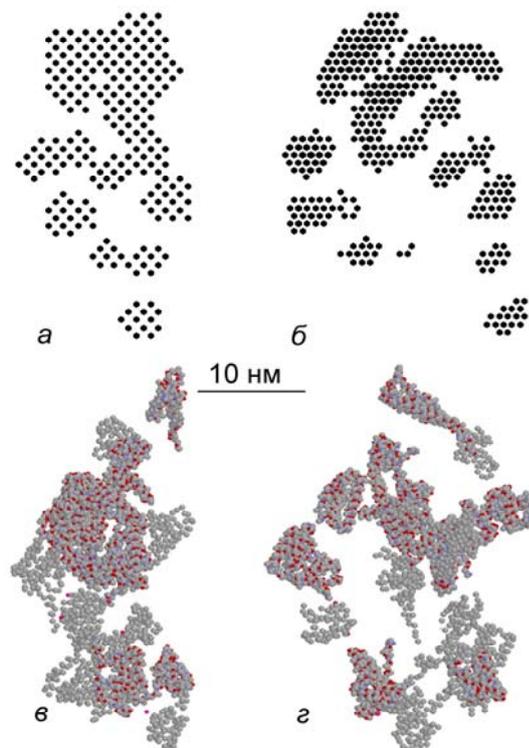


Рисунок 2.46 - Среднестатистическая модель белковой фазы рибосомы, полученная путем объединения 24 независимых решений (а,б) и ее сравнение с известной кристаллической структурой (с,д).

Таким образом, метод восстановления структуры представленной пространственным расположением виртуальных атомов применен к набору данных малоуглового нейтронного и рентгеновского рассеяния от растворов рибосомы с вариацией контраста. Построена среднестатистическая модель пространственного расположения белковой фазы молекулы, которая отличается от известной кристаллографической модели. Определенные отличия в моделях могут служить дополнительной информацией для корректировки кристаллографической структуры.

### 2.2.5 Исследование морфологии макромолекул иммуноглобулинов и ревматоидного фактора человека по данным малоуглового рассеяния от раствора

Иммуноглобулины представляют собой семейство гликопротеинов, содержащихся в организме млекопитающих. Основная функция большинства из них состоит в специфическом распознавании и связывании чужеродных объектов (антигенов) и осуществлении эффекторных функций [173]. У большинства высших млекопитающих обнаружено пять классов иммуноглобулинов, которые отличаются размерами молекул, зарядом, аминокислотным составом и содержанием углеводов. Иммуноглобулин G (**IgG**) является главным классом, на долю которого приходится 70-75% общего количества сывороточных иммуноглобулинов. На долю **IgM** приходится примерно 10% общего пула иммуноглобулинов сыворотки.

Ревматоидный иммуноглобулин M (**IgM-RF**) относится к группе ревматоидных факторов (**RF**), которые являются аутоантителами против иммуноглобулина G собственного организма. Полученные за последнее время данные позволяют утверждать, что IgM-RF, как и другие RF, играет важную роль в патогенезе ревматоидного артрита и таких аутоиммунных заболеваний, как системная красная волчанка, синдром Сьогрена, а также при хронических инфекционных заболеваниях. Одновременно с патологической функцией RF играют заметную роль и в нормальном функционировании иммунной системы – участвуют в выведении иммунных комплексов, в процессинге антигена В-клетками, в становлении первичного антительного репертуара [174]. Отсюда вытекает особый интерес к исследованию связей между структурой и функцией RF (в особенности, IgM-RF, который доминирует среди RF по ряду аспектов). Установление взаимосвязи между структурой и патологической функцией IgM-RF является актуальной проблемой, связанной с разработкой подходов к

диагностике и лечению ревматоидного артрита и других аутоиммунных заболеваний.

В последние годы получена пространственная структура нескольких интактных моноклональных IgG, с разрешением 0,28 – 0,4 нм [175-177,192]. Иммуноглобулины других классов, такие как IgA, IgD, IgE и IgM, до сих пор не удалось закристаллизовать и их молекулярная структура исследуется методами электронной микроскопии, атомно-силовой микроскопии (АСМ), ЯМР и другими физико-химическими методами. Большинство методов предполагают специальную обработку образцов, что, в свою очередь, может приводить к изменению структуры исследуемых молекул.

В структурном отношении как иммуноглобулин М (IgM), так и ревматоидный IgM-RF являются большими макромолекулами, состоящими из пяти IgG-подобных мономеров (Рисунок 2.47).

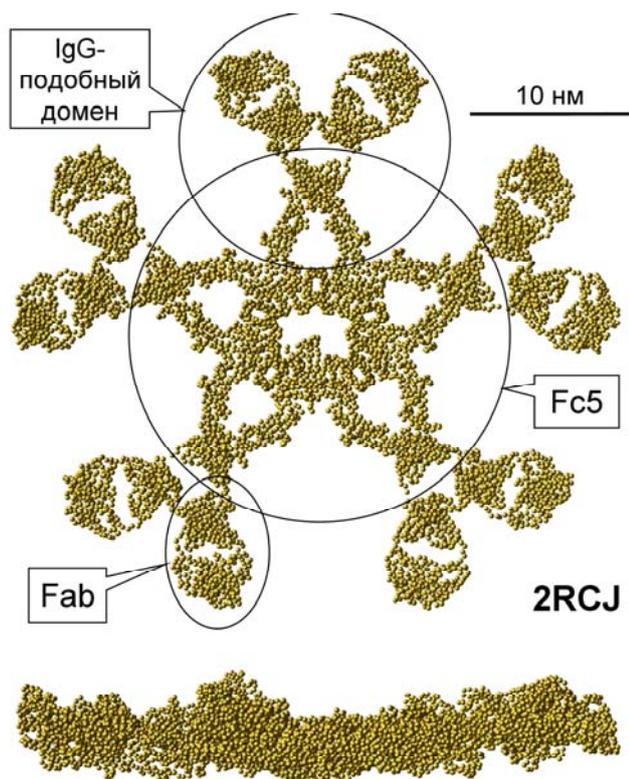


Рисунок 2.47 - Модель 2RCJ молекулы IgM по литературным данным [178]. Структура показана в 2-х ориентациях.

Пространственное строение IgG было ранее изучено с разрешением 2.8–4.0 Å благодаря возможности их кристаллизации и использованию кристаллографических рентгеновских методов [177]. Как IgM, так и IgM-RF не кристаллизуются в силу высокой подвижности доменов, а представленная на Рисунке 2.47 плоская пентамерная модель структуры **2RCJ** была предложена Перкинсом и соавторами [178] на основании анализа данных малоуглового рентгеновского рассеяния от растворов IgM известной кристаллической структуры доменов молекул IgG. Эта модель хорошо согласуется как с гидродинамическими и седиментационными данными, так и с электронными микрофотографиями молекул иммуноглобулина [179-181].

Однако до настоящего времени принципиальных структурных различий между IgM и IgM-RF на уровне их моноклональных представителей не обнаружено. Предыдущие исследования [182] позволили предположить, что различия следует искать в первую очередь в Fab-областях, поскольку именно там расположены антиген-связывающие центры, определяющие специфичность функционирования антитела.

Наши исследования, которые будут рассмотрены далее, выявили некоторые структурные различия между Fab-областями в IgM и IgM-RF [A27,184], а в [185] методом электронного парамагнитного резонанса (ЭПР) с использованием спин-метки показано, что Fab-фрагмент неревматоидного IgM обладает жесткой структурой. Метод спин-метки оказался весьма эффективным инструментом для обнаружения даже незначительных изменений как общей, так и локальной конформационной подвижности макромолекул, в том числе и в иммуноглобулинах [186]. Кроме того, метод ЭПР, как и малоугловое рентгеновское рассеяние, позволяет исследовать макромолекулы в растворе при условиях, близких к естественным. Из работ [187,188] следует, что в Fab-RF существуют два центра связывания антигена. Один из них предназначен для взаимодействия с антигеном и соответствует обычной антительной специфичности данного Fab-RF. Другой центр связывает только аутоантиген.

Учитывая сказанное, в данной работе было предпринято комплексное изучение структурных особенностей IgM, IgM-RF, Fab и Fab-RF-фрагментов методами малоуглового рентгеновского рассеяния (МУР), спин-метки и масс-спектрометрии (МС), что позволило получить новые данные.

Плоская структура макромолекулы IgM, как было показано в разделе 2.2.3.2, не может быть восстановлена без привлечения дополнительной информации. В качестве такой информации в данной работе использовали наличие оси симметрии 5-го порядка. Для определения устойчивости определения формы частицы был проведен ряд численных экспериментов, результаты которых рассмотрены в следующем разделе.

### 2.2.5.1 Исследование устойчивости определения формы молекулы IgM

Для определения стабильности восстановления формы была рассчитана кривая рассеяния от структуры 2RCJ, показанной на Рисунке 2.47. Так как влияние случайных шумов в силу высокой гладкости кривой малоуглового рассеяния незначительно, их влияние не моделировали.

Учитывая факт зависимости решения от диапазона экспериментальных данных, продемонстрированный в разделе 2.2.3.2, численные эксперименты проводили по диапазонам данных, соответствующих числу шенноновских каналов (2.28)  $N_S = 12 - 25$ , которое типично для экспериментальных кривых рассеяния от данных образцов.

Здесь следует сделать **важное для практики замечание**. Вариация деталей формы от решения к решению для структур с большим количеством мелких деталей приводит к тому, что усреднение решений с помощью программы DAMAVER приводит к наиболее вероятному объему, слишком далекому от самих решений ( $NSD > 0.8$ ), тогда как решения отличаются друг от друга не более, чем на  $NSD = 0.7$ . Особенно сильно этот эффект проявляется при увеличении диапазона данных до  $N_S > 20$ .

При моделировании и поиске решений по данным малоуглового эксперимента область поиска назначали в виде цилиндрического объема, имеющего форму диска с отношением диаметр/толщина около 3 - 4. Это обусловлено тем, что молекулы IgM имеют плоскую структуру и сферическая область поиска имеет избыточное число узлов, что приводит к неоправданному увеличению их числа и, соответственно, к значительному увеличению машинного времени. Форма IgM восстанавливается в сферической области в виде плоского тела, как показано на Рисунке 2.48(а,б), и пропорции цилиндрической области были выбраны по результатам поиска в сфере с учетом возможного разброса решений по толщине структур.

Результаты восстановления показаны на Рисунке 2.49 и 2.50. Видно, что при увеличении углового диапазона данных рассеяния восстанавливаемые формы приобретают мелкие детали, которые передают "ажурность" структуры, но их расположение практически хаотично. Общий характер формы низкого

разрешения при этом сохраняется. В любом случае совпадение между исходной кривой рассеяния и рассчитанными по найденным моделям не превышает  $R = 0.0010 \pm 0.0001$  и соответствует Рисунку 2.48(a).

По результатам расчетов около 300 модельных вариантов были подобраны оптимальные значения весовых коэффициентов штрафных членов в (2.64) для рыхлости структуры  $w_L = 0.001$  и  $w_C = 0.005$ . Эти значения использовали для расчетов с помощью модифицированной программы DAMMIN.

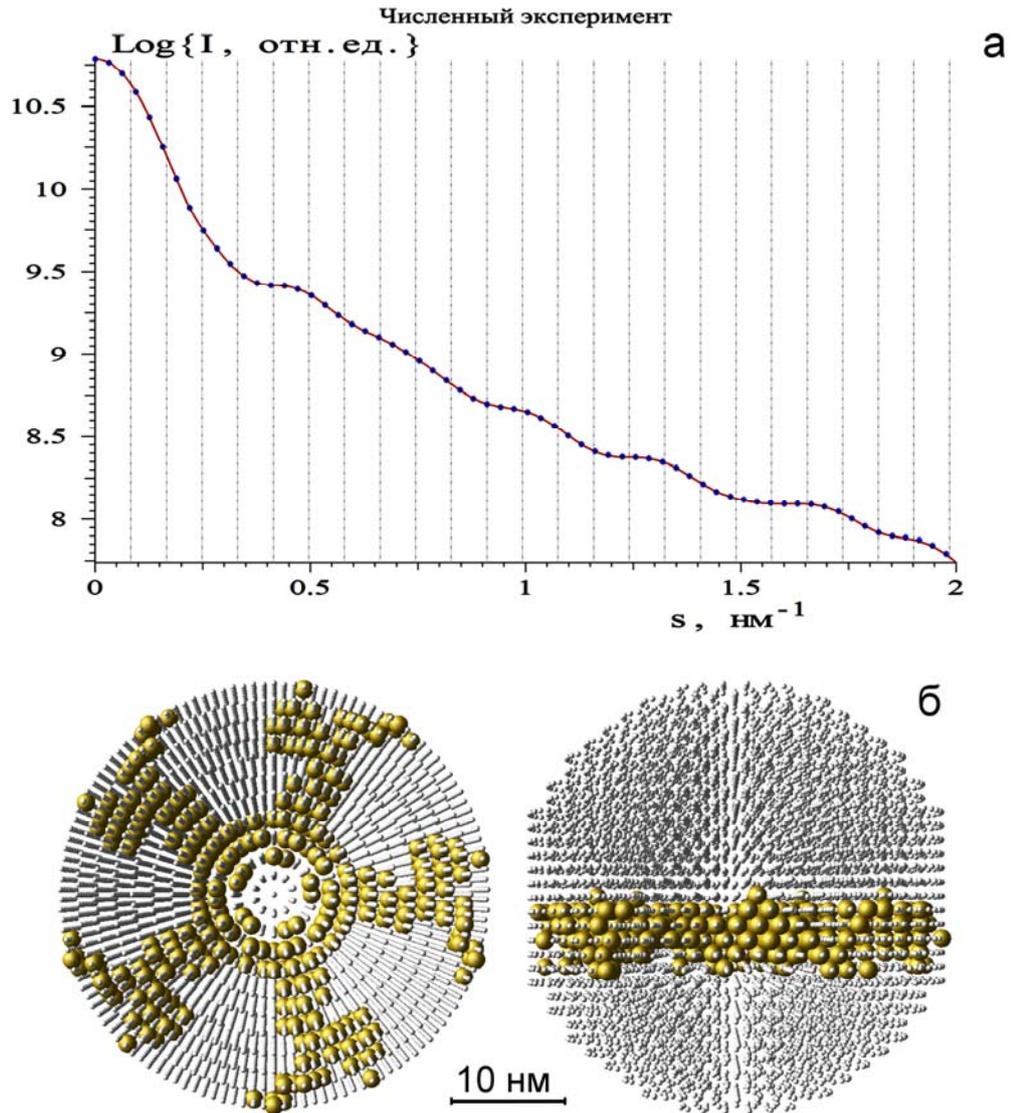


Рисунок 2.48 - Пример восстановления формы молекулы IgM в сферической области поиска диаметром 38 нм по теоретической кривой рассеяния от известной структурной модели. а): сравнение исходной интенсивности рассеяния от 2RCJ (точки) и рассеяния от найденных решений (сплошная линия). Вертикальными пунктирными линиями отмечены позиции шенноновских каналов. б): найденная структура в 2-х ориентациях: желтые шарики - частица, серые - область растворителя.

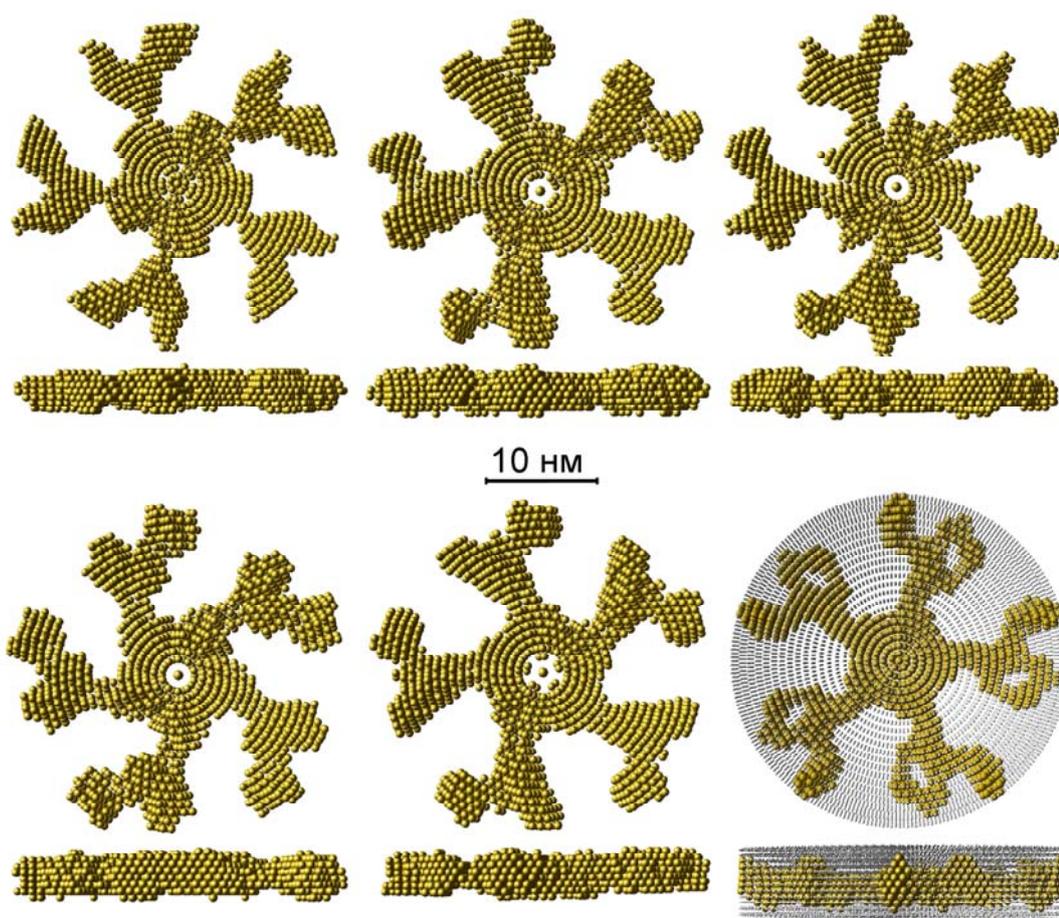


Рисунок 2.49 - Пример шести восстановлений формы 2RCJ (Рисунок 2.47) по модельным кривым рассеяния при числе шенноновских каналов  $N_S = 12$ . NSD структур  $0.6 \pm 0.1$ . Структура справа внизу показана вместе с областью поиска (серые точки). Структуры здесь и на следующих рисунках показаны в двух ориентациях.

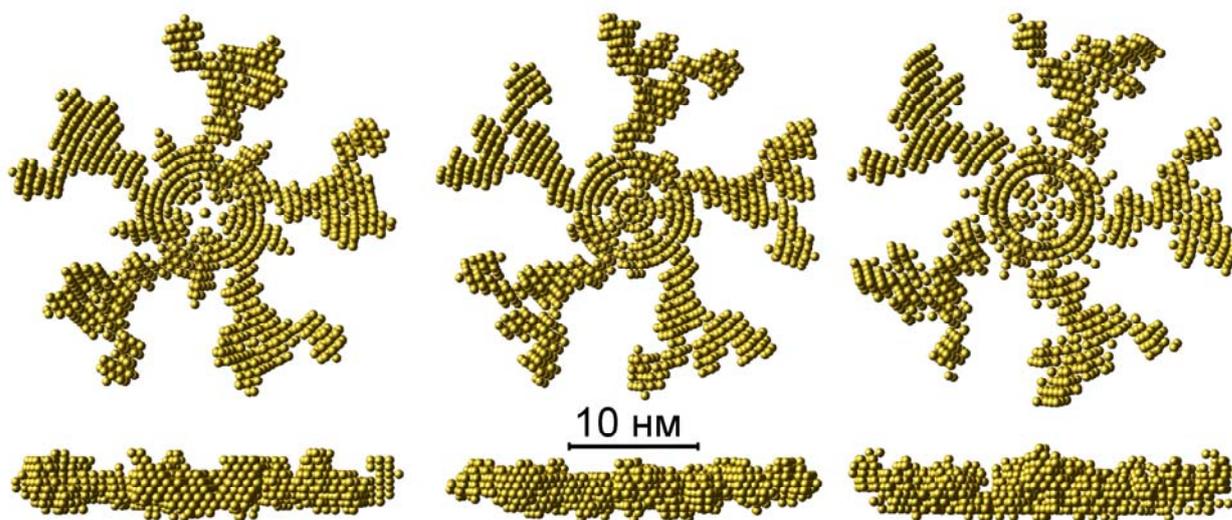


Рисунок 2.50 - Пример восстановлений формы 2RCJ (Рисунок 2.47) по модельным кривым рассеяния при числе шенноновских каналов  $N_S = 25$ . NSD структур  $0.65 \pm 0.10$ .

### 2.2.5.2 Результаты исследования растворов молекул иммуноглобулинов и ревматоидного фактора

Данная работа была выполнена сотрудниками Института органической химии РАН, Института молекулярной биологии РАН, Института биоорганической химии РАН, Москва и Гематологического научного центра РАМН. Измерения малоуглового рентгеновского рассеяния и анализ данных проводились в Европейской лаборатории молекулярной биологии (EMBL с/o DESY, Гамбург) и в Институте кристаллографии РАН [A27, A31, A44, A61, A63]. Автором были проведены измерения рассеяния, обработка и интерпретация основной части полученных данных, модификация программ и разработка ряда новых процедур анализа интенсивности малоуглового рассеяния.

Методики, использованные в данной работе, кратко описаны ниже.

#### Получение растворов иммуноглобулинов.

Для получения растворов иммуноглобулина IgG использовали препарат фирмы "Sigma" (США). Моноклональный IgM человека и моноклональный IgM-RF (болезнь Вальденштрема, в дальнейшем обозначены как образцы IgM-RF1, IgM-RF2 и IgM-RF3) были получены по разработанным ранее методикам выделения из плазмы крови человека [188]. Для приготовления растворов этих белков с концентрациями от 2.5 до 14.0 мг/мл использовали трис-буфер pH 8,3-8,4 фирмы "Merck" (Германия). Концентрацию белка в растворе измеряли спектрофотометрически, используя величины поглощения  $A_{280,1cm}^{1\%} = 12.0$  для IgM и IgM-RF, и 14.0 для IgG. Уровень агрегации белка в растворе оценивали с помощью аналитической ультрацентрифуги "Spinco E" фирмы Beckman (США). Полученные величины содержания агрегатов не превышали 5%. Непосредственно перед экспериментом все растворы пропускали через мембранный фильтр с величиной пор 0.45 мкм фирмы "Schleicher & Schuell" (Германия). Препараты оказались полностью гомогенными по данным иммуноэлектрофореза с анти- $\gamma$ ,

анти- $\mu$  и антицельной антисыворотками, а также по данным аналитического ультрацентрифугирования.

Образцы IgM-RF $\eta$  выделены из плазмы крови потенциального больного с необычайно высокой ревматоидной активностью IgM $\kappa$  – 32600 IU/ml (при норме 0 - 20).

Fab- и Fab-RF-фрагменты были получены соответственно из IgM и IgM-RF по методу [188]. Они также были проанализированы методом аналитического ультрацентрифугирования и оказались полностью гомогенными. Непосредственно перед введением спиновой метки для исследования методом ЭПР образцы Fab- и Fab-RF-фрагментов были сконцентрированы с помощью миниконцентратора Minicon B-15 (фирма Amicon, США).

#### Измерение малоуглового рассеяния.

Часть данных малоуглового рентгеновского синхротронного рассеяния была получены на лабораторном дифрактометре АМУР-К (разработка СКБ института кристаллографии РАН, Москва) при фиксированной длине волны излучения  $\lambda$ , равной 0.1542 нм (CuK $\alpha$  линия) с блочным коллиматором по схеме Кратки. Угловой диапазон измерений составлял  $0,12 < s < 9,0 \text{ нм}^{-1}$ . Основную часть образцов измеряли на станции X33 в Европейской лаборатории молекулярной биологии (EMBL) на накопительном кольце синхротрона DORIS III, (DESY, Гамбург, Германия). Необходимость использования синхротронного излучения определялась высокой интенсивностью рентгеновского пучка, что позволило получить данные с хорошим отношением сигнала к шуму, которого невозможно добиться на лабораторных установках с рентгеновскими трубками из-за слабой рассеивающей способности растворов биомолекул. Угловой диапазон измерений составлял  $0.1 \text{ нм}^{-1} < s < 5.0 \text{ нм}^{-1}$ , при  $\lambda = 0.15 \text{ нм}$ . Для того, чтобы перекрыть указанный угловой диапазон, измерения проводили при двух расстояниях образца от детектора - 2.9 м и 4 м. Для получения полной картины рассеяния оба набора данных объединяли. Образцы для измерения помещали в 1 мм кювету со слюдяными окнами толщиной 20 мкм (станция X33) или в кварцевый капилляр

диаметром 1 мм со стенками 10 мкм (АМУР-К). Измерения нормировали относительно интенсивности падающего пучка и делали соответствующие поправки на чувствительность детектора. После вычитания рассеяния от кюветы с чистым растворителем результирующие кривые нормировали относительно концентраций растворов.

#### Определение интегральных параметров молекул IgM и IgM-RF в растворе.

Радиусы инерции молекул  $R_g$  вычисляли по данным МУР, используя приближение Гинье, а также определяли из функции распределения по расстояниям  $p(r)$ , как описано в предыдущих разделах.

#### Масс-спектрометрия Fab и Fab-RF фрагментов.

Определение молекулярных масс как самих Fab и Fab-RF фрагментов, так и продуктов их расщепления трипсином проводилось методом масс-спектрометрии. Целью эксперимента было обнаружение предполагаемых различий в масс спектрах как самих Fab и Fab-RF фрагментов, так и продуктов их деградации трипсином. Трипсинолиз препаратов Fab и Fab-RF проводили однотипно по следующей методике. Белок переводили диализом в 0.3 М трис-НСI буфер рН 8.0, содержащий 6М гуанидин-хлорид и 5 мМ дитиотрейит и инкубировали 30 мин при 60°C. Затем раствор охлаждали до комнатной температуры и добавляли иодацетамид до концентрации 40мМ. После инкубации в темноте в течение 60 мин при комнатной температуре разбавляли водой квалификации milli-Q в 4 раза и добавляли бычий ТРСК-трипсин, инкубировали при 37° в течение 12 ч в темноте, разбавляли в 2 раза 0.2%-ной ТФУ и обессоливали на картридже для твердофазной экстракции. Съемку масс-спектров проводили на приборе Ultraflex Tof/Tof (Bruker) при способе ионизации MALDI-TOF-MS. В качестве матрицы использовали 2,5-дигидробензойную кислоту. Метод нанесения образца на мишень – dried-droplet.

#### Введение спиновой метки и измерения спектров электронного парамагнитного резонанса.

Были использованы следующие препараты: спиновая метка 2,2,6,6-тетраметил-4-дихлортриазиниламинопиперидин-1-оксил (синтезирован А.Ю. Мишариным); Сефадекс G-25 средний (фирма Pharmacia, Швеция); 0.01 M *трис*-HCl буфер, pH 8.3 – 8.4, содержащий 0.1M NaCl и 0.02% NaN<sub>3</sub> (буфер А). Введение спиновой метки в образцы Fab-фрагментов проводилось по следующей методике: 2 мкл раствора спиновой метки в этаноле (0.1M), приготовленного при комнатной температуре, были добавлены к 200 мкл раствора белка в буфере А (10 мг/мл). Диализованный раствор был пропущен через колонку с Сефадексом G-25 средний (0.5 × 20 см) в буфере А при скорости элюирования 15 мл/ч, фракции элюата по 150 мкл. Для каждой фракции были измерены спектры ЭПР, а для пиковых фракций – еще и при различных температурах и вязкости среды.

Спиновая метка в Fab-RF-фрагменты была введена аналогичным путем. Воспроизводимость данных, представленных в настоящей работе, была подтверждена повторным проведением экспериментов.

Измерения проводили на ЭПР-спектрометре Varian E-104A (США) при амплитуде высокочастотной модуляции магнитного поля 1-2 Гс, мощности СВЧ 5 мВт, при различных значениях коэффициента усиления и температурах 1 и 20°C. Развертка магнитного поля составляла 100 и 200 Гс при времени сканирования 8 и 4 мин, соответственно. Крайние широкие пики ЭПР спектров были измерены отдельно при высокочастотной модуляции амплитуды 2 – 4 Гс и с более высоким усилением для определения расстояния между максимумами высокопольных и низкопольных пиков.

В данной работе проводили сравнение релаксационных свойств в растворе глобулярных макромолекул Fab- и Fab-RF-фрагментов, спин-меченных по пептидной части. В методе спин-метки информацию о структуре макромолекулы получают из анализа двух величин: времени вращательной корреляции  $\tau$  (в нс) и параметра упорядоченности  $S$ . Величина  $\tau$  характеризует пространственную подвижность спин-меченого участка и позволяет определить его молекулярную массу. Чем меньше  $\tau$ , тем меньше молекулярная масса спин-меченого объекта.

Величина  $S$  характеризует ближайшее окружение спин-метки и может принимать значения  $\leq 1$ . Если  $S = 1$ , спин-метка полностью неподвижна относительно белковой глобулы и чем меньше  $S$ , тем более подвижна спин-метка.

Для монодисперсных фракций обоих спин-меченых препаратов было проведено исследование температурно-вязкостной зависимости спектров ЭПР, аналогично [183]. Вязкость белковых растворов варьировали добавлением сахарозы. Величину вязкости контролировали с помощью рефрактометра Carl Zeiss (Йена, Германия). На основании измерения расстояний между крайними широкими пиками в спектрах ЭПР, записанных при 1 и 20°C и переменной вязкости среды, были проанализированы зависимости указанного расстояния от вязкости среды для обоих спин-меченых препаратов.

Эффективную величину вращательной корреляции  $\tau$  рассчитывали по формуле

$$\tau S = 197 \left( \frac{2\bar{A} - 2A'}{S} \right)^{-1.39}, \quad (2.66)$$

где  $S$  – параметр упорядоченности (связанный с частичным усреднением магнитных тензоров спин-метки),  $2\bar{A}$  – точка пересечения с осью ординат прямых участков зависимостей расстояний между крайними широкими пиками в спектре ЭПР  $2A'$  от температурно-вязкостного параметра  $(T/\eta)^{0.74}$ .

Предварительно, по точке пересечения этих прямых с осью ординат рассчитывали параметр упорядоченности  $S$ :

$$S = \frac{2\bar{A} - 2a_0}{2A_{ZZ} - 2a_0}, \quad (2.67)$$

где  $a_0$  – изотропная константа сверхтонкой структуры нитроксила, определяемая как расстояние между соседними острыми пиками в спектрах ЭПР свободной спин-метки в растворе,  $2A_{ZZ}$  – расстояние между крайними широкими пиками в спектрах ЭПР спин-меченого препарата при 77°K.

### Интегральные структурные параметры иммуноглобулинов.

Средние значения молекулярных масс иммуноглобулинов для образцов с различными концентрациями и при различных условиях съемки (длина камеры 4 м и 2.9 м) по данным малоуглового рассеяния составляли : IgG -  $146 \pm 10$  кДа; IgM -  $760 \pm 40$  кДа; IgM-RF -  $690 \pm 20$  кДа. Согласно литературным данным, молекулярный вес IgG и IgM, соответственно, равен 146 кД и 970 кД [191]. Несмотря на значительную ошибку в полученных значениях молекулярных масс, величина этого параметра для ревматоидного фактора оказалась в целом систематически немного меньше, чем для IgM. Надо сказать, что заниженные оценки масс могут свидетельствовать и о меньшей степени агрегации молекул IgM-RF в растворе. Радиусы инерции, рассчитанные по экспериментальным кривым, отражают ту же тенденцию: для молекулы ревматоидного фактора  $R_g$  меньше, чем для IgM. Найденные с помощью программы GNOM максимальные размеры молекул  $D_{\text{макс}}$  равны  $17 \pm 1$  нм,  $38 \pm 0.5$  нм и  $38 \pm 2$  нм, соответственно, для IgG, IgM и IgM-RF. Средние значения найденных радиусов инерции  $R_g$  для вышеприведенных условий составляют соответственно: для IgG  $5.9 \pm 0.1$  нм, для IgM  $11.4 \pm 0.6$  нм и для IgM-RF  $10.9 \pm 0.4$  нм. Согласно литературным данным, средние значения  $R_g$  составляют соответственно для IgG  $5.84 \pm 0.04$  нм [175], для IgM  $12.0 \pm 0.2$  нм [192].

### Структура IgG в растворе.

Молекулы IgG имеют типичную для антител структуру и состоят из 4-х полипептидных цепей - двух одинаковых легких цепей и двух одинаковых тяжелых цепей (с переменными и константными областями), которые образуют в целом 12 доменов и формируют две Fab- и одну Fc-области. Присущая иммуноглобулинам подвижность может дать целый спектр молекулярных конформаций. В соответствии с кристаллографическими данными [177], молекула IgG в кристалле может иметь как симметричные, так и асимметричные T и Y конформации. В растворе антитела могут принимать промежуточные между этими основными конформациями формы. На Рисунке 2.51 представлены кривые

малоуглового рентгеновского рассеяния от раствора IgG вместе с кривыми рассеяния, рассчитанными для опубликованных кристаллических структур IgG<sub>1k</sub> (Mab61.1.3, код структуры 1IGY) и IgG<sub>2ak</sub> (Mab231, код структуры 1IGT) [192,193]. Значение  $\chi$  для 1IGY, равное 2,1, меньше, чем для 1IGT (2,8). Значительное отклонение этих значений от единицы связано с тем, что статичная структура высокого разрешения содержит много деталей, тогда как в растворе мы наблюдаем молекулы, находящиеся одновременно в разных конформациях, и усредненная кривая рассеяния уже не может соответствовать кристаллической модели.

Меньшее значение критерия согласия для структуры 1IGY позволило предположить, что в растворе форма молекул IgG в ближе к Y-конформации.

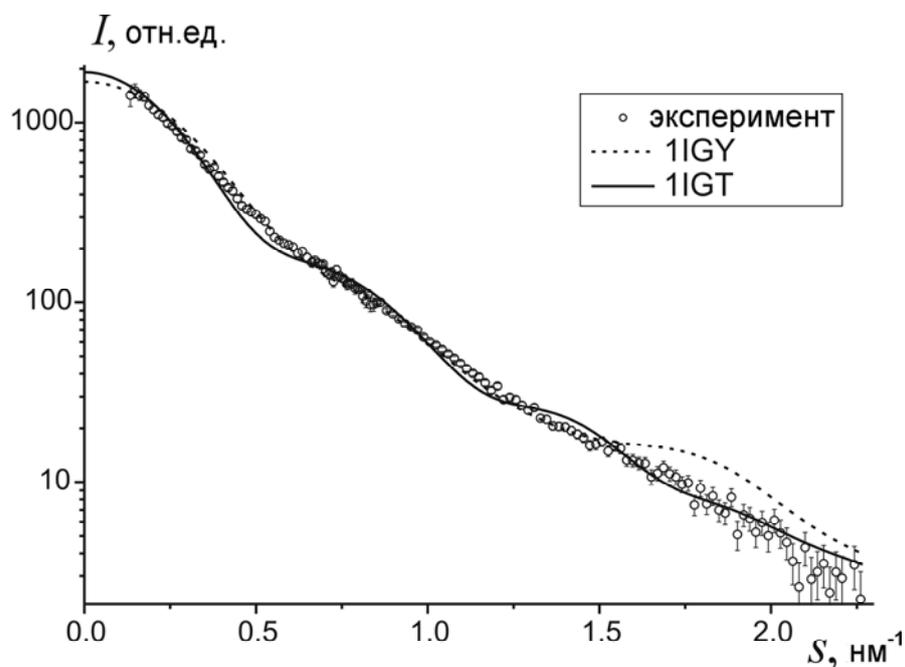


Рисунок 2.51 - Сравнение экспериментальной кривой малоуглового рассеяния от раствора иммуноглобулина IgG (кружки с разбросом ошибок) и кривых, рассчитанных от известных структур высокого разрешения: IgG<sub>1k</sub> Mab61.1.3, код структуры 1IGY (пунктир) и IgG<sub>2ak</sub> Mab231, код структуры 1IGT (сплошная кривая).

Для проверки этого предположения было проведено определение формы частиц IgG с помощью программы DAMMIN по полученным экспериментальным данным рассеяния. Максимальный размер макромолекул IgG был оценен из

функции распределения по расстояниям  $p(r)$ . Этот параметр оказался равен 17 нм. Для определения устойчивости решения было проведено несколько независимых восстановлений структуры белка. Для восстановления формы IgG в растворе при решении использовалась произвольно расположенная ось симметрии второго порядка. Одно из типичных решений вместе с кристаллографической моделью представлены на Рисунке 2.52 в двух взаимно перпендикулярных ориентациях. Для сравнения на том же рисунке представлена форма IgG, восстановленная без какого-либо ограничения на анизометрию частицы. Соответствие модельных кривых рассеяния экспериментальным данным показано на Рисунке 2.53.

Полученные структуры, включая случай восстановления без ограничений на симметрию макромолекул отличаются от атомной модели 1IGY, большими линейными размерами: 17 нм вместо 14.7 нм. Найденный размер соответствует максимальному расстоянию в структуре 1IGT 16,4 нм с учетом поправки на толщину гидратной оболочки 0,3 нм. Это позволяет сделать предположение, что в растворе молекулы IgG имеют более раскрытую конформацию, чем 1IGY. Найденные программой DAMMIN V-образные модели отличаются, в свою очередь, от формы 1IGT и ближе к Y-типу. Рассеяние от полученных шариковых моделей совпадает с экспериментом без систематических отклонений с  $\chi = 0.8 \pm 0.1$  (Рисунок 2.53). Экспериментальная кривая отличается от теоретических профилей рассеяния, рассчитанных для атомных структур, большей гладкостью. Это можно объяснить тем, что, как уже было отмечено, в растворе одновременно присутствуют молекулы, находящиеся в разных конформациях.

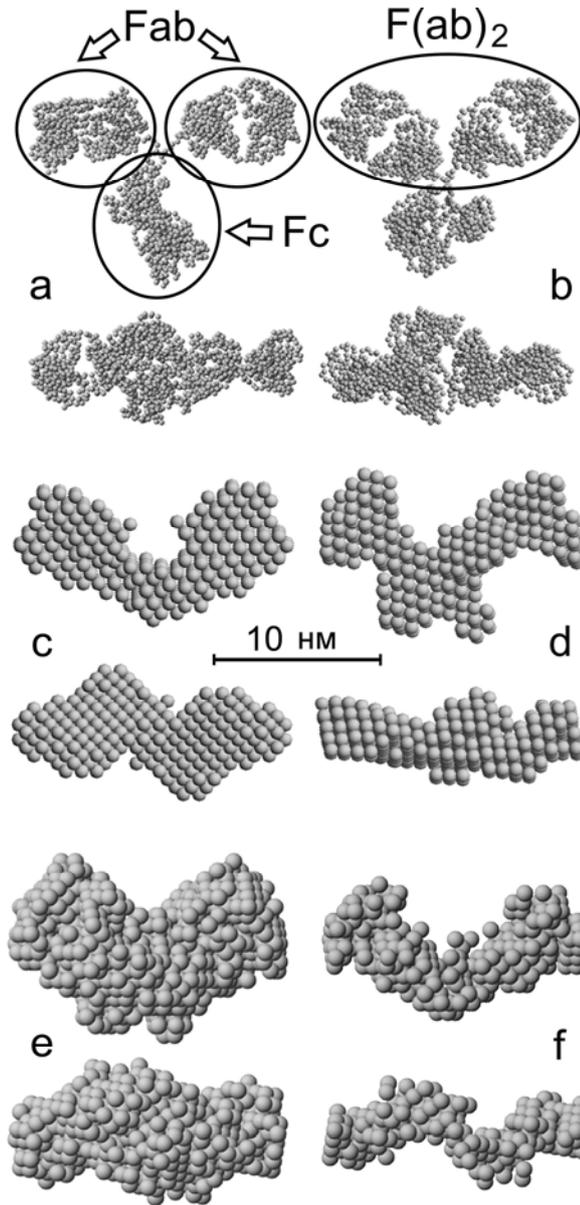


Рисунок 2.52 - Известные кристаллографические модели молекул: (a)  $\text{IgG}_{2ak}$  (код структуры 1IGT [193]), (b)  $\text{IgG}_{1k}$  (код структуры 1IGY [192]), (c) структура, восстановленная по данным малоуглового рассеяния с учетом оси симметрии 2-го порядка и (d) модель, восстановленная без дополнительной симметрии. (e) и (f) - область перекрытия решений и наиболее вероятный объем. Каждая структура показана в двух перпендикулярных ориентациях.

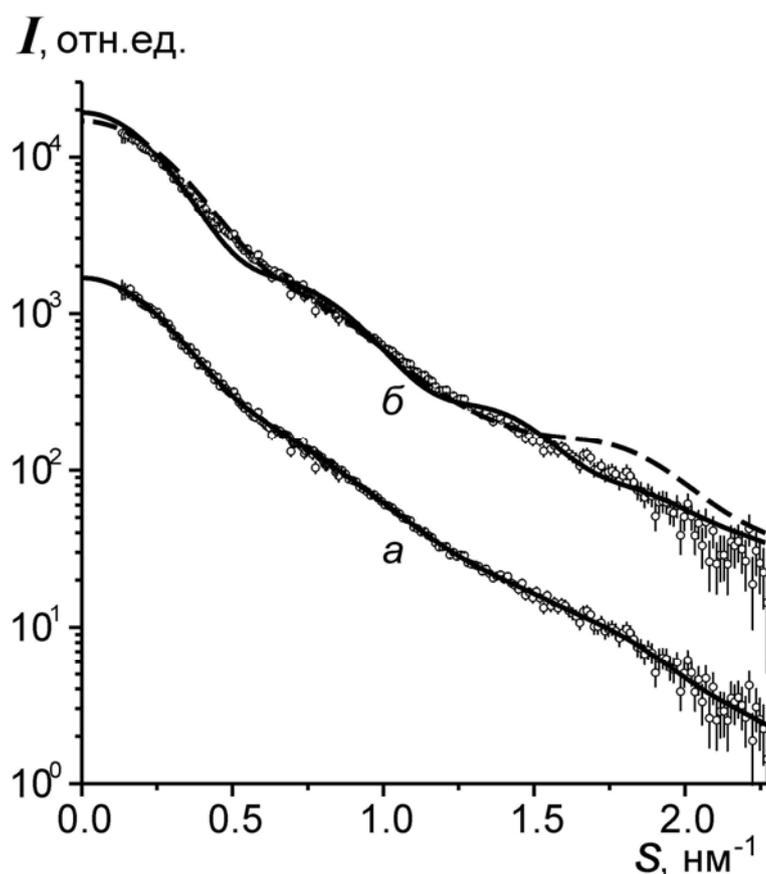


Рисунок 2.53 – (а): сравнение экспериментальных данных рассеяния раствором IgG (кружки с разбросом ошибок) с типичной кривой рассеяния от модельной структуры IgG (сплошная линия) рассчитанной по программе DAMMIN (см. Рисунок 2 b, c). (б): сравнение той же экспериментальной кривой с теоретическим расчетами по известным моделям (см. Рисунок 2.51).

#### Структуры IgM и IgM-RF в растворе.

На Рисунке 2.54 а представлены экспериментальные данные рассеяния от растворов IgM и IgM-RF в растворе, а также теоретическая кривая рассеяния, рассчитанные от симметричной пентамерной атомной модели IgM 2RCJ с помощью программы CRY SOL [194] из пакета ATSAS [A39]. В области малых углов ( $s$  до  $1 \text{ nm}^{-1}$ ) видно относительно неплохое согласие рассчитанных и экспериментальных данных. Значительное расхождение при больших углах определяется рассеянием на внутренней структуре частиц симметричной модели 2RCJ. В растворе симметрия молекулы вследствие ее гибкости может нарушаться и экспериментальная кривая рассеяния оказывается более монотонной из-за усреднения.

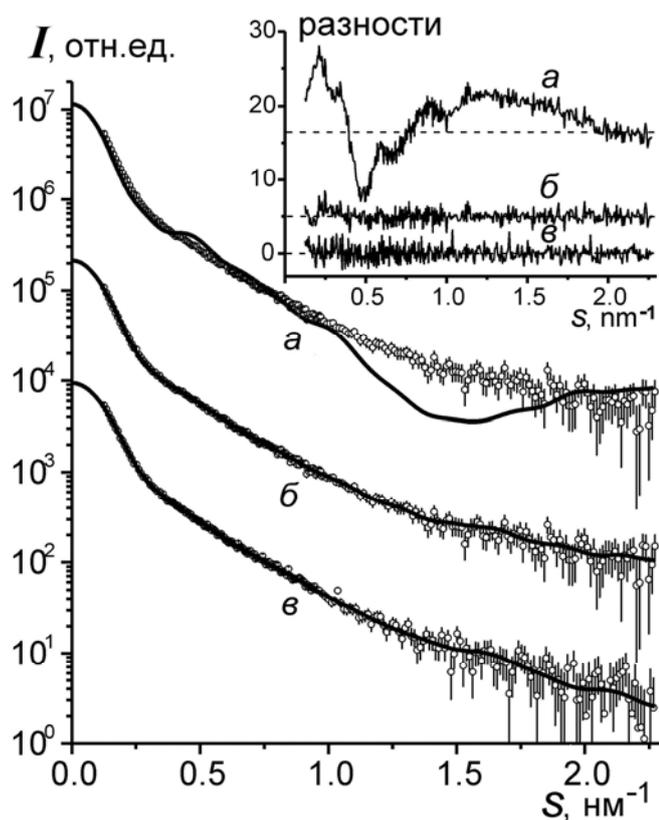


Рисунок 2.54 - Сравнение экспериментальной кривой малоуглового рассеяния раствором IgM (*a*, кружки с разбросом ошибок) с интенсивностью, рассчитанной по программе CRY SOL от атомной модели 2RCJ предложенной в [178] (*a*, сплошная линия). Сравнение нормированных экспериментальных кривых рассеяния от растворов IgM (*б*, кружки) и IgM-RF (*в*, кружки) с теоретическими кривыми, рассчитанными от найденных моделей (сплошные линии). Для наглядности сравниваемые пары кривых смещены по вертикали. На врезке: амплитуды разностных кривых для соответствующих пар, нулевые уровни показаны пунктиром.

Некоторые из типичных восстановлений симметричной структуры IgM, полученные с помощью программы DAMMIN по данным малоуглового рассеяния показаны на Рисунке 2.55. В целом, найденная при минимальных допущениях о структуре, модель IgM неплохо согласуется с предложенной С. Перкинсом моделью 2RCJ.

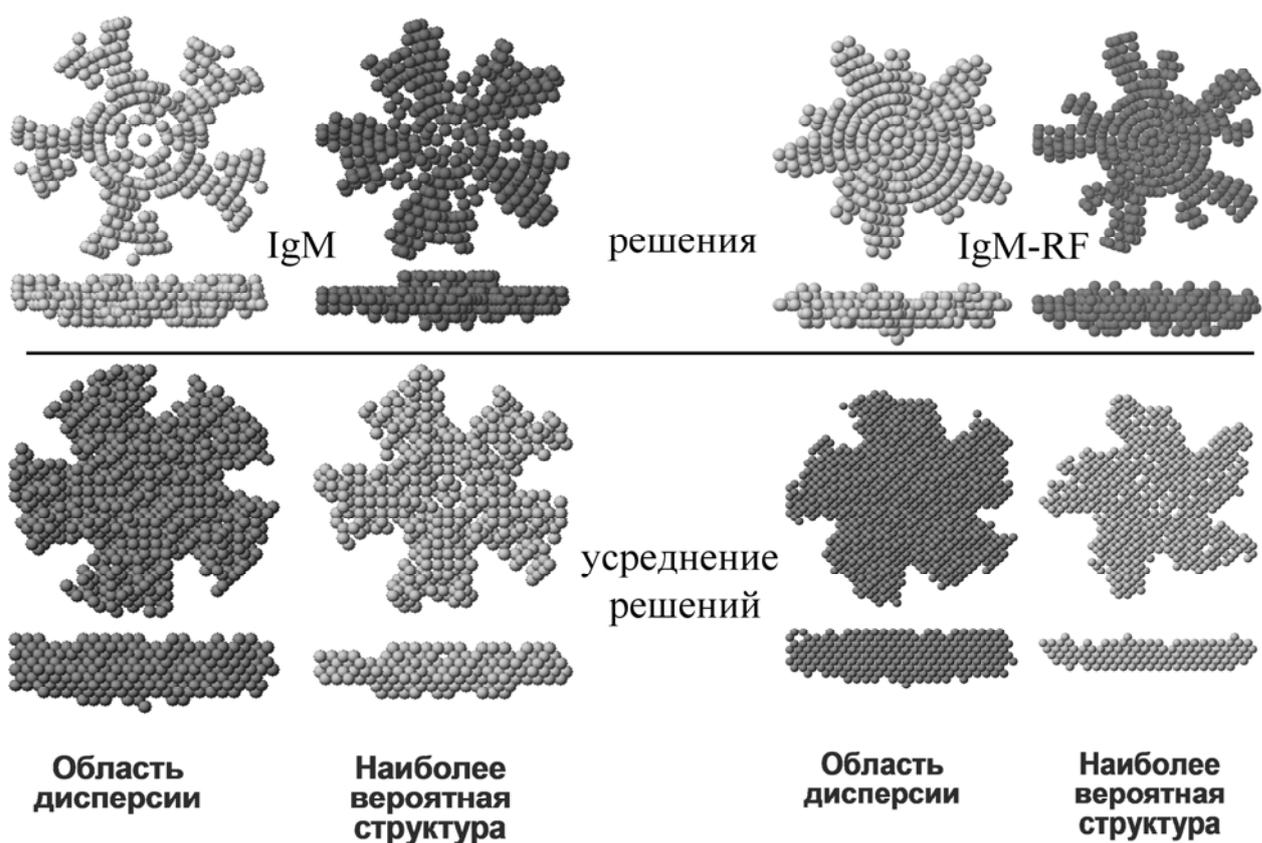


Рисунок 2.55 - Сравнение моделей IgM (левая половина рисунка) и IgM-RF (правая половина). Две верхние строчки – по два типичных решения в двух ориентациях, две нижние строчки - результаты усреднения с помощью программы DAMAVER по 30 независимо полученным решениям.

На Рисунке 2.55 (справа) представлен один из результатов восстановления структуры IgM-RF. Во всех восстановленных структурах  $F(ab)_2$  области молекулы оказались асимметричны: один из Fab-фрагментов каждого IgG-подобного домена (см. схему на Рисунке 2.47) имеет или неполный объем, или вследствие большей гибкости Fab-RF, средняя по времени структура молекулы оказывается менее заполненной на больших расстояниях от центра. Это согласуется с меньшими значениями радиуса инерции IgM-RF по сравнению с IgM и со сделанным предположением о неодинаковости в структурном и функциональном плане Fab-областей в молекуле ревматоидного фактора. Разница в заселенности Fab-областей моноклональных IgM и ревматоидного IgM-RF была также ранее подтверждена с помощью метода дифференциальной сканирующей микрокалориметрии [195]. Сравнение в упомянутой работе параметров тепловой

денатурации Fab и  $(Fc)_5$  фрагментов IgM и IgM-RF позволило ее авторам сделать вывод о том, что  $(Fc)_5$ -фрагменты обоих белков термодинамически мало отличаются, в то время как их Fab-фрагменты различаются характером взаимодействия между доменами.

Поиск более детализированных структур с помощью модифицированной программы DAMMIN, а также программ моделирования структуры последовательностью аминокислотных остатков GASBOR и интерактивного моделирования доменами MASSHA [196] подтвердил ранее полученные результаты, за исключением того, что максимальный размер молекул ревматоидного фактора оказался ближе к размерам IgM, несмотря на меньший радиус инерции. Был обработан ряд как ранее измеренных кривых рассеяния, так и интенсивности, полученные от новых образцов. Экспериментальные данные показаны на Рисунке 2.57, типичные результаты восстановления формы частиц - на Рисунках 2.58 и 2.59.

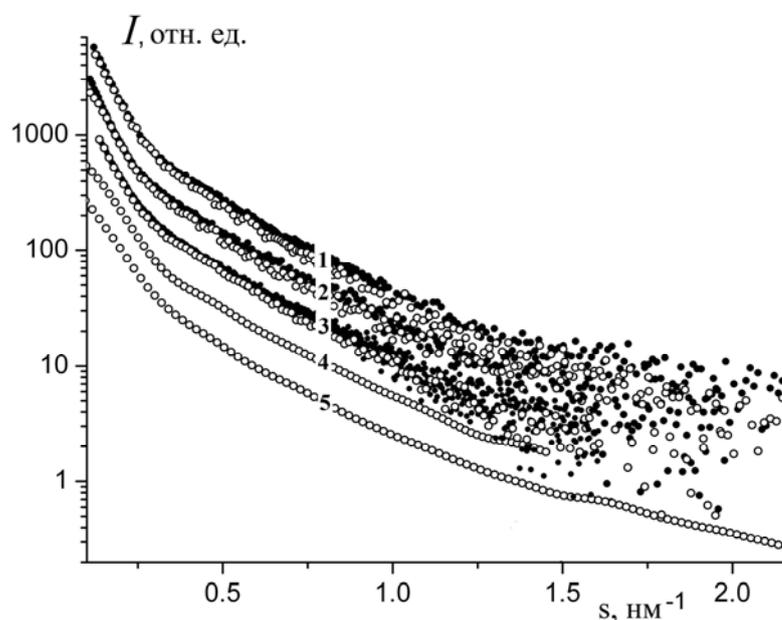


Рисунок 2.57 - Сравнение экспериментальных данных рассеяния растворами IgM (1 - 3, черные точки) и IgM-RF (кружки). Для большей наглядности пары кривых, измеренных в одинаковых условиях, смещены по вертикали.

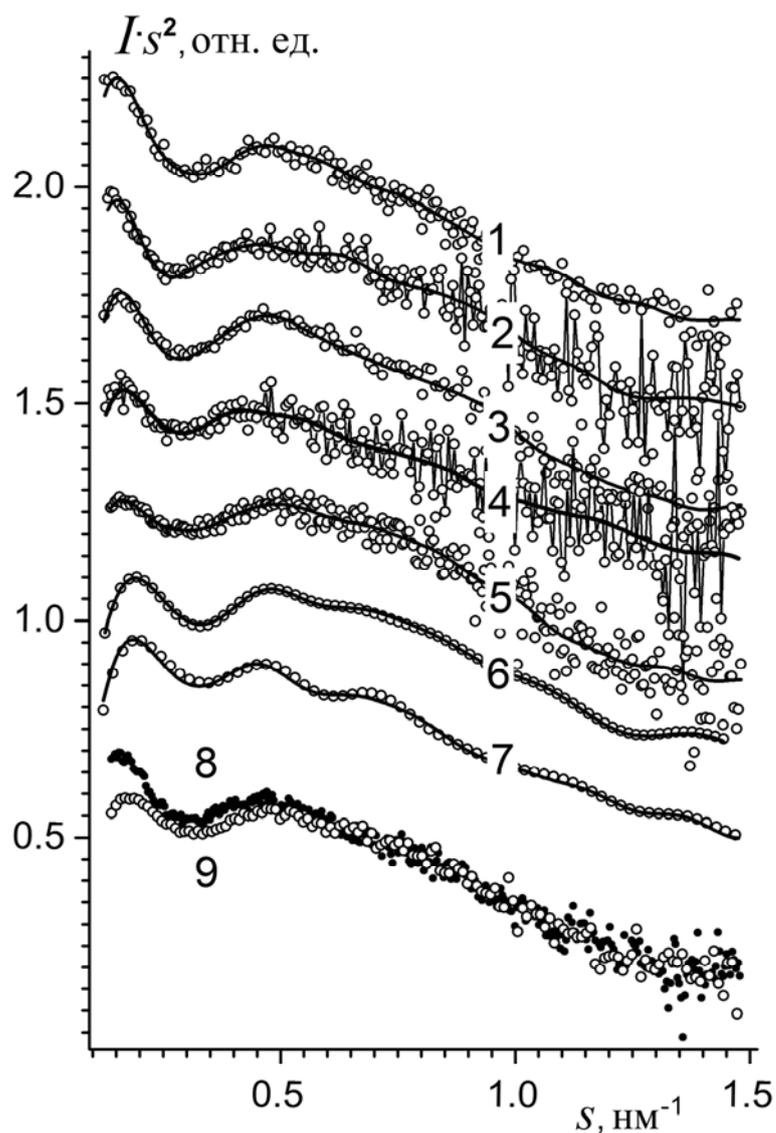


Рисунок 2.58 - Экспериментальные интенсивности (точки) и рассеяние от найденных структурных моделей (гладкие сплошные линии) в координатах Кратки для растворов (1-3): IgM1, IgM2, и IgM3; (4-7): IgM-RF1, IgM-RF2, IgM-RF3, и IgM-RFq, соответственно. Кривая (8): усредненная интенсивность рассеяния для всех образцов IgM. (9): усредненная интенсивность для всех образцов ревматоидных факторов (кружки). Пары сравниваемых кривых сдвинуты по вертикали для наглядности. Интенсивности (8, 9) в области  $0,5 < s < 1,0 \text{ нм}^{-1}$  практически идентичны, тогда как рассеяние от молекул ревматоидного фактора в начальной области  $s < 0,5 \text{ нм}^{-1}$  систематически меньше рассеяния от интактного иммуноглобулина М.

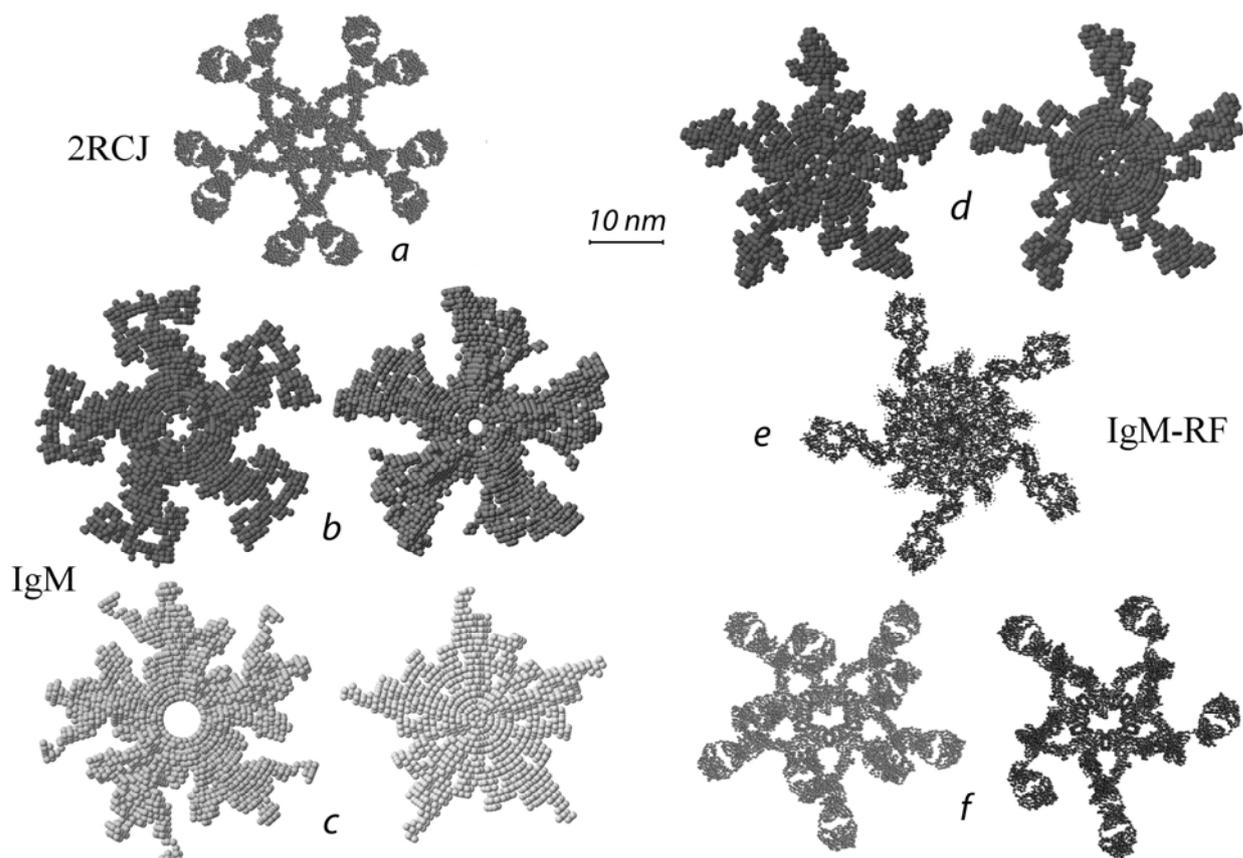


Рисунок 2.59 - Сравнение типичных моделей, найденных по данным рассеяния, с моделью Перкинса 2RCJ (а). (b): два решения IgM (DAMMIN). (c, d): 4 решения для IgM-RF (DAMMIN). (e): модель из аминокислотных остатков (программа GASBOR). (f): моделирование доменами, из которых состоит структура 2RCJ.

На Рисунке 2.60 показано сравнение усредненных по всем расчетам контуров функций распределения парных расстояний. Расчет для структуры 2RCJ проводили путем прямого подсчета количеств расстояний между атомами с помощью разработанной программы DAMPRIS. Видно, что распределение для образцов интактного иммуноглобулина IgM в среднем близко по своему характеру распределению, найденному для 2RCJ, при том, что экспериментальные кривые показывают больший максимальный размер молекул, чем диаметр структуры 2RCJ. Это свидетельствует о подвижности Fab-фрагментов в растворе и увеличении, как следствие, среднестатистического диаметра всей структуры. Однако, данный вывод нельзя считать окончательным и надежным, так как аналогичный эффект увеличения кажущегося диаметра может дать возможная примесь агрегатов макромолекул, которая в данном случае

слишком незначительна, чтобы их влияние на рассеяние можно было достоверно детектировать.

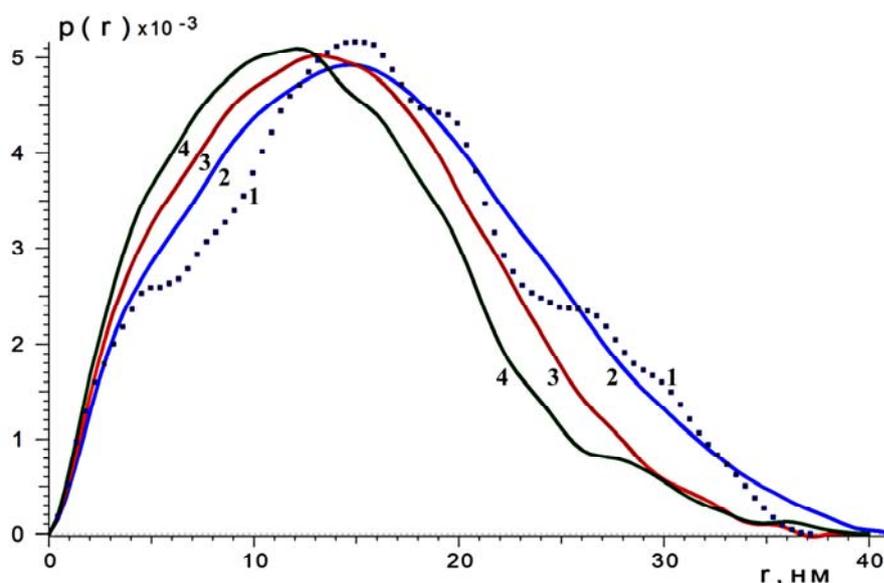


Рисунок 2.60 - Усредненные функции парных расстояний для (1) - структуры 2RCJ, точки; (2) - данных для IgM; (3) - образцов ревматоидного фактора IgM-RF1, IgM-RF2 и IgM-RF3 (болезнь Вальденстрема); (4) - образца IgM-RFq (потенциальный больной).

Таким образом, основное различие между молекулами IgM и IgM-RF, определяемое по данным рассеяния, заключено в меньшем количестве отрезков большой длины, которые можно вписать в структуру ревматоидного фактора. Наиболее наглядно этот факт демонстрирует различие между функциями радиального распределения. Для их оценки была написана программа расчета количества атомов (шариков в структурных моделях)  $N_r$ , находящихся на расстоянии  $r$  от центра тяжести частицы и усреднения распределений по разным измерениям и сериям расчетов (10-30 решений для каждого измерения). Были объединены результаты для образцов, перечисленных в подписи под Рисунком 2.58. Результаты расчетов показаны на Рисунке 2.61. Распределение для молекулы IgM по данным рассеяния (Рисунок 2.61,2) близко к распределению, усредненному по серии модельных восстановлений (Рисунки 2.49 — 2.50) и значительно отличается от распределения, рассчитанного для структуры 2RCJ

(Рисунок 2.61(1)), в области больших расстояний от центра, что согласуется с предположением о гибкости частицы в растворе. Отличие распределений для IgM-RF еще значительнее, что говорит о большей подвижности периферийных доменов ревматоидного фактора.

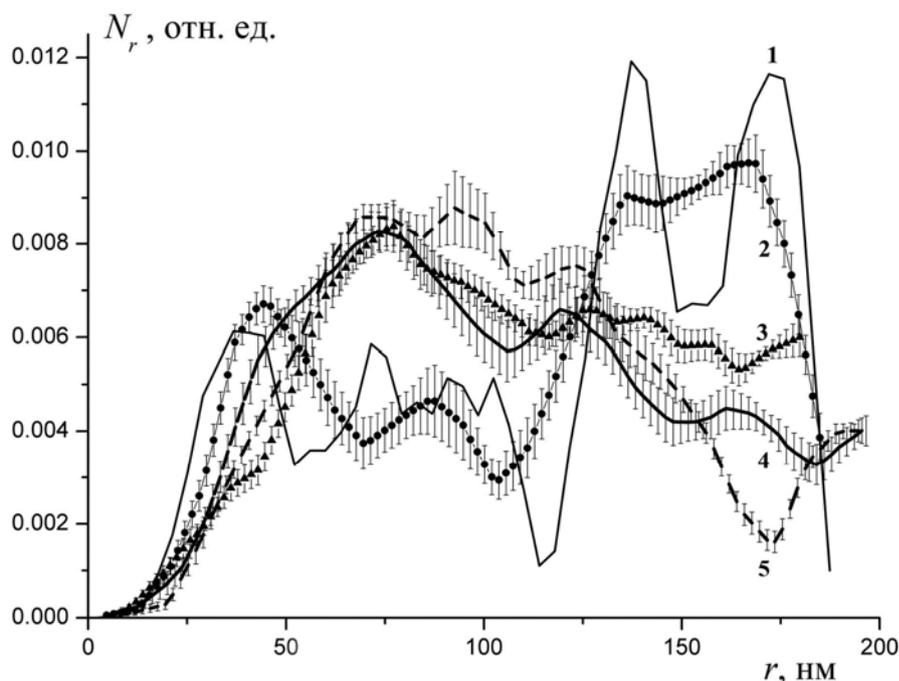


Рисунок 2.61 - Усредненные функции радиального распределения числа структурных элементов (атомов углерода в 2RCJ или шариков в моделях, найденных по данным малоуглового рассеяния) в зависимости от расстояния от центра тяжести частиц. (1): 2RCJ; (2): усредненные модельные решения на Рисунке 2.49 - 2.50; (3): данные от растворов IgM1-IgM3; (4): образцы IgM-RF1 - IgM-RF3; (5): IgM-RFq. Разбросы распределений рассчитаны по 10-30 решениям для каждого образца и отражают численную стабильность расчетов структурных моделей.

Моделирование молекул иммуноглобулина и ревматоидных факторов с помощью фрагментов.

Сравнение экспериментальных кривых рассеяния с интенсивностями, рассчитанными от известных кристаллических структур высокого разрешения, позволяет оценивать различие в строении макромолекул в кристалле и растворе, а также проверять адекватность гипотетических моделей молекул. Расчет интенсивности малоуглового рассеяния от атомных структур проводили с помощью программы CRY SOL. Конструирование структур из фрагментов осуществляли с помощью программы моделирования MASSHA [196] (Рисунок 2.59(f)). Эффект асимметрии периферийных областей хорошо виден и в решении для ревматоидного IgM<sub>k</sub>, полученного с помощью программы GASBOR (Рисунок 2.59(e)).

Таким образом, сравнение форм молекул IgM и IgM-RF, восстановленных по данным малоуглового рассеяния различными методами, привело к выводу о большей относительной подвижности фрагментов молекул ревматоидного фактора, анализ радиальных распределений плотности показал большую подвижность периферийных областей Fab-фрагментов. Эта подвижность может быть обусловлена, в свою очередь, увеличенной гибкостью этих фрагментов, исследование которых представлено ниже.

Спектры ЭПР спин-меченых Fab и Fab-RF фрагментов.

После введения спин-метки в молекулы Fab- и Fab-RF-фрагментов были записаны спектры ЭПР (Рисунок 2.62a) [A44].

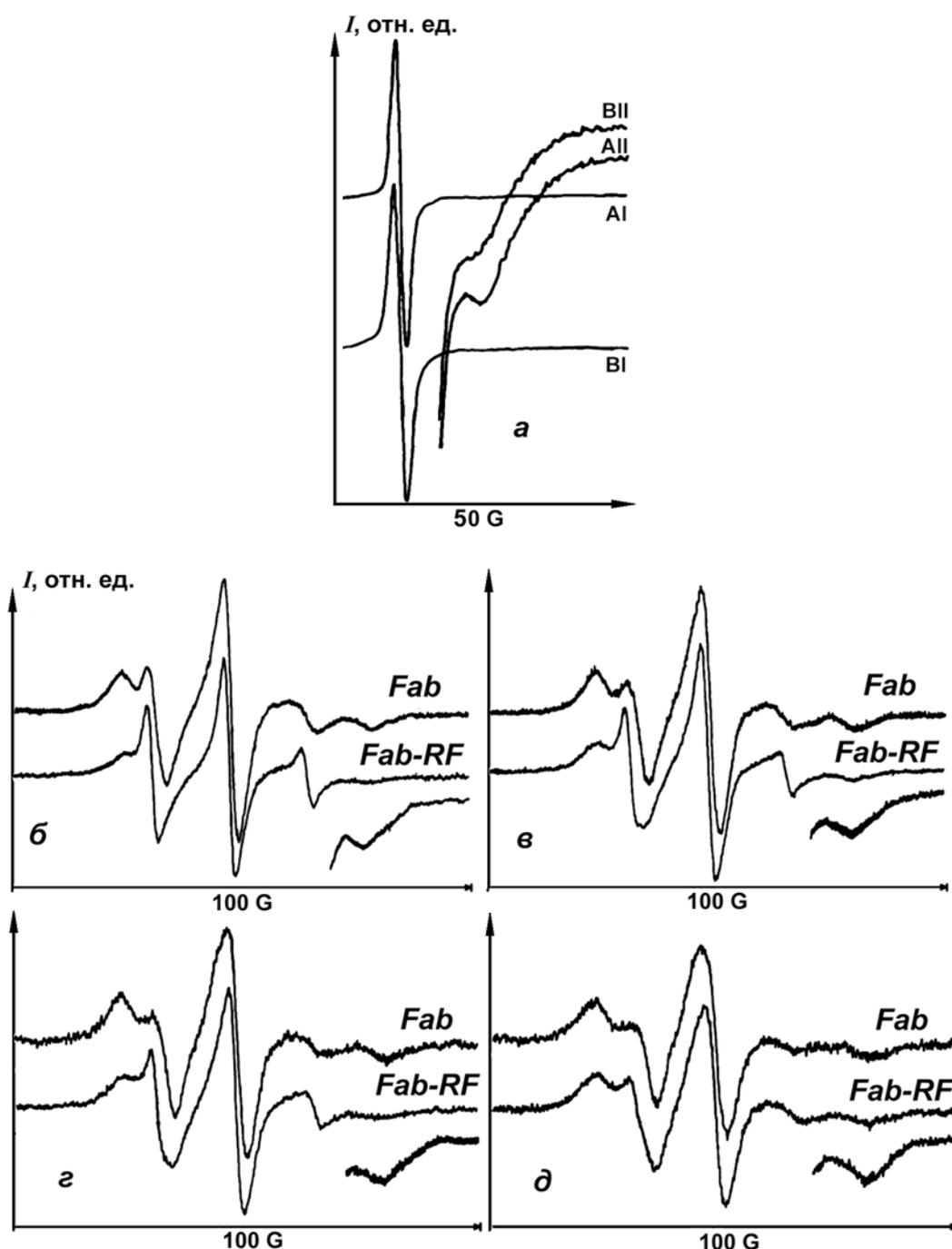


Рисунок 2.62 - Сравнение высокопольных пиков ЭПР спектров при слабом (I) и сильном (II) усилении: А – спин-меченые Fab-фрагменты; В – спин-меченые Fab-RF-фрагменты. Измерения проведены до диализа при температуре 20° (а). ЭПР спектры пиковых фракций (после диализа) спин-меченых Fab- и Fab-RF-фрагментов при различных температурах и вязкостях среды: 20°С (б, в), 1°С (z, д), 0% сахарозы (б, z), 33% сахарозы (в, д).

Как видно из рисунка, высокопольные пики в ЭПР спектрах заметны при большом усилении уже при 20°С даже до диализа, что свидетельствует о

ковалентном связывании спин-метки с белком. Уже на этой стадии заметны различия в спектрах. Спектры ЭПР, записанные после диализа спин-меченых образцов (Рисунок 2.62б-2д), еще более наглядно показывали наличие крайних широких пиков. При этом спектр ЭПР от свободной спин-метки дает заметный вклад в суммарный спектр. Поэтому диализованные спин-меченые препараты были подвергнуты дополнительной гель-фильтрации на колонке с сефадексом G-25-средний с полным ЭПР мониторингом собираемых фракций. Было получено хорошее разделение спин-меченых препаратов и остатков непрореагировавшей спин-метки. При этом следует отметить, что спин-меченый Fab-RF-фрагмент заметно дольше удерживается на колонке по сравнению со спин-меченым Fab-фрагментом.

Для пиковых фракций обоих спин-меченых препаратов было проведено исследование температурно-вязкостной зависимости спектров ЭПР. На основании измерения расстояний между крайними широкими пиками в спектрах ЭПР, записанных при 1 и 20<sup>0</sup>С и переменной вязкости среды, были измерены зависимости указанного расстояния от вязкости среды для обоих спин-меченых препаратов. По формулам (2.66) и (2.67) рассчитанные величины времени вращательной корреляции  $\tau$  и параметра упорядоченности  $S$  составили при 20<sup>0</sup>С: для Fab – 21±2 и 0.60±0.01, а для Fab-RF – 11±1 и 0.59±0.01 соответственно.

Моделирование Fab-фрагментов по данным малоуглового рентгеновского эксперимента.

Кривые малоуглового рентгеновского рассеяния Fab-фрагментами нормального и ревматоидного иммуноглобулинов представлены на Рисунке 2.63а [А44].

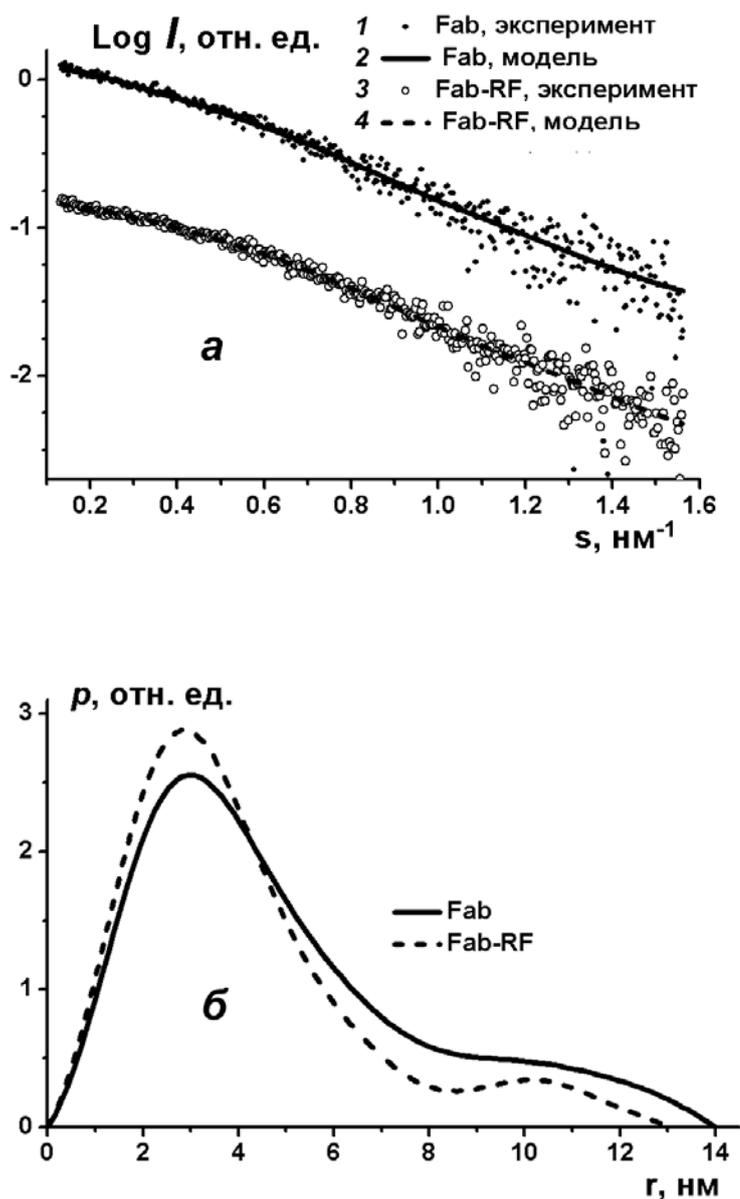


Рисунок 2.63 - Кривые малоуглового рентгеновского рассеяния от растворов Fab-фрагментов нормального и ревматоидного иммуноглобулинов и модельные кривые, рассчитанные от найденных структурных моделей, показанных на Рисунке 2.64 (а, в): 1 – Fab, 2 – модельная кривая для Fab, 3 – Fab-RF, 4 – модельная кривая для Fab-RF. Кривые разнесены попарно по вертикали для лучшей визуализации (а). Распределения по расстояниям внутри молекул, рассчитанные по данным рассеяния (б).

В области углов до  $0.5 \text{ nm}^{-1}$  имеются небольшие отличия в форме кривых, что является отражением отличия структур этих макромолекул, что подтверждается значениями радиуса инерции  $R_g$ , максимального размера  $D_{max}$  и

объема  $V$  для Fab- и Fab-RF-фрагментов, рассчитанных по кривым рассеяния с использованием приближения Гинье (уравнение 2.4) и программы GNOM (уравнение 2.6):

для Fab-фрагмента:  $R_g = 2.9 \pm 0.1$  нм,  $D_{max} = 10.8 \pm 0.3$  нм,  $V = 88 \pm 2$  нм<sup>3</sup>;

для Fab-RF-фрагмента:  $R_g = 2.7 \pm 0.1$  нм,  $D_{max} = 10.0 \pm 0.3$  нм,  $V = 82 \pm 2$  нм<sup>3</sup>.

Эти различия хотя и не являются значительными, тем не менее, они вполне достоверны, так как ошибка определения геометрических характеристик составляла не более 3%.

Результаты восстановления формы Fab-фрагментов молекул IgM и IgM-RF в растворе показаны на Рисунке 2.64, а соответствующие им модельные кривые рассеяния, рассчитанные для этих фрагментов с помощью программы DAMMIN – на Рисунке 2.63а (сплошные линии 2 и 4). Полученные модели близки по форме. Однако практическое отсутствие внутренней полости для модели Fab-RF и более компактное строение этой макромолекулы говорит о некоторых структурных перестройках в ней по сравнению с нормальным Fab-фрагментом. Наличие в моделях "хвостов" является артефактом и, по-видимому, связано с присутствием некоторого количества агрегатов в растворе, рассеяние от которых увеличивает оценку максимального размера макромолекул в монодисперсном приближении. Действительно, максимальный размер молекул Fab и Fab-RF значительно больше (Рисунок 2.63б) по сравнению с данными кристаллографических исследований, в соответствии с которыми  $R_g = 2.6$  нм, а  $D_{max} = 8.9$  нм. Головная же часть восстановленных моделей молекул близка по форме к известной кристаллической структуре.

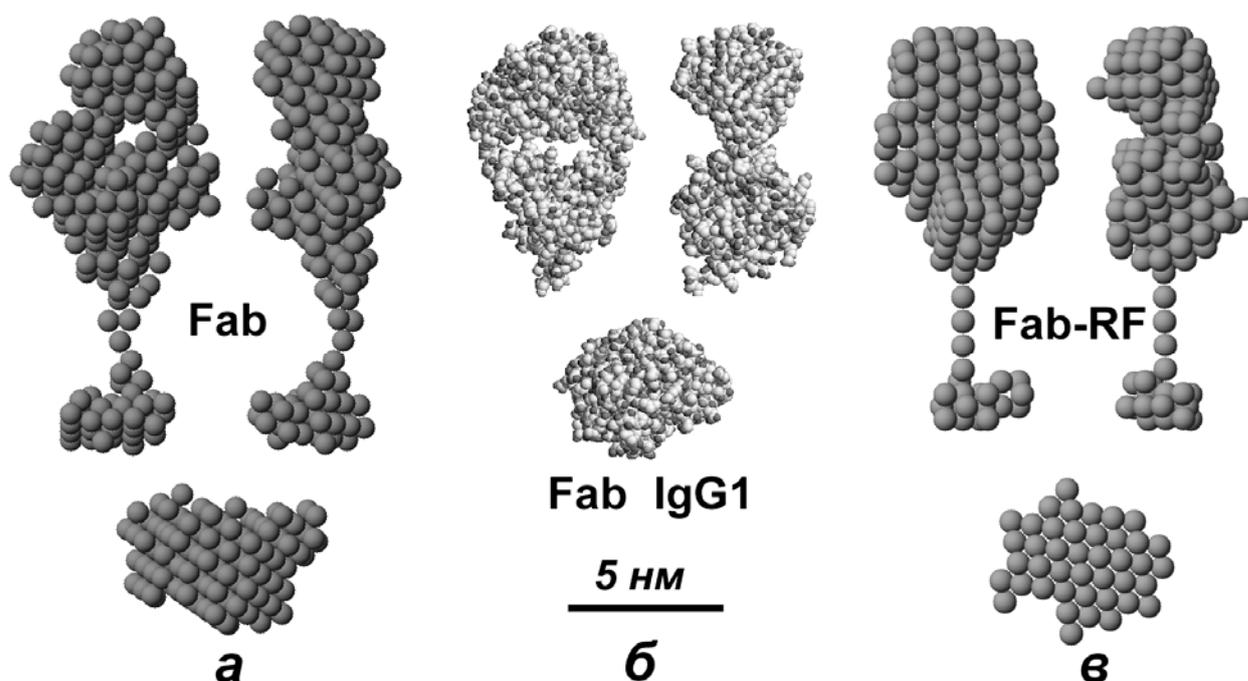


Рисунок 2.64 - Восстановленные формы макромолекул Fab-фрагментов нормального (*a*) и ревматоидного (*в*) иммуноглобулинов программой DAMMIN. Структуры показаны в трех взаимно перпендикулярных ориентациях. Для сравнения показан Fab-фрагмент, выделенный из структуры 1IGY.PDB [193] (*б*).

Для проверки надежности сделанных выводов была проведена серия численных экспериментов по восстановлению формы Fab-фрагмента с использованием модельной интенсивности рассеяния, рассчитанной для известной кристаллической структуры (*б*) Рисунка 2.64. На Рисунке 2.65 представлены результаты моделирования, полученные в тех же условиях расчетов, что и для экспериментальных данных при доле димера в модельной смеси равной 10% вес. Интенсивность рассеяния от атомных структур были рассчитаны с помощью программы CRY SOL.

Из рисунка видно, что форма распределения, рассчитанная по данным рассеяния от смеси похожа на распределения, показанные на Рисунке 2.63(*б*). Найденные структуры по своему характеру аналогичны структурам, представленным на Рисунке 2.64.

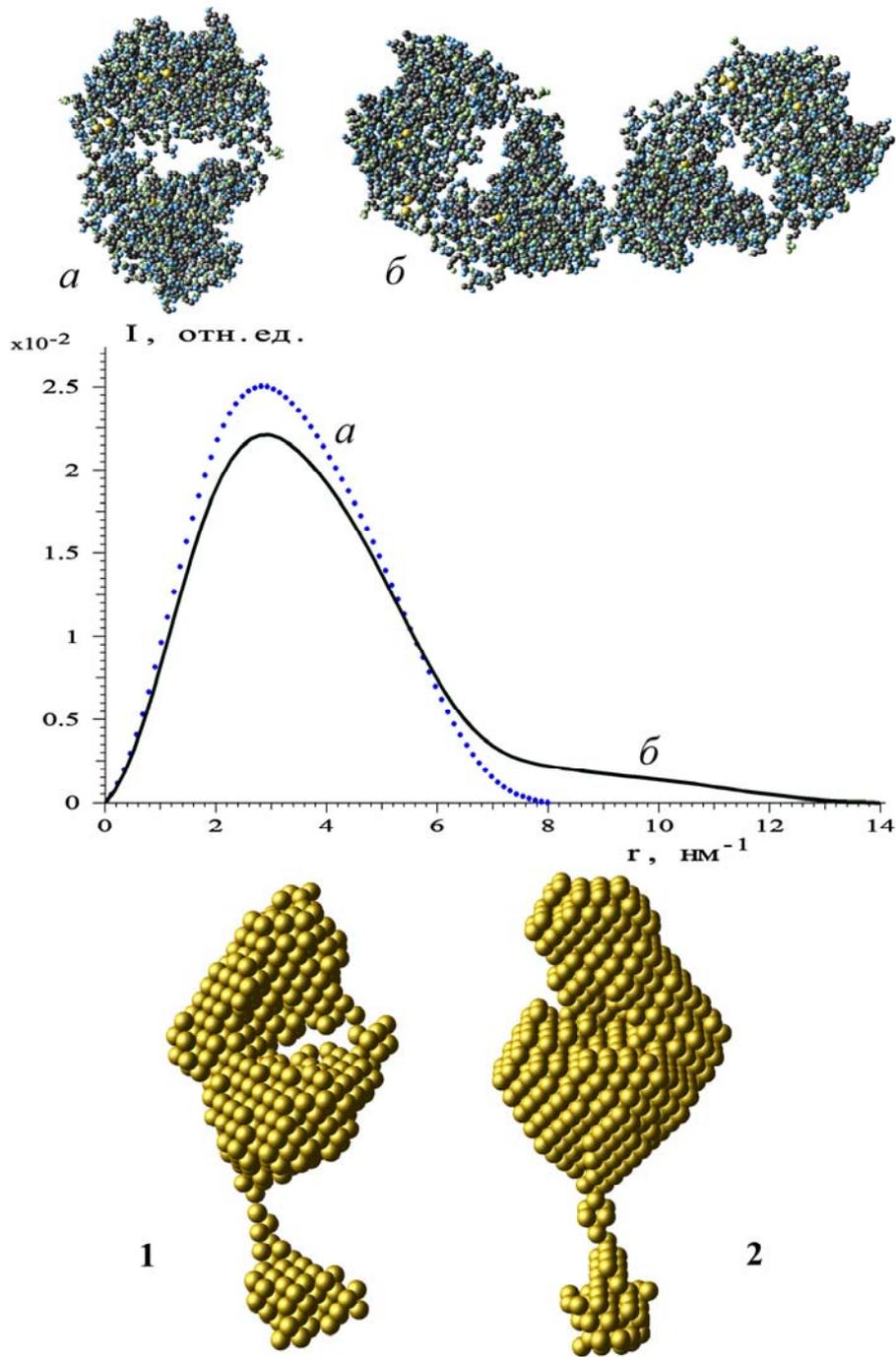


Рисунок 2.65 - Результаты определения формы доменов Fab (а) при наличии в модельной смеси 10% димеров (б). На графике показано сравнение распределений по расстояниям, рассчитанные для раствора мономера (а) и для смеси (б). Внизу: два типичных решения, полученных с помощью модифицированной программы DAMMIN по данным малоуглового рассеяния от модельной смеси.

Определение молекулярных характеристик Fab- и Fab-RF-фрагментов с помощью масс-спектрологии.

Как следует из Рисунка 2.66 (а,б), полученные препараты IgM и IgM-RF в нативном состоянии практически гомогенны и их молекулярная масса (порядка 50 кДа) очень близка к величинам, указанным в литературе. Правда, при этом молекулярная масса Fab-RF меньше массы Fab примерно на 380 Да, что хотя и достоверно, но не принципиально: спектр смеси обоих препаратов не выявил расщепления основного массового пика. Однако после трипсинолиза разница масс-спектров Fab и Fab-RF уже представляется весьма значительной.

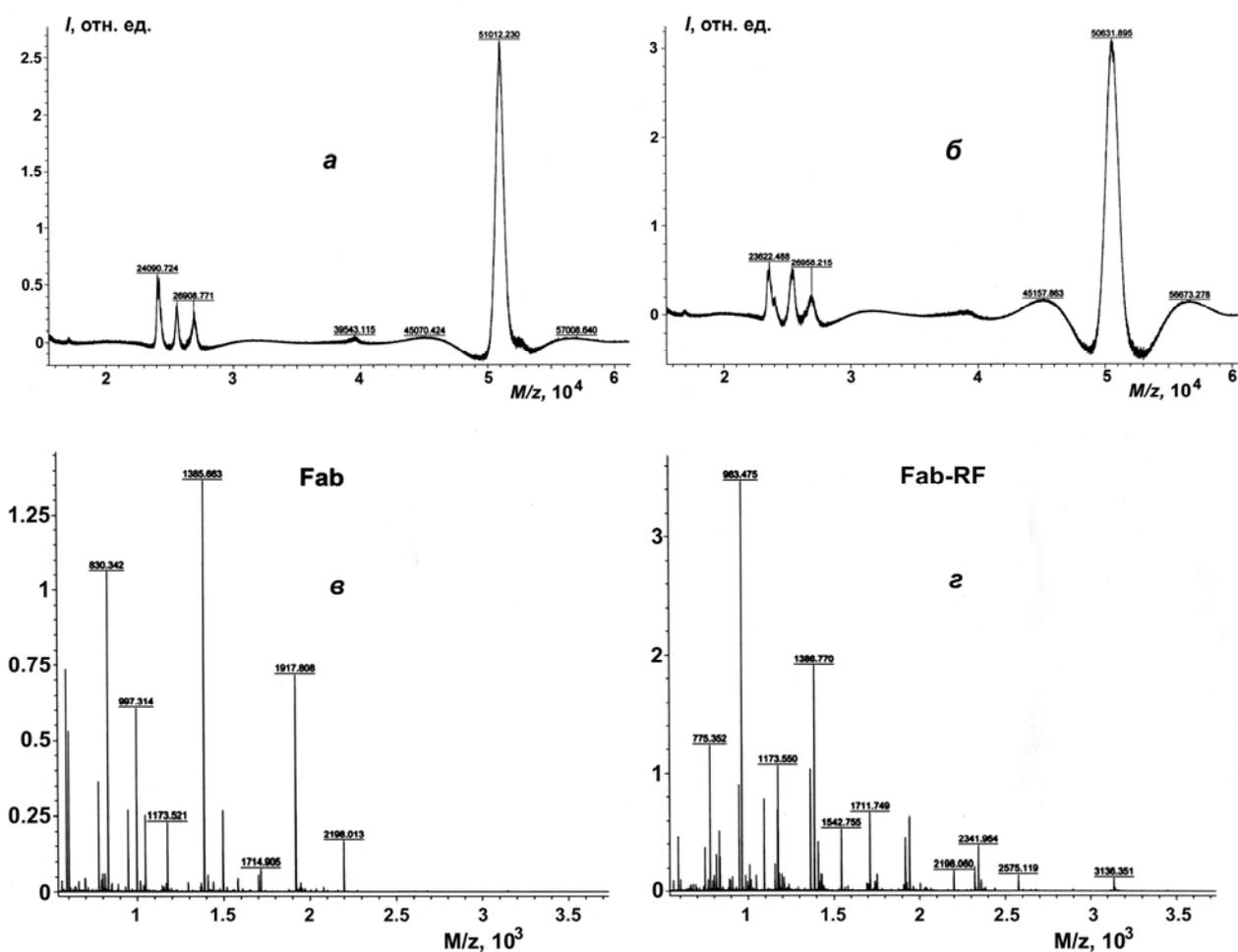


Рисунок 2.66 - Масс-спектры Fab (а), Fab-RF (б) и фрагментов, полученных методом расщепления трипсином (в, г, соответственно).

Известно, что трипсин гидролизует связи Lys-X и Arg-X, за исключением случаев, когда X = Pro (полная негидролизуемость) и X = Asp, Gly, Cys (частичная

негидролизуемость). С этой точки зрения проведенный нами анализ первичных структур обоих фрагментов показал, что теоретически после трипсинолиза из Fab должно образоваться 33 пептида, а из Fab-RF – 31. Из Рисунка 2.66 следует, что в реальности из Fab получается только 15 пептидных фрагментов, а из Fab-RF – 28. Заметим, что трипсинолиз протекает теоретически в условиях полной денатурации белка, т.е. полного разворачивания глобулярной структуры. При этом из Fab-RF образовалось количество пептидов, весьма близкое к ожидаемому, а из Fab - только около половины. Из этого следует, что в использованных условиях трипсинолиза пространственная структура Fab-RF оказалась гораздо лабильнее, чем в случае Fab, что согласуется с данными метода спин-метки. Метод ЭПР выявил лишь гибкость междоменных областей Fab-RF, тогда как метод МС показал еще и лабильность структуры всех глобул, образующих этот фрагмент. Таким образом, МС спектры также свидетельствуют о возможных структурных перестройках в макромолекуле Fab-RF-фрагмента, которые определяются внутренними связями и изменяют подвижность отдельных ее участков.

Основным выводом из результатов экспериментов является свидетельство повышенной гибкости и подвижности Fab-участков молекул IgM-RF по сравнению с IgM при незначительных различиях в форме и молекулярной массе фрагментов. Величина параметра упорядоченности  $S$ , характеризующая ограничение подвижности спин-метки, только незначительно различается для обоих препаратов. Это свидетельствует о том, что объем угловых переориентаций спин-метки в обоих белках одинаков. Однако спектры ЭПР спин-меченых Fab- и Fab-RF-фрагментов (Рисунок 2.62) имеют серьезные различия: при большом усилении видно, что пик VII сдвинут к острому высокопольному пику от свободной метки больше, чем пик AII. Этот сдвиг говорит о том, что в спин-меченном Fab-RF-фрагменте метка более расторможена, чем в спин-меченном Fab-фрагменте. В то же время эффективные величины  $\tau$  для Fab- и Fab-RF-фрагмента отличаются почти вдвое и равны  $21 \pm 2$  нс и  $11 \pm 1$  нс, соответственно. Величина 21 нс соответствует жесткой структуре и молекулярной массе ( $\approx 50$  кДа)

Fab-фрагмента. В то же время, найденная эффективная величина времени вращательной корреляции  $\tau$ , равная для Fab-RF-фрагмента 11 нс, должна соответствовать величине молекулярной массы около 25 кДа. Такая молекулярная масса могла бы наблюдаться при условии жесткости структуры этого фрагмента. Кроме того, столь резкое уменьшение молекулярной массы в данном случае не имеет места, так как данные малоуглового рентгеновского рассеяния и масс-спектропии свидетельствуют только о незначительном изменении формы, размера и массы этих биополимеров. Результаты исследования этими методами указывают на изменения внутренней структуры и большей подвижности отдельных частей ревматодного Fab-фрагмента. Следовательно, можно предположить, что молекула Fab-RF обладает значительно большей междоменной подвижностью и внутриглобулярной лабильностью по сравнению с нормальным Fab-фрагментом. Это и обуславливает столь резкое уменьшение величины  $\tau$ . Этот вывод подтверждается и другими данными. Результаты рентгеноструктурного анализа [187] показывают, что макромолекула Fab-RF на участках  $V_{\mu} - C_{\mu}1$  и  $V_L - C_L$  имеет значительные изгибы. Кроме того, И.И. Протасевич и др., исследуя Fab- и Fab-RF-фрагменты методом дифференциальной сканирующей микрокалориметрии, отмечали, что эти фрагменты различаются характером междоменных контактов именно на указанных выше участках [195]. В пользу предположения о большей гибкости структуры Fab-RF-фрагмента говорит и тот факт, что при отделении непрореагировавшей спин-метки на колонке с сефадексом спин-меченый Fab-RF-фрагмент удерживался на колонке заметно дольше, чем спин-меченый Fab-фрагмент. Такое влияние гибкости макромолекулы на ее время прохождения через колонку при гель-фильтрации теоретически описано в [197].

Сравнение последовательностей аминокислотных остатков междоменных областей  $V_{\mu} - C_{\mu}1$  и  $V_L - C_L$  обоих фрагментов также указывает на справедливость выдвигаемого предположения о большей гибкости Fab-RF. Известно, что гибкость структуры в глобулярных белках связана с нерегулярными участками (петлями) и обусловлена наличием в них аминокислотных остатков Pro и Gly

[198]. Действительно, гибкие шарнирные участки в молекулах IgG1 и IgG4 весьма богаты аминокислотными остатками Pro [199]. В Fab-фрагменте неревматоидного IgM в нерегулярном междоменном участке  $V_{\mu} - C_{\mu}1$  (остатки 122-130) Pro отсутствует и имеется только один Gly, а в междоменном участке  $V_L - C_L$  (остатки 105-112) нет ни одного из указанных аминокислотных остатков [200]. Такая аминокислотная последовательность определяет упомянутую выше жесткость неревматоидного Fab-фрагмента в целом. Напротив, соответствующие участки ревматоидного Fab-фрагмента (аминокислотные остатки 120-130  $\mu$ -цепи и 105-112 L-цепи), исследованные методом рентгеноструктурного анализа, имеют каждый по одному аминокислотному остатку Pro и Gly [201]. Эти участки находятся приблизительно посередине глобулы Fab-фрагмента, разделяя ее на две подвижные друг относительно друга половины.

*Выводы.* Таким образом, анализ литературных и полученных экспериментальных данных подтверждает предположение о большей гибкости глобулярной структуры Fab-RF-фрагмента в отличие от жесткой структуры неревматоидного Fab при практически одинаковой молекулярной массе и конформации макромолекул обоих типов в растворе. Обнаруженное структурно-динамическое свойство Fab-RF-фрагмента может пролить свет на особенности взаимодействия Fab-RF-областей в составе целой макромолекулы IgM с антигеном, т.е. с иммуноглобулином IgG собственного организма. Полученные результаты важны для более полного понимания процессов протекания аутоимунных заболеваний и для разработки новых способов их диагностики и лечения, особенно учитывая возможность связывания еще одного антигена ревматоидным Fab-фрагментом.

Работа по исследованию морфологии молекул иммуноглобулинов и ревматоидных факторов в растворе представляет собой пример комплексного подхода, в котором метод малоуглового рассеяния играет ключевую роль.

### 2.3 Анализ данных малоуглового рассеяния от смесей полидисперсных частиц

В разделе 2.1 был рассмотрен принцип расчета распределений по размерам рассеивающих неоднородностей в образце методом моделирования интенсивности малоуглового рассеяния набором формфакторов (уравнение (2.9)).

Расчет распределений проводится при априори заданных формах частиц, так как одновременное определение распределений и формы представляет собой математически недоопределенную задачу. В некоторых случаях (если возможно получение образцов с узким распределением, не превышающим по полуширине 5-10%) определение формы частиц может быть проведено с помощью выше рассмотренных методов.

Малоугловое рассеяние от смеси, содержащей различные типы невзаимодействующих компонентов, представляет собой сумму парциальных интенсивностей рассеяния этими компонентами, взвешенных согласно их объемным долям с поправкой на межчастичную интерференцию (2.10). Если рассеивающие частицы (поры, или неоднородности другого типа) имеют определенные формы, но различаются по размерам и шириной распределения по размерам (полидисперсностью), то эти характеристики могут быть параметризованы и определены вместе с их объемными долями с помощью метода наименьших квадратов. Интенсивность рассеяния от  $k$ -ой компоненты смеси (системы частиц одинаковой формы, но разных размеров) является изотропной функцией, зависящей от вектора рассеяния  $s$  и может быть представлена в виде:

$$I_k(s) = \int_0^{\infty} N_k(R) [v_k(R) \Delta\rho_k(R)]^2 i_k(s, R) S_k(s) dR, \quad (2.68)$$

где  $N_k(R)$  - функция распределения по размерам,  $\Delta\rho_k(R)$ ,  $v_k(R)$  и  $i_k(s, R)$  обозначают контраст, объем и нормализованную интенсивность рассеяния (квадрат формфактора) частицы радиусом  $R$  (эти функции определяются формой и внутренней структурой частицы, при этом  $i_k(0, R)=1$ ). Эффекты межчастичной интерференции

учитывают путем введения функции структурного фактора  $S_k(s)$  который в данной работе рассчитывали в приближении потенциала жестких сфер Перкуса-Йовика [202, 203, A34]. Полный объем компоненты равен

$$V_k = \int_0^{\infty} v_k(R) N_k(R) dR. \quad (2.69)$$

Для того, чтобы на графике распределения вклад больших частиц был заметнее, вместо распределения по числу частиц можно использовать распределение по объемам (будем называть его **парциальным распределением**, а соответствующую  $I_k(s)$  (2.68) **парциальной интенсивностью**):

$$D_k(R) = N_k(R) v_k(R). \quad (2.70)$$

Учитывая (2.69, 2.70), уравнение (2.68) можно переписать в нормированном виде

$$I_k(s) = V_k \frac{\int_0^{\infty} D_k(R) v_k(R) [\Delta\rho_k(R)]^2 i_k(s, R) dR}{\int_0^{\infty} D_k(R) dR} \quad (2.71)$$

В дальнейшем будем предполагать, что функция распределения нормирована:

$$\int_0^{\infty} D_k(R) dR = 1. \quad (2.72)$$

Определим интенсивность рассеяния от смеси как линейную комбинацию парциальных интенсивностей с соответствующими объемными долями  $V_k$ :

$$I(s) = \sum_{k=1}^K V_k I_k(s) \quad (2.73)$$

где интенсивность рассеяния компонентами, соответственно

$$I_k(s) = \int_0^{\infty} D_k(R) v_k(R) [\Delta\rho_k(R)]^2 i_k(s, R) S_k(s) dR. \quad (2.74)$$

Аналитическое выражение для парциальной функции объемного распределения  $D_k(R)$  в данной работе представлено нормированным распределением Шульца [204,205], описываемым двумя параметрами, средним значением  $R_{0k}$  и дисперсией  $\Delta R_k$ :

$$D_k(R) = G(R, R_{0k}, \Delta R_k) = \left( \frac{z+1}{R_{0k}} \right)^{z+1} \frac{R^z}{\Gamma(z+1)} \exp \left[ - \frac{(z+1)R}{R_{0k}} \right] \quad (2.75)$$

где  $z = (R_{0k} / \Delta R_k)^2 - 1$ .

Чтобы определить функции  $\Delta \rho_k(R)$ ,  $v_k(R)$ ,  $i_k(s, R)$  и  $S_k(s)$ , необходимо иметь априорное знание о форме частиц компонент. В данной работе использовали формфакторы [126] однородной сферы с контрастом, или плотностью (электронной или массовой)  $\rho$

$$i_{sphere}(s, R) = \rho^2 V^2 [A_{sphere}(sR)]^2 = \rho^2 \left( \frac{4}{3} \pi R^3 \right) \cdot \left[ \frac{\sin(sR) - sR \cdot \cos(sR)}{(sR)^3} \right]^2, \quad (2.76)$$

двухфазной концентрической сферы с внешним радиусом  $R_2$ , плотностью внешней части  $\rho_2$ , внутренним радиусом  $R_1$  и плотностью ядра  $\rho_1$ :

$$i_{shell}(s, R_1, R_2) = \left\{ \rho_1 [V_1 A_{sphere}(s, R_1) - V_2 A_{sphere}(s, R_2)] + \rho_2 V_2 \cdot A_{sphere}(s, R_2) \right\}^2, \quad (2.77)$$

$$V = \frac{4}{3} \pi R^3, \quad R_2 > R_1,$$

трехосного эллипсоида с полуосями  $a, b, c$ :

$$i_{ell}(s, a, b, c) = \rho^2 \int_0^1 \int_0^1 A_{sphere}^2 \left\{ s, \left[ \begin{array}{l} a^2 \cos^2 \left( \frac{\pi x}{2} \right) + \\ + b^2 \sin^2 \left( \frac{\pi x}{2} \right) \cdot (1 - y^2) + \\ + c^2 y^2 \end{array} \right]^{1/2} \right\} dx dy, \quad (2.78)$$

кругового коаксиального двухфазного цилиндра длиной  $H$ , внутренним радиусом  $R_1$ , внешним  $R_2$ , и плотностями  $\rho_1$  и  $\rho_2$ , соответственно:

$$i_{cyl}(s, R) = \int_0^1 \left[ W \cdot \frac{\sin(t)}{t} \right]^2 \cdot dx, \quad t = \frac{s \cdot H \cdot x}{2},$$

$$W = \frac{1}{1-g^2} \cdot \{ \rho_1 \cdot V_1 \cdot Q(s \cdot R_1 \cdot y) + \rho_2 \cdot [V_2 \cdot Q(s \cdot R_2 \cdot y) - V_1 \cdot Q(s \cdot R_1 \cdot y)] \}, \quad (2.79)$$

$$g = \frac{R_1}{R_2} < 1, \quad y = (1-x^2)^{1/2}, \quad Q(u) = \frac{2 \cdot J_1(u)}{u}, \quad V = \pi \cdot R^2 \cdot H,$$

$J_1(u)$  - функция Бесселя.

В программе были предусмотрены и другие формфакторы, такие, как параллелепипед, составное тело из двух сфер с заданным расстоянием между центрами, и т.д.

Для расчета интенсивности рассеяния автором была разработана библиотека программ, включающая специальные функции и процедуры численного интегрирования. Особое внимание было уделено точности и скорости расчетов. Численные эксперименты показали сильную зависимость скорости сходимости поиска решения от "шумов вычислений", поэтому в качестве процедур численного интегрирования были применены хорошо зарекомендовавшие себя высокоточные адаптивные алгоритмы Гаусса-Кронрода и Tanh-sinh квадратура [206], оптимизированные для обеспечения максимальной скорости вычислений.

Для определения объемных долей и размерных параметров частиц решали задачу подгонки данных эксперимента полидисперсной моделью с помощью нелинейной минимизации функционала наименьших квадратов, аналогично (2.31):

$$\min_{V_k, R_{0,k}, \Delta R_k, \Delta \rho_k} \left\{ \frac{\sum_{i=1}^N [(I_{\text{exp}}(s_i) - \xi \cdot I(s_i)) \cdot W(s_i)]^2}{\sum_{i=1}^N [I_{\text{exp}}^2(s_i) \cdot W^2(s_i)]} \right\}, \quad k = 1, \dots, K \quad (2.80)$$

$$W(s_i) = [\tilde{I}_{\text{exp}}(s_i)]^p, \quad p = 1, -\frac{1}{2}, -\frac{2}{3}, -\frac{3}{4},$$

где  $N$  число экспериментальных точек,  $K$  - число парциальных распределений (компонентов),  $\Delta\rho_k$  - контраст, а  $\tilde{I}_{\text{exp}}(s_i)$  есть сглаженная интенсивность. Нормировка на сглаженную экспериментальную интенсивность позволяет проводить поиск решения в относительных единицах интенсивности рассеяния, а применение совмещающего множителя  $\zeta$  (2.32) обеспечивает сглаживание многих локальных минимумов целевой функции, делая нахождение решения задачи практически возможным. Данный факт был подтвержден результатами нескольких сотен модельных расчетов по теоретическим данным рассеяния от систем смесей полидисперсных однородных и полых сферических и цилиндрических частиц.

Для назначения оптимального динамического диапазона данных рассеяния применяют весовую функцию  $W(s)$ , соответствующую степенному преобразованию интенсивности рассеяния (2.80), то есть невязку можно рассчитывать в шкалах  $I(s)$ ,  $I^{1/2}(s)$ ,  $I^{1/3}(s)$  или  $I^{1/4}(s)$  при отношении  $I_{\text{max}}/I_{\text{min}} = 10-50$ . Использование в качестве сглаживающей функции (при расчете  $W$ ) оконной функции Стеклова [207]

$$\tilde{I}_{\text{exp}}(s_i) = \frac{1}{h} \int_{i-h/2}^{i+h/2} I_{\text{exp}}(s_i) \cdot (\Delta s), \quad (2.81)$$

где  $\Delta s$  - шаг угловой шкалы и  $h$  - ширина окна сканирования, позволяет эффективно работать с неэквидистантной угловой сеткой, в отличие от родственного метода бегущей средней. Полиномиальные и частотные сглаживающие алгоритмы в данном случае, из-за опасности получения отрицательных значений  $\tilde{I}_{\text{exp}}(s_i)$  при малых, углах применять не рекомендуется. Задача сглаживания - уменьшение амплитуды шумов измерений, без внесения существенной систематической ошибки (контролируется по величине критерия автокорреляции остатков Дарбина-Ватсона) и артефактов в виде отрицательной интенсивности. Выполнение этих критериев обеспечивается выбором ширины окна сглаживания  $h$ , которое обычно захватывает 5-15 экспериментальных точек, если шумы измерений носят случайный характер.

Поиск параметров распределений в программе MIXTURE проводится с помощью многовариантного оптимизационного программного пакета OPTIS [A23]. Как было отмечено, такой подход имеет значительное преимущество перед методом расчета, основанном на поиске гладкого контура методом регуляризации.

Автором был разработан другой вариант программы (POLYMIX), в которой применен метод минимизации NL2SOL (см. Раздел 2.2.2). Применение этого алгоритма позволило сократить число вычислений целевой функции в 5-10 раз. Применение высококачественных алгоритмов численного интегрирования в расчетах интенсивности рассеяния от формы частиц позволило в несколько раз повысить точность аппроксимации данных рассеяния.

Была проведена еще одна существенная модификация программы, которая позволила повысить устойчивость расчетов распределений с большой дисперсией размеров частиц. Эта модификация состоит в том, что расчет невязки (2.73) проводят на неэквидистантной угловой сетке. В случае эквидистантной сетки, в процессе поиска решения, возможно совпадение периода осцилляций модельных кривых рассеяния от частиц с периодом сетки и, как следствие, резкое уменьшение численного значения градиента целевой функции, приводящее к необоснованно большой величине шага по параметрам модели. Это приводит к нарушению хода поиска и большой зависимости решения от стартовых значений параметров модели. Практика показала, что хорошие результаты дает сетка с шагом, увеличивающимся по закону геометрической прогрессии. Дополнительное преимущество данного подхода состоит в том, что нет необходимости задавать частую угловую сетку на больших углах рассеяния, в области, где основной вклад в рассеяние дают малые частицы и ширина шенноновского канала  $N_S$  (2.28) намного больше, чем в малых углах. В программе POLYMIX реализовано определение ширины шенноновского канала по радиусам инерции  $R_g$  (в предположении сферических частиц, которое, как показала практика, дает удовлетворительные результаты и в случае анизометричных тел), рассчитанным по касательным к кривой рассеяния в координатах  $\ln(I)$  от  $s^2$  в начальном участке

и в области максимальных углов. Величины шагов назначают как  $\frac{\pi}{3 \cdot R_g}$ , где коэффициент 3 выбран на основании анализа результатов моделирования.

По найденным величинам шагов программа автоматически рассчитывает угловую сетку с увеличивающимся в геометрической прогрессии угловым интервалом. Расчет невязки на такой сетке обладает еще одним преимуществом, так как позволяет лучше подогнать моделью начальный участок интенсивности рассеяния, в котором располагается значительное количество точек, что улучшает качество аппроксимации.

Для проверки работы программ поиска распределений были разработаны программы расчета интенсивности рассеяния от полидисперсных систем однородных и полых сферических, эллиптических и цилиндрических частиц, по аналитическим выражениям (2.73 - 2.79) при заданных формфакторах, параметрах и относительных долях распределений. Для перекрестной проверки реализации расчетных формул и подбора оптимальных параметров алгоритмов были также разработаны программы генерации смесей полидисперсных частиц методом шариковых моделей с заданными распределениями по размерам. Расчет формфакторов проводится в этом случае по формулам (2.60, 2.61), аналогично тому, как это реализовано в программе DAMMIN. По результатам около 0 экспериментов были выбраны оптимальные для большинства практических случаев значения параметров алгоритмов поиска и правила установки стартовых значений параметров моделей. Для большинства полидисперсных систем рекомендуется использовать от 2 или 3 компоненты, в общем случае сферических частиц, если априорная оценка возможной анизотрии рассеивающих неоднородностей (отношение максимального размера формы неоднородности или частицы к минимальному) не превышает 2-3. В результате поиска по сферическим формфакторам мы получаем оценку распределения по радиусам эквивалентных по объему сфер, которая, как правило, адекватно передает действительное распределение. Цифры точности в общем случае указать трудно, но относительная ошибка в оценках параметров распределений для модельных

систем с анизотропией до 3:1 редко превышала 20%. На практике, точность необходимо исследовать путем моделирования ожидаемых параметров и сравнения вариантов решений. В качестве критериев сравнения можно использовать коэффициент корреляции контуров распределений и другие критерии, рассмотренные в первой части диссертационной работы.

Моделирование показало, что число обусловленности задачи поиска распределений довольно велико, более  $10^4 - 10^5$  при 2-х компонентах и  $10^5 - 10^7$  при  $K = 3$ . Это приводит к возможной сильной зависимости результата от стартового приближения. Чтобы повысить надежность ответа необходимо, как и в случаях задач рассмотренных ранее, проводить "мультистарт": получать набор решений, не менее 10, при различных стартовых значениях параметров, варьируя их случайным образом с дисперсией 30-50% относительно приближения, найденного автоматически из оценок максимального и минимального радиусов инерции (уравнение (2.4)) рассеивающих неоднородностей, рассматриваемых как "частицы". Конечно, эти оценки являются приближенными, их точность в случае систем с экспоненциальным или фрактальными распределениями может быть хуже 50%, что и определяет величину вариации стартовых параметров в численных экспериментах. Особенно велика ошибка в оценке радиуса инерции малых частиц, так как график Гинье проводят в области максимальных углов рассеяния, в которых определяющий вклад может вносить фон атомного и межмолекулярного рассеяния. Поэтому компоненту с малым средним радиусом распределения вводят именно для учета такого фона без специальной структурной интерпретации результата.

На Рисунке 2.67 показаны примеры анализа данных рассеяния от образца битума. Рассеяние в данной системе обусловлено в основном наличием кластеров неопределенных соединений, которые расположены довольно близко друг от друга, о чем говорит широкий интерференционный пик в области  $2.5 \text{ nm}^{-1}$ . Поэтому при расчетах вводили структурные факторы для двух компонент (использовали 3 системы сферических частиц, 3-я система описывала фон рассеяния от малых образований с размерами менее 1 нм) в приближении Перкуса-Йовика, параметры

которых также были предметом поиска. Комментарии даны в подписи под рисунком. Из рисунка видно преимущество предложенного подхода с расчетом невязки на неэквидистантной угловой сетке.  $\chi^2$  для эквидистантной сетки был 19.0, для неэквидистантной 3.0. Отличие его от 1.0 связано с осциллирующим поведением функции структурного фактора, которое свойственно аппроксимации Перкуса-Йовика. К сожалению, более общие выражения для функций межчастичной интерференции в до настоящего времени получить не удалось.

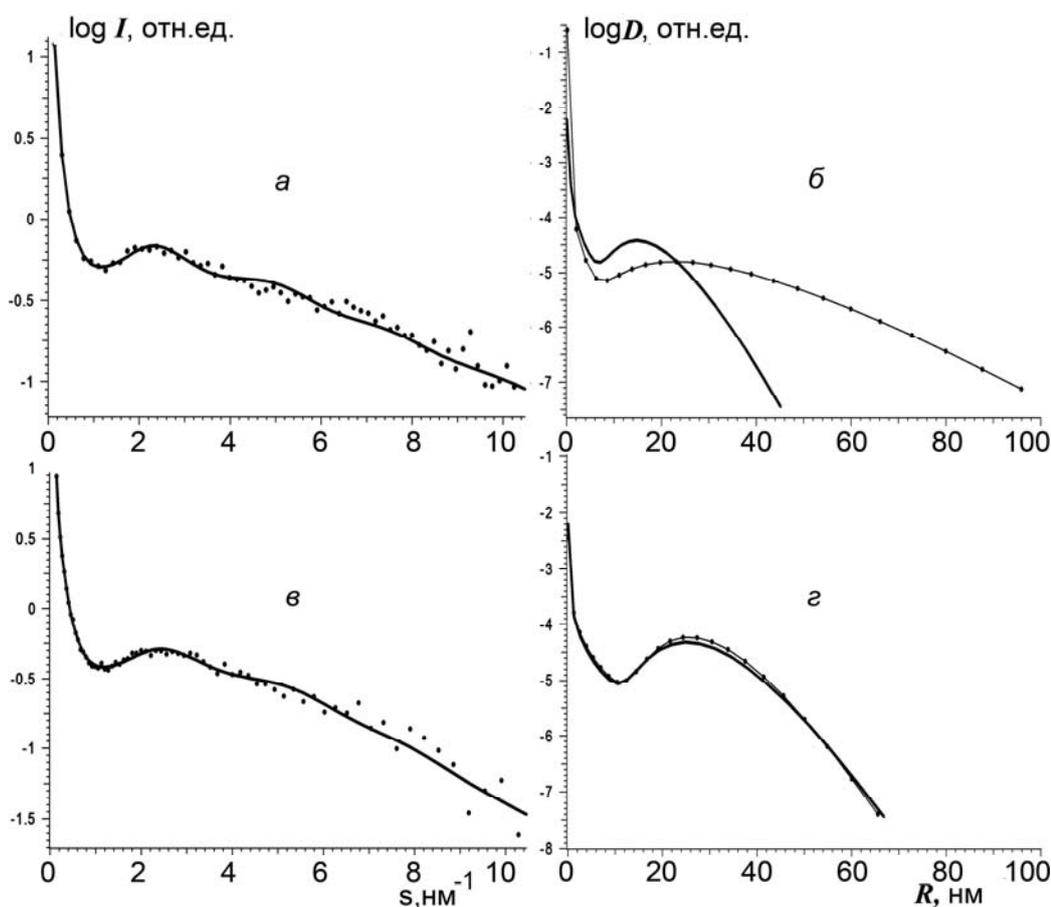


Рисунок 2.67 - Сравнение распределений по размерам кластеров непердельных органических соединений, рассчитанных по данным малоуглового рассеяния от образца природного битума из месторождения г. Грозного. Слева показаны экспериментальные (точки) и модельные (сплошные линии) интенсивности рассеяния на угловых сетках, использованных при расчете невязки:

**а)** - эквидистантная, **в)** - предложенная в работе неэквидистантная.

Справа - показаны по два решения, полученные при разных стартовых аппроксимациях (наборы стартовых значений одинаковы для **б)** и **г)**). Модельные кривые рассеяния от двух показанных решений на графиках неразличимы, несмотря на различия распределений. Осцилляции на кривых рассеяния в области средних углов связаны с поведением структурного фактора, рассчитанного в приближении Перкуса-Йовика для потенциала жестких сфер.

Программы поиска распределений MIXTURE и POLYMIX интенсивно использовали в исследованиях образцов различной природы: самоорганизующихся полимеров [A38,A41,A43,A50,A57,A62], биоцеллюлозы [A42,A53], металлоорганических комплексов [A33,A51], мицеллярных частиц [A40], в исследованиях композитных материалов, содержащих металлические и другие неорганические наночастицы [A24,A25,A32,A37,A45–A49,A52,A54–A56,], неорганических пленочных систем [A60].

### 2.3.1 Применение метода моделирования распределений: анализ микроэмульсий

Методом малоуглового рентгеновского рассеяния (МУР) исследована микроструктура трехкомпонентных микроэмульсий вида вода/поверхностно-активное вещество ди-2-этил-гексил-сульфосукцинат натрия (АОТ)/изооктан при различных температурах (15-60°C), концентрациях воды (2-20 объемных %), с разным молярным отношением вода/АОТ (25-45). Для анализа данных МУР был разработан общий подход моделирования экспериментальной интенсивности в виде суммы вкладов от рассеяния на полидисперсных системах частиц, имеющих простую геометрическую форму с учетом эффектов межчастичной интерференции. Для расчета интенсивностей рассеяния была выбрана трехфазная модель, состоящая из полидисперсных цилиндрических и двух типов сферических частиц. Обнаружено, что в диапазоне температур 30-40°C происходит значительное увеличение доли цилиндрических частиц. Проведенное моделирование дало возможность определить характерные размеры, форму и полидисперсность частиц в исследуемых системах, что важно при исследовании особенностей фазовых переходов в трехкомпонентных микроэмульсиях вида вода/АОТ/изооктан.

Микроэмульсии являются важными представителями самоорганизующихся систем. Они представляют собой смесь трех компонент: полярной и неполярной жидких фаз (обычно это вода и углеводород) и поверхностно-активного вещества. Микроэмульсионная фаза является изотропной и термодинамически устойчивой при комнатной температуре. В этой фазе молекулы воды объединяются в сферические капли, окруженные мономолекулярным слоем поверхностно-активного вещества. В данной работе в качестве поверхностно-активного вещества использовался ди-2-этил-гексил-сульфосукцинат натрия (АОТ). Молекула АОТ состоит из гидрофильной сульфатной группы и двух неполярных углеводородных хвостов. Площадь, занимаемая молекулой АОТ в слое на поверхности капли равна приблизительно  $0,55 \text{ нм}^2$ . Размер одной капли зависит

от молярного отношения воды к АОТ -  $W_0$ . Для типичного значения  $W_0=25$  одна микроэмульсионная капля представляет собой сферу радиуса 4,2 нм, содержащую приблизительно 10000 молекул воды и покрытую 400 молекулами АОТ.

Модель поперечного сечения капли, использованная в работе, показана на Рисунке 2.68.

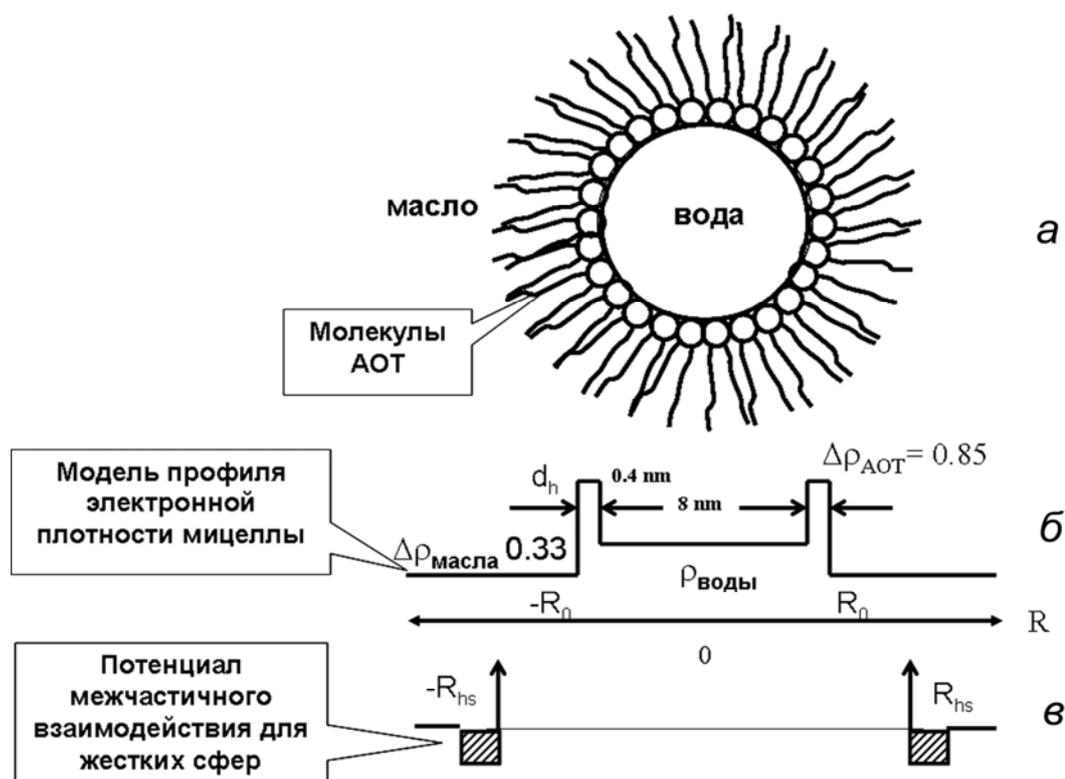


Рисунок 2.68 - Поперечное сечение модели микроэмульсионной капли (а), профиль электронной плотности  $\rho(r)$  (контраст в  $e/\text{Å}^3$ ) (б) и межчастичный потенциал жестких сфер с поверхностной адгезией (в). Заштрихован участок притяжения потенциала.  $d_h$  - размер головной части молекулы АОТ,  $R_{\text{hs}}$ -радиус жесткой сферы,  $R_0$ -размер микроэмульсионной капли.

Капли не являются идентичными, их полидисперсность по размеру равна примерно 20%. Полидисперсность является результатом тонкого баланса противоположных вкладов в свободную энергию от таких величин, как энергия двойного заряженного слоя, поверхностное натяжение и энтропия системы. Отличительным свойством микроэмульсий АОТ является возможность изменения в большом диапазоне концентраций водяных капель без существенных изменений их среднего размера и полидисперсности.

В данной работе представлены результаты анализа данных малоуглового рентгеновского рассеяния микроэмульсией в широком диапазоне температур (от 15 до 60 °С), объемных концентраций воды (от 2 до 20%) и молярном отношении  $W_0=25-45$ , полученные с помощью разработанного нами общего метода расчета кривых интенсивности от многокомпонентной системы с учетом эффекта межчастичной интерференции.

Результаты подгонки экспериментальных данных двухкомпонентной моделью для микроэмульсии с молярным отношением вода/АОТ 25 и 45 при концентрации воды 2% показаны штриховой линией на Рисунке 2.69. Как видно из рисунка, с помощью 2х-компонентной модели не удается добиться полного согласия между экспериментальными и модельными интенсивностями особенно при больших значениях  $s$ . Введение в модель третьей фазы - сферических агрегатов из молекул АОТ с размерами от 1,0 до 1,5 нм позволило получить значительно лучшее согласие между экспериментальными и теоретическими интенсивностями. Таким образом, к набору варьируемых параметров добавились объемная доля малых сфер  $V_{мсф}$ , средний радиус  $R_{мсф}$  и полидисперсность  $dR_{мсф}$  сферических агрегатов АОТ.

Использование дополнительной компоненты позволило значительно улучшить согласие между экспериментальными кривыми и модельными кривыми рассеяния, что видно из рисунков, где сплошными линиями показаны рассчитанные с помощью трехкомпонентной модели интенсивности рассеяния. Результаты подгонки трехкомпонентной моделью приведены в Таблице 2.6. Как видно из таблицы, в исследуемых микроэмульсионных системах между 30 и 40 °С происходит структурный переход с образованием цилиндрических мицелл, причем если для микроэмульсии с молярным отношением  $W_0=25$  образование цилиндрической фазы сопровождается уменьшением доли малых сферических АОТ агрегатов, то для микроэмульсии с  $W_0=45$  ее образование сопровождается уменьшением доли больших микроэмульсионных капель (вода+АОТ).

Общим результатом для всех серий является непрерывное постепенное уменьшение среднего радиуса больших сферических капель ( $R_{сф}$ ) и среднего

радиуса цилиндрических мицелл ( $R_{цмл}$ ), в то время как отношение  $R_{сф}/R_{цмл} \approx 0,7 \div 0,8$  остается практически неизменным. Полидисперсность больших сферических капель составляет приблизительно 20% и слабо меняется при повышении температуры. Степень полидисперсности цилиндрических мицелл увеличивается при температурах 40-45 °С, как следует из таблицы 1.

Если при небольших концентрациях воды (для 2 и 5 объемных %) межчастичная интерференция слабо влияет на форму кривой рассеяния, то для больших концентраций (15-20%) ее учет оказывается существенным. На Рисунке 2.70 штриховыми линиями показаны кривые интенсивности для микроэмульсии с концентрацией воды 20%, рассчитанные без учета межчастичного взаимодействия. Видно, что при низких температурах экспериментальные кривые сильно отличаются от модельных, особенно при низких температурах. Для описания взаимодействия между микроэмульсионными каплями нами был взят потенциал жестких сфер с поверхностной адгезией, впервые предложенный в [202]. Данный потенциал описывается двумя параметрами: радиусом жесткой сферы ( $R_{hs}$ ), определяющим отталкивающую часть потенциала и параметром поверхностной адгезии, ответственным за короткодействующее притягивание. Вид потенциала показан на Рисунке 2.68. Радиус жесткой сферы  $R_{hs}$  превышает средний эффективный радиус  $R_0$  микроэмульсионной капли из-за длинных углеводородных хвостов (0,8 нм) молекул АОТ.

После учета межчастичного взаимодействия нам удалось добиться хорошего согласия экспериментальных данных с модельными во всем диапазоне температур. Кривые интенсивности, рассчитанные с учетом межчастичного взаимодействия, показаны на Рисунках 2.69 и 2.70 сплошными линиями. Параметр адгезии для  $W_0=25$  системы увеличивается с повышением температуры, тогда как для  $W_0=45$  системы параметр адгезии пренебрежимо мал во всем диапазоне температур (Таблица 2.6).

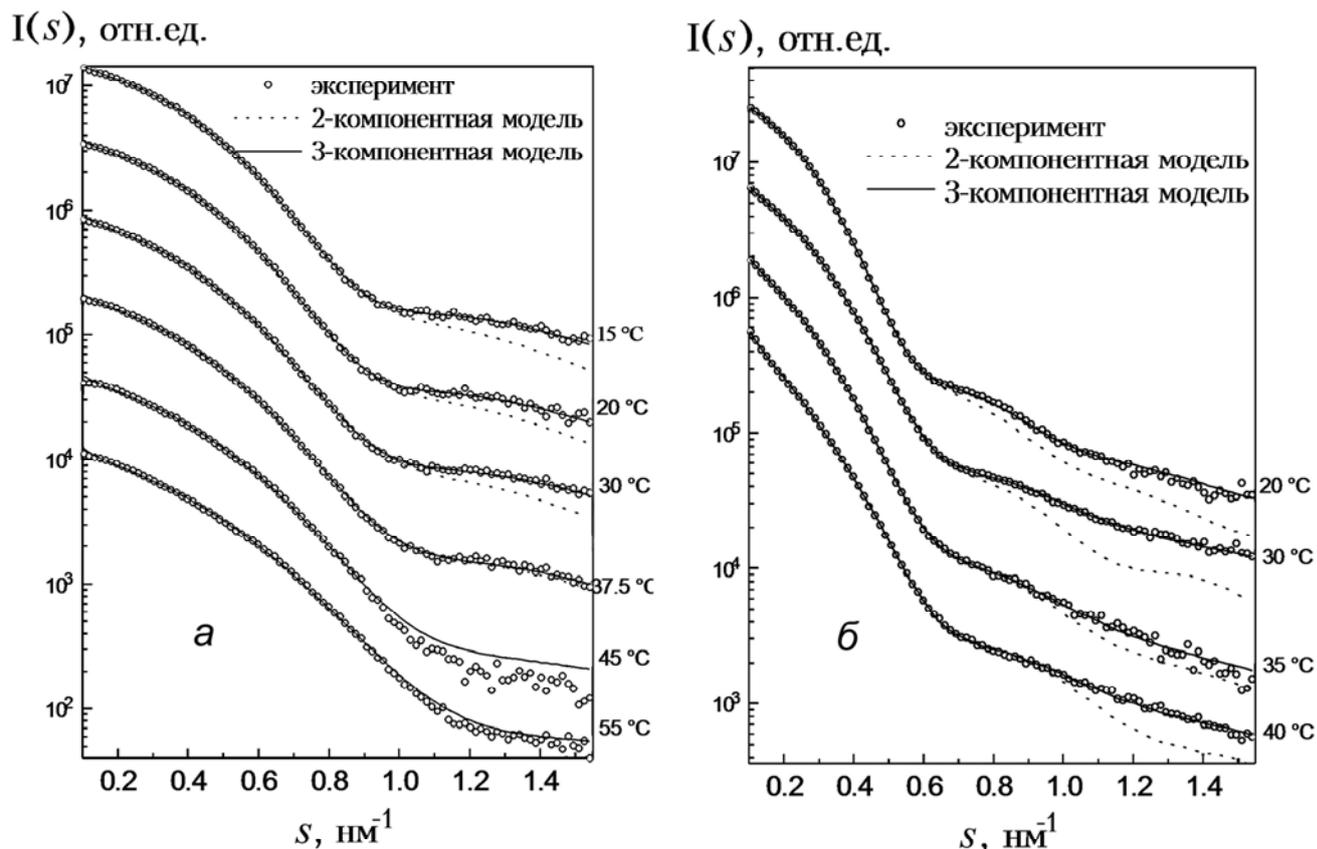


Рисунок 2.69 – Экспериментальные интенсивности (светлые кружочки) малоуглового рассеяния микроэмульсии АОТ/вода/изооктан с концентрацией воды 2%. Рисунок слева: диапазон температур 15-55 °C,  $W_0=25$ . Рисунок справа: диапазон 20-40 °C,  $W_0=45$ . Интенсивности рассчитанные в рамках 2х-компонентной модели показаны штриховыми линиями, а в рамках 3х-компонентной модели - сплошными линиями.

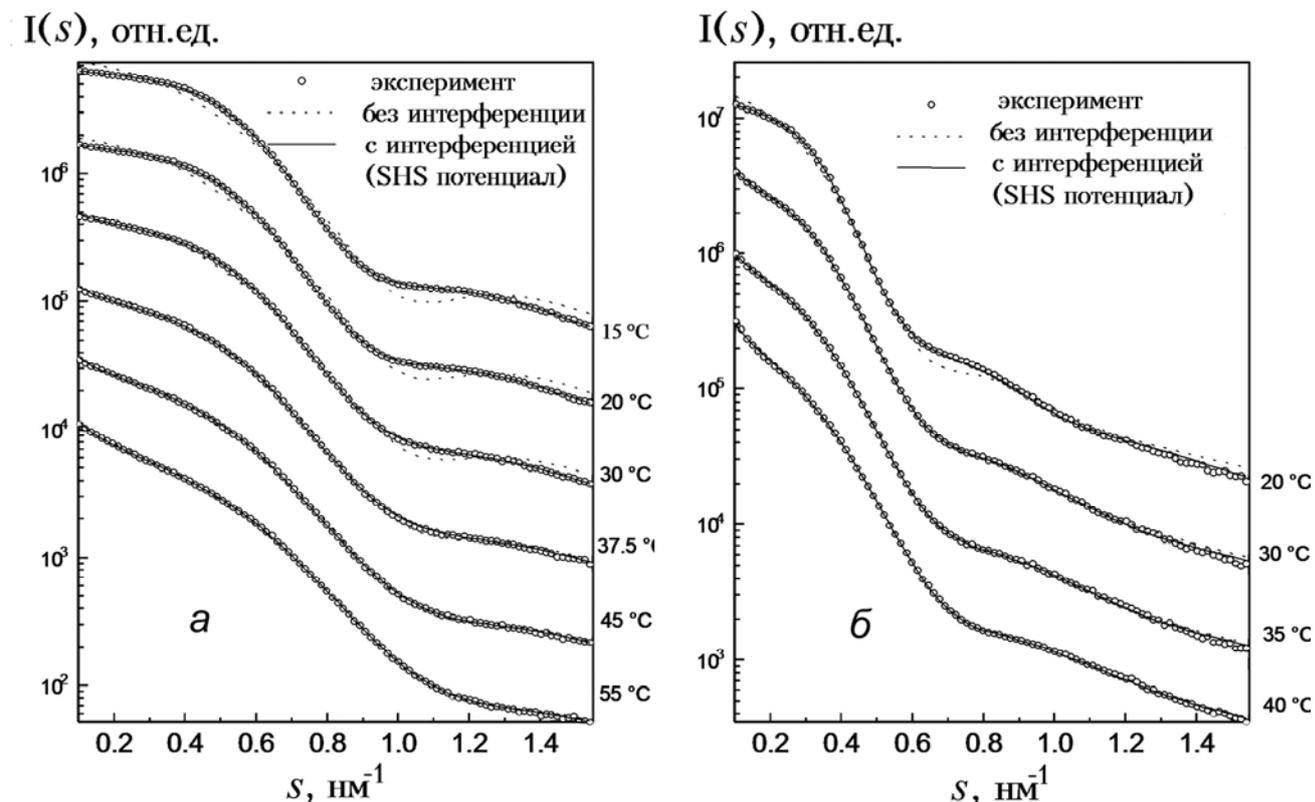


Рисунок 2.70. Экспериментальные интенсивности (светлые кружочки) малоуглового рассеяния микроэмульсии АОТ/вода/изооктан с концентрацией воды 20%. Рисунок слева: диапазон температур 15-55 °С,  $W_0=25$ . Рисунок справа: диапазон 20-40 °С,  $W_0=45$ . Интенсивности рассчитанные в рамках 3х-компонентной модели без учета интерференции показаны штриховыми линиями, а с учетом эффекта межчастичной интерференции - сплошными линиями.

Таким образом, в данной работе исследовались трехкомпонентные вода-АОТ-изооктан микроэмульсии с помощью малоуглового рентгеновского рассеяния в широком диапазоне температур (15-55 °С), концентрации воды (2-20%) и молярного отношения вода/АОТ ( $W_0=25-45$ ). Было найдено, что между 30 и 40 °С в исследуемой системе происходит структурный переход от сферических микроэмульсионных капель к нитевидным цилиндрическим мицеллам. С помощью 3х-компонентной модели, учитывающей эффект межчастичной интерференции, удалось адекватно описать экспериментальные наборы данных, снятые при различных условиях, и количественно охарактеризовать поведение

основных параметров системы при структурном переходе. Наши результаты хорошо согласуются с предыдущими исследованиями малоуглового рентгеновского и нейтронного рассеяния, а также с измерениями проводимости, диэлектрической проницаемости, вязкости и флуоресценции. Найденные структурные параметры частиц даны в Таблицах 2.6 (а-г). В этих таблицах  $P$  означает объемную долю компоненты.

Таблица 2.6а. АОТ/Н<sub>2</sub>О/изооктан  $W_0=25$ ,  $\varphi_w=2\%$ 

T, °C	$P_{сф}$	$P_{цил}$	$P_{мсф}$	$R_{сф}$ , нм	$dR_{сф}$ , нм	$R_{цил}$ , нм	$dR_{цил}$ , нм
15	0,68	0,07	0,25	4,2	0,8	3,1	0,10
20	0,79	0,07	0,14	4,1	0,9	3,2	0,10
30	0,70	0,09	0,21	4,0	0,9	3,1	0,10
37,5	0,84	0,11	0,05	3,9	0,9	3,0	0,10
45	0,79	0,21	0,00	3,8	0,8	2,5	0,10
55	0,77	0,23	0,00	3,6	0,9	2,4	0,10

Таблица 2.6б АОТ/Н<sub>2</sub>О/изооктан  $W_0=45$ ,  $\varphi_w=2\%$ 

T, °C	$P_{сф}$	$P_{цил}$	$P_{мсф}$	$R_{сф}$ , нм	$dR_{сф}$ , нм	$R_{цил}$ , нм	$dR_{цил}$ , нм
20	0,73	0,05	0,22	6,9	1,5	6,1	0,1
30	0,52	0,14	0,34	6,3	1,4	5,2	0,7
35	0,52	0,32	0,16	6,4	1,7	5,1	0,9
40	0,22	0,48	0,30	5,7	1,6	5,0	1,0

Таблица 2.6в АОТ/Н<sub>2</sub>О/изооктан  $W_0=25$ ,  $\varphi_w=20\%$ 

T, °C	$P_{сф}$	$P_{цил}$	$P_{мсф}$	$R_{сф}$ , нм	$dR_{сф}$ , нм	$R_{цил}$ , нм	$dR_{цил}$ , нм	$\tau$
15	0,85	0,02	0,13	4,2	0,8	2,8	0,1	2,76
20	0,84	0,04	0,12	4,1	0,8	2,8	0,1	0,99
30	0,92	0,05	0,03	4,1	0,9	3,1	0,1	0,41
37,5	0,81	0,12	0,07	4,0	1,0	2,9	0,1	0,84
45	0,84	0,16	0,00	3,9	1,1	2,8	0,1	0,46
55	0,61	0,39	0,00	3,7	0,8	2,3	0,4	0,17

Таблица 2.6г АОТ/Н<sub>2</sub>О/изооктан  $W_0=45$ ,  $\varphi_w=20\%$ 

T, °C	$P_{сф}$	$P_{цил}$	$P_{мсф}$	$R_{сф}$ , нм	$dR_{сф}$ , нм	$R_{цил}$ , нм	$dR_{цил}$ , нм	$\tau$
20	0,82	0,03	0,15	7,0	1,6	5,8	0,1	100
30	0,72	0,18	0,10	6,6	1,6	4,8	0,6	100
35	0,60	0,27	0,13	6,5	1,7	4,6	0,7	100
40	0,42	0,42	0,16	6,0	1,5	4,4	0,9	100

В заключение можно показать сравнение результатов расчета распределения для кривой рассеяния Рисунка 2.70б, 35° с помощью программы косвенного преобразования Фурье с регуляризацией GNOM [169] и программы

моделирования распределениями Шульца MIXTURE. Рисунок 2.71а демонстрирует достаточно хорошее соответствие результатов расчетов, если в программе MIXTURE вводить двухкомпонентную модель. Осцилляции на распределении, полученном с помощью программы GNOM связаны с эффектами обрыва при фурье-преобразовании, от которых избавиться на практике невозможно. Распределение, найденное программой MIXTURE представляет собой "сглаженное" распределение GNOM, что показывает эффективность подхода с использованием суперпозиции аналитических распределений. Однако, трехкомпонентная модель, как было сказано выше, показывает лучшее согласие с экспериментальными данными и типичный вид получаемых распределений приведен на Рисунке 2.71б. Неустойчивость решения может быть продемонстрирована путем расчета при разных весовых функциях, используемых при вычислении невязки в (2.80). Из Рисунка 2.71б видно, что если средний радиус компонент остается достаточно стабильным, вариациям подвергается величина дисперсии их распределений.

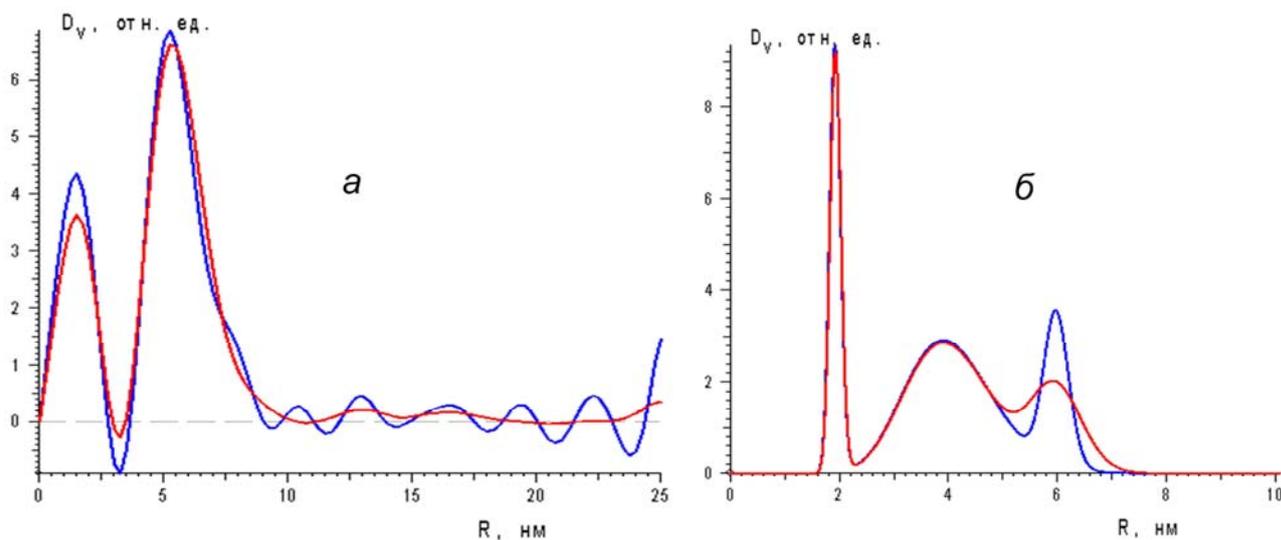


Рисунок 2.71 – Результаты расчета объемного распределения по радиусам компонент. а): Синяя кривая – расчет методом косвенного преобразования Фурье (программа GNOM [169]), красная кривая – программа MIXTURE. б): Сравнение решений программы MIXTURE при двух способах расчета невязки: синяя кривая – вес с  $p = 1/2$ , красная кривая –  $p = 1$  (формула 2.80).

В заключение раздела можно привести выводы, аналогичные сделанным в конце раздела 1, а именно – при решении плохо обусловленных обратных задач, часто в условиях неопределенности, для оценивания надежности и стабильности решения необходимо проводить дополнительные численные эксперименты с варьированием не только входных данных, но и параметров алгоритмов поиска решения.

## Выводы

1. Разработаны алгоритмы и программы решения обратных задач анализа наборов аддитивных неотрицательных спектров смесей, основанные на сингулярном разложении матрицы данных. Разработанные алгоритмы включают оценку числа независимых компонентов в смесях, определение контуров индивидуальных спектров и способы исследования устойчивости решений. Показана возможность анализа систем с числом априори неизвестных компонентов более 3-4, в зависимости от качества экспериментальных данных. Разработанные подходы реализованы в виде компьютерных программ и применены для анализа ИК спектров смесей органических соединений, колебательных спектров жидкой воды, анализа серий измерений интенсивности малоуглового рентгеновского рассеяния.

2. Разработанные способы разложения аддитивных спектров впервые применены, в частности, для анализа и коррекции данных малоуглового нейтронного рассеяния с вариацией контраста, по которым были определены структурные модели селективно дейтерированных частиц рибосомы 70S *E-Coli* и 30S *Thermus Thermophilus*.

3. С помощью модельных расчетов проведен анализ единственности и устойчивости решения задач поиска формы частиц в монодисперсных системах по данным малоуглового рентгеновского рассеяния. Определены параметры анизометрии простых геометрических форм, являющиеся пограничными для устойчивых решений.

5. Программа поиска шариковых моделей частиц по данным малоуглового рассеяния модифицирована способами автоматического определения оптимальных значений параметров и режимов алгоритма поиска, что позволило уменьшить время расчетов в 3-5 раз и уменьшить разброс решений. С помощью разработанного программного обеспечения по данным малоуглового рассеяния от раствора определены формы молекул иммуноглобулина М и ревматоидного фактора человека. Показано, что различие между макромолекулами состоит в

меньшей заселенности периферийных участков молекул ревматоидных факторов. Выдвинута гипотеза о вероятной большей гибкости периферийных фрагментов ревматоидного фактора, что независимо подтверждено другими методами исследований.

6. Разработана модифицированная схема поиска размерных распределений частиц по данным малоуглового рассеяния с повышенной устойчивостью решений. Повышение устойчивости достигнуто путем расчета невязки между экспериментальными и модельными интенсивностями рассеяния на неравномерной угловой сетке с определением шага сетки из оценок максимального и минимального радиусов инерции частиц в системе. Проведен анализ ряда систем с наночастицами – металлических катализаторов на органических и углеродных носителях, органико-неорганических композитов, полимерных систем.

## Приложение

### Программное обеспечение для решения задач анализа смесей и данных малоуглового рассеяния

#### П.1 Базовый диалоговый пакет оптимизации

Основные идеи, реализованные в программе OPTIS:

1) Модульность: в программе реализован набор известных алгоритмов, которые максимально эффективны в своем классе и обеспечена легкость добавления новых алгоритмов. Из методов нулевого порядка: координатный спуск, метод деформируемого многогранника Нелдера-Мида, метод Монте-Карло, методы моделирования отжига для схемы Монте-Карло и деформируемого многогранника. Из градиентных методов: методы сопряженных градиентов (2 схемы), методы переменной метрики (3 схемы), в том числе с простыми ограничениями на переменные, метод решения задач нелинейных квадратов на основе модифицированной схемы Левенберга-Маркардта, усиленной применением схемы уточнения матрицы вторых производных по алгоритму переменной метрики с учетом простых ограничений на переменные.

2) Возможность смены схемы минимизации в процессе поиска решения.

3) Диалоговый режим. При этом диалог по выбору оператора может переключаться в пассивный или активный режим. Пассивный режим означает, что программа ведет оператора, задавая вопросы, требующие обязательного ответа, при активном режиме оператор отдает команды программе, которая пытается их выполнить, а в случае невозможности выполнения временно переходит в пассивный режим, до тех пор, пока требуемое действие не станет выполненным. Этот режим обеспечивает максимальную эффективность диалога при исследовании сложных для минимизации функций.

4) Определение точности вычисления функции. Правильная оценка точности критична для градиентных методов минимизации.

5) Автоматическое определение параметров методов минимизации:

величин оптимальных начальных шагов по переменным, величин конечных разностей для расчета градиентов и вторых производных, критериев окончания поиска. Все параметры оцениваются, исходя из найденной точности вычисления функции для текущих значений аргументов.

б) Возможность исследования качества полученного решения путем построения сечений (одномерных "проколов") целевой функции в окрестности текущей точки по координатным осям или вдоль заданного направления.

## **П.2 Пакет для анализа наборов аддитивных спектров смесей**

Пакет программ основан на диалоговой программе минимизации нелинейных функций OPTIS.

1) Реализована оценка числа компонентов путем расчета сингулярных чисел матрицы спектроскопических данных и статистического анализа последовательности элементов левых сингулярных векторов.

2) Поиск спектров компонентов с заданием числа компонентов, размерности пространства решений, относительных весов для максимальной гладкости индивидуальных спектров, их различия, неотрицательности.

3) Диалоговый режим.

4) Вывод графической информации для принятия оператором решений о выборе метода поиска решения, определения числа компонентов и оценки качества решения.

5) Предложена схема определения случайной составляющей ошибок измерений методом Фурье-сглаживания с анализом спектра и статистических тестов на случайность и независимость остатков. Это позволяет подгонять модели по критерию хи-квадрат и проводить анализ приборных ошибок, разделяя вклады систематических и случайных ошибок. (Обычно доля случайной составляющей априори неизвестна, так как складывается из нескольких причин).

## **П.3 Пакет для поиска оболочечной модели частицы по данным рассеяния**

Основан на решении задачи наименьших квадратов со штрафными членами, отражающими требования неотрицательности функции формы и степень допустимой "негладкости" оболочки частицы.

Вклад автора: применение эффективного метода минимизации для плохо обусловленной задачи: алгоритма Левенберга-Маркардта, усиленного применением схемы переменной метрики с учетом простых ограничений на переменные и схемы вычисления конечно-разностных приращений из оценок ошибок вычисления функции.

#### **П.4 Пакет анализа данных малоуглового рассеяния от многокомпонентной структуры**

Программа MONSTER для поиска оболочечной модели многокомпонентной частицы по данным нейтронного рассеяния с вариацией контраста.

Вклад автора: метод коррекции экспериментальных данных на основе анализа числа компонентов в наборе кривых малоуглового рассеяния по левым сингулярным векторам матрицы данных. Данные корректируются путем поиска множителей и аддитивных констант для каждой экспериментальной кривой до достижения требуемого для теоретической модели числа базисных функций, определяющих весь экспериментальный набор данных.

#### **3.5 Пакет для моделирования состава полидисперсных смесей частиц по данным малоуглового рассеяния**

1) Две программы для поиска распределений частиц по размерам при известных формфакторах.

Программа MIXTURE на пакете OPTIS и предназначена для исследования поведения целевой функции в процессе поиска решения с целью получения решения, соответствующего глобальному минимуму. Программа POLYMIX имеет то же ядро целевой функции, но работает в автоматическом режиме по алгоритму Левенберга-Маркардта, усиленного применением схемы переменной

метрики с учетом простых ограничений на переменные и предназначена для рутинных вычислений.

### **3.6. Вспомогательные алгоритмы**

Оценка машинных констант

Оценка ошибок вычислений

Новый метод сглаживания зашумленных данных с двойной регуляризацией и коррекцией на основе статистического анализа остатков.

**Литература****Список цитируемой литературы**

1. Malinowski, E.R. Factor Analysis in Chemistry / E.R. Malinowski, D.G. Howery. – New York: R.E. Krieger Publishing Company, 1980. – 251 p. – ISBN–10 : 0–894–64343–6, ISBN–13 : 978–0–894–64343–9.
2. Ramos, L.S. Chemometrix / L.S. Ramos, K.R. Beebe, W.P. Carey, E. Sanchez, B. C. Erickson, B.E. Wilson, L.E. Wangen, B.R. Kowalski // Anal. Chem. –1986. –V. 58 No. 5. –P. 294R – 315R.
3. Delaney, M. Chemometrix / M. Delaney // Anal. Chem. – 1984. – V. 56, – No. 5. – P. 261R–277R.
4. Teicher H. Identificality of Mixtures of Product Measures / H. Teicher // The Annals of Mathematical Statistic. – 1967. – V. 38, – P. 1300 – 1302.
5. Fogarty, M.P. Multicomponent Photochemical Studies by Rapid Scanning Fluorescence Spectroscopy / Fogarty M.P., Warner I.M. // Appl. Spectr. –1980. – V. 34. – No. 4. – P.438–445.
6. Appellof, C.J. Tree–Dimensional Rank Annihilation for Multi–Component Determinations / C.J. Appellof, Davidson E.R. // Anal. Chim. Acta. – 1983. – V. 146. – P. 9–14.
7. Гречушников, Б.Н. Разложение сложного спектрального контура на составляющие методом преобразования Фурье / Б.Н. Гречушников, И.Н. Калинкина // ЖПС. – 1984. – Т. 41. – № 2. – С.249–251.
8. Gold, H.S. Principal Component and Decomposition Analysis of Multicomponent Mixtures of Carcinogenic Fluorofores / H.S. Gold, G.T. Rasmussen, J.A. Mercer–Smith, D.G. Whitten, R.P. Buck // Anal. Chim. Acta. – 1980. – V. 122. – No. 2. – P. 171–178.
9. Ильин, М.И. Метод анализа непрерывных перекрывающихся спектров / М.И.Ильин, С.В.Валгин, В.М.Соболев., Ю.В.Шариков // Завод. лаб. – 1987. – Т.53. – № 9. – С. 32 – 35.

10. Comon, P. Independent Component Analysis: a new concept? / P. Comon // *Signal Processing*. – 1994. – V. 36. – No. 3. – P. 287–314.
11. Hyvärinen, A. Independent Component Analysis / A. Hyvärinen, J. Karhunen, E. Oja. – New York: John Wiley & Sons, 2001. – 421 p. ISBN 978–0–471–40540–5.
12. Hyvärinen, A. Independent Component Analysis: Algorithms and Application / A. Hyvärinen, E. Oja. // *Neural Networks*. – 2000. – V. 13. – No. 4–5. – P. 411–430.
13. Stone, J.V. A Brief Introduction to Independent Component Analysis / J.V. Stone. // *Encyclopedia of Statistics in Behavioral Science* / S.D. Everitt, D.C. Howell (ed.). – Chichester: John Wiley & Sons, 2005. – V. 2. – P. 907–912. ISBN–10: 0–470–86080–4, ISBN–13: 978–0–470–86080–9.
14. Lee, T.–W. Independent component analysis: Theory and applications / T.–W. Lee. – Boston: Kluwer Academic Publishers, 1998. – 442 p. – ISBN: 0 7923 8261 7.
15. J.F. Cardoso, J.F. Blind signal separation – statistical principles / J.F. Cardoso // *Proc. of the IEEE*. – 1998. – V. 9. – No. 10. – P. 2009 – 2025.
16. Rodionova, O.Ye. Progress in Chemometrics Research / O.Ye. Rodionova, A.L. Pomerantsev // *Progress In Chemometrics Research* / A.L. Pomerantsev (ed.). New York: NovaScience Publishers, 2005. – 324 p. – P. 43–64. ISBN: 1–59454–257–0.
17. Mas, S. Chromatographic and spectroscopic data fusion analysis for interpretation of photodegradation processes / S. Mas S, R. Tauler, A. de Juan // *Journal of Chromatography A*. – 2011. – V. 1218. – No. 51. – P. 9260–9268.
18. Hurier, G. MILCA: A Maximum Internal Linear Component Analysis for the extraction of spectral emissions / G. Hurier, S.R. Hildebrandt, J.F. Macias–Perez. // *Astronomy & Astrophysics manuscript no. milca\_v5*. 2013. – V. 5. – P. 1–13. (USBN <http://arxiv.org/pdf/1007.1149.pdf>).
19. Stögbauer, H. Least dependent component analysis based on mutual information / H. Stögbauer, A. Kraskov, S. A. Astakhov, P. Grassberger // *Phys. Rev.* – E 70. – 2004. – P. 066123–066140.
20. Astakhov, S.A. Monte Carlo Algorithm for Least Dependent Non–Negative Mixture Decomposition / S. A. Astakhov, H. Stögbauer, A. Kraskov, P. Grassberger // *Anal. Chem.* – 2006. – V. 78. – P. 1620–1627.

21. Lawton, W.H. Self Modeling Curve Resolution / W.H. Lawton, E.A. Sylwestre // *Technometrix*. – 1971. – V. 13. – No. 3. – P.617–633.
22. Фок, М.В. Разделение сложных спектров на индивидуальные полосы при помощи обобщенного метода Аленцева / М.В. Фок // *Труды ФИАН*. – 1972. – Т. 59. – С. 3–24.
23. Антипова–Каратаева, И.И. О возможности предварительной оценки эффективности разложения сложных спектров смесей методом Аленцева–Фока / И.И. Антипова–Каратаева // *ЖПС*. – 1984. – Т.41, – № 3. – С. 461–466.
24. Антипова–Каратаева, И.И. Определение спектров индивидуальных органических соединений по спектрам смесей с помощью оптимизированного математического метода Аленцева–Фока / И.И. Антипова–Каратаева, Н.С. Софронова // *ЖАХ*. – 1985 – Т. 40, – № 6. – С. 1103–1108.
25. Ohta, N. Estimating Absorption Bands of Component Dyes by Means of Principal Component Analysis / N. Ohta // *Anal. Chem.* – 1973. – V. 45. – No. 3. – P. 553 – 557.
26. Borgen, O.S. An Extention of the Multivariate Component–resolution Method to Three Components / O.S. Borgen, B.R. Kowalski // *Anal. Chim. Acta.* – 1985. – V. 174. – P. 1–26.
27. Ohta, N. Estimating Absorption Bands of Component Dyes by Means of Principal Component Analysis / N. Ohta // *Anal. Chem.* –1973. – V. 45. – No. 3. – P. 553– 557.
28. Meister, A. Estimation of Component Spectra by the Principal Components Method / A. Meister // *Anal. Chim. Acta.* – 1984. –V. 161. – P. 149–161.
29. Martens, H. Factor Analysis of Chemical Mixtures / H. Martens // *Anal. Chim. Acta.* – 1979. – V. 112. – P. 423–442.
30. Туров, Ю.П. Применение факторного анализа к исследованию смесей / Ю.П. Туров / 5 Всесоюзная конференция по использованию вычислительных машин в спектроскопии молекул и химических исследованиях. – Новосибирск, 1980. – Сб. тезисов. – С. 182–183.

31. Туров, Ю.П. Разложение сложного спектрального контура на индивидуальные составляющие / Ю.П. Туров, О.Н. Вылегжанин / 5 Всесоюзная конференция по использованию вычислительных машин в спектроскопии молекул и химических исследованиях. – Новосибирск, 1980. – Сб. тезисов. – С.181.
32. Туров, Ю.П. Применение метода главных компонент при масс-спектрометрическом исследовании смесей органических соединений / Ю.П. Туров, О.Н. Вылегжанин / 5 Всесоюзная конференция по использованию вычислительных машин в спектроскопии молекул и химических исследованиях. – Новосибирск, 1980. – Сб. тезисов. – С. 184.
33. Pastrnak, J. To the Problem of Decomposition of Complex Optical Spectra into Individual Contributions / J. Pastrnak // *Czech. J. Phys.* – 1987. – В 37. – No. 8. – P. 933–941.
34. Kawata, S. Advanced Algorithm for Determining Component Spectra Based on Principal Component Analysis / S. Kawata, H. Komeda, K. Sasaki, S. Minami // *Appl. Spectr.* – 1985. – V. 39. – No. 4, – P. 610–614.
35. Friedrich, H.B. Combinations of Orthogonal Spectra to Estimate Component Spectra in Multicomponent Mixtures / H.B. Friedrich, J.-P. Yu // *Appl. Spectr.* – 1987. – V. 47. – No. 2. – P. 227–234.
36. Уиллиамс, У.Т. Методы иерархической классификации / У.Т. Уиллиамс, Дж.Н. Ланс // *Статистические методы для ЭВМ* / К. Энслейн, Э. Рэлстон, Г.С. Уилф (ред.), перевод с англ. М.Б. Малютова. – М. : Наука, – 1986. – 464 с. – С. 269–300.
37. Thijssen, P.C. Optimal Designs with Information Theory in Least-Squares Problems / P.C. Thijssen, G. Kateman, H.C. Smit // *Anal. Chim. Acta.* – 1984. – V. 157. – P. 99–115.
38. Kawata, S. Advanced Algorithm for Determining Component Spectra Based on Principal Component Analysis / S. Kawata, H. Komeda, K. Sasaki, S. Minami // *Appl. Spectr.* – 1985. – V. 39. – No. 4, – P. 610–614.
39. Evans, J.C. Automatic Analysis of Mixed Spectra / J.C. Evans, P.H. Morgan // *Anal. Chim. Acta.* – 1981. – V. 133. – P. 329–338.

40. Gillette, P.C. Computer-assisted Spectral Identification of Unknown Mixtures / P.C. Gillette, J.B. Lando, J.L. Coenig // *Appl. Spectr.* – 1982. – V. 36. – No. 6. – P. 661–665.
41. Rossi, T.M. Rank Estimation of Excitation–Emission Matrices Using Frequency Analysis of Eigenvectors / T.M. Rossi, I.M. Warner // *Anal. Chem.* – 1986. – V. 58. – No. 4. – P. 810–815.
42. Knorr, F.J. Separation of Mass Spectra of Mixtures by Factor Analysis / F.J. Knorr, J.H. Futrell // *Anal. Chem.* – 1979. – V. 51. – No. 8. – P. 1236–1241.
43. Malinowski, E.R. Factor Analysis for Isolation of the Raman Spectra of Aqueous Sulfuric Acid Components / E.R. Malinowski, R.A. Cox, U.L. Haldna // *Anal. Chem.* – 1984. – V. 56. – No. 4. – P. 778–781.
44. Lawton, W.H. Self Modeling Curve Resolution / W.H. Lawton, E.A. Sylwestre // *Technometrix.* – 1971. – V. 13. – No. 3. – P. 617–633.
45. Sharaf, M.A. Extraction of Individual Mass Spectra from Gas Chromatography–Mass Spectrometry Data of Unresolved Mixtures / M.A. Sharaf, B.R. Kowalski // *Anal. Chem.* – 1981. – V. 53. – No. 3. – P. 518–522.
46. Gilbert, R.A. Application of Factor Analysis to the Resolution of Overlapping XPS Spectra / R.A. Gilbert, J.A. Llewellyn, W.E., Jr. Swartz, J.W. Palmer. // *Appl. Spectr.* – 1982. – V. 36. – No. 4. – P. 428–430.
47. Chen, J.–H. Reconstruction of Mass Spectra of Components of Unknown Mixtures Based on Factor Analysis / J.–H. Chen, L.–P. Hwang // *Anal. Chim. Acta.* – 1981. – V. 133. – P. 271–281.
48. Malinowski, E.R. Obtaining the Key Set of Typical Vectors by Factor Analysis and Subsequent Isolation of Component Spectra / E.R. Malinowski // *Anal. Chim. Acta.* – 1982. – V. 134, – P. 129–137.
49. Беляев, Л.С. Решение сложных оптимизационных задач в условиях неопределенности / Л.С. Беляев / Новосибирск : Наука, 1978. – 128 с.
50. Эляшберг, М.Е. Молекулярный спектральный анализ и ЭВМ. / М.Е. Эляшберг, Л.А. Грибов, В.В. Серов / М. : Наука, 1980. – 308 с.

51. Иберла, К. Факторный анализ / К. Иберла / Перевод с нем. В. М. Ивановой. – М. : Статистика, 1980. –398 с.
52. Bulmer, J.T. Factor Analysis as a Component to Band Resolution Techniques. I. The Method and its Application to Self–Association of Acetic Acid / J.T. Bulmer, H.F. Shurvell // *The Journal of Physical Chemistry*. – 1973. – V. 77. – No. 2. – P. 256–262.
53. Davis, J.M. Statistical Method for Estimation of Number of Components from Single Complex Chromatograms: Theory, Computer–Based Testing, and Analysis of Errors / J.M. Davis, J.C. Giddings // *Anal. Chem.* – 1985. – V. 57. – No. 12. – P. 2169–2177.
54. Davis, J.M. Statistical Method for Estimation of Number of Components from Single Complex Chromatograms: Application to Experimental Chromatograms / J.M. Davis, J.C. Giddings // *Anal. Chem.* – 1985. – V. 57. – No. 12. – P. 2178–2182.
55. Лоули, Д. Факторный анализ как статистический метод. / Д. Лоули, А. Максвелл / Перевод с англ. – М. : Мир, 1967. – 145 с.
56. Калинкина, И.Н. Метод определения числа компонентов по спектрам смесей / И.Н. Калинкина, Б.Н. Гречушников // *ЖПС*. – 1985. – Т.42, – № 2. – С. 328–331.
57. Culler, S.R. Factor Analysis Applied to a Silane Coupling Agent on E–Glass Fiber System / S.R. Culler, P.C. Gillette, H. Ishida, J.L. Koenig // *Appl. Spectr.* – 1984. – V. 38. – No. 4. – P. 495–500.
58. Knorr, F.J. Resolution of Multicomponent Fluorescence Spectra by an Emission Wavelength – Decay Time Data Matrix / F.J. Knorr, J.M.Harris // *Anal. Chem.* – 1981. – V. 53. – No. 2. – P. 272–276.
59. Shrager, R.I. Titration of Individual Components in a Mixture with Resolution of Difference Spectra, pKs, and Redox Transitions / R.I. Shrager, R.W. Hendler // *Anal. Chem.* – 1982. – V. 54. – No. 7. – P.1147–1152.
60. Lawson, Ch.L. Solving Least Squares Problems / Ch.L. Lawson, R.J. Hanson / New Jersey. : Prentice–Hall, Inc., Englewood Cliffs, 1974. – 338 p. – ISBN: 0–89871–356–0.

61. Morgan, D.R. Spectral Absorption Pattern Detection and Estimation. I. Analytical Techniques / D.R. Morgan // *Appl. Spectr.* – 1977. – V. 31. – No. 5. – P. 404–415.
62. Иванов, И.А. Статистические алгоритмы цифровой обработки данных ИК–спектрометрического анализа. / И.А Иванов. / Препринт № 4 Института спектроскопии АН СССР, отделение общей физики и астрономии / г. Троицк Московской области, 1982. – 28 с.
63. Денисов, Г.С. К вопросу о спектроскопическом определении числа компонент в системе с равновесиями / Г.С. Денисов, Б.С. Терушкин // *Опт. и спектр.* – 1981. – Т.51, – № 6. – С. 1118–1120.
64. Hirschfeld, T. Computer Resolution of Infrared Spectra of Unknown Mixtures / T. Hirschfeld // *Anal. Chem.* – 1976. – V. 48. – No. 4. – P. 721–723.
65. Джонсон, Н. Статистика и планирование эксперимента в технике и науке. Методы обработки данных / Н. Джонсон, Ф. Лион, перевод с англ. – М. : Мир, 1980. – 612 с. Перевод изд.: Johnson, N.L. *Statistics and Experimental Design in Engineering and the Physical Sciences*. V. 1 / N.L. Johnson, F.C. Leone – 2–nd ed. – New York, John Wiley & Sons, 1977. – 496 p. ISBN: 0471444898 / 0–471–44489–8.
66. Иванова, В.М. Математическая статистика / В.М. Иванова, В.Н. Калинина, Л.А. Нешумова, И.О. Решетникова – М.: Высшая школа, 1981. – 371 с.
67. Durbin, J. Testing for Serial Correlation in Least–Squares regression III. / J. Durbin, G.S. Watson // *Biometrika.* – 1971. – V. 58. – P. 1–19.
68. Rasmussen, G.T. Principal Component Analysis of the Infrared Spectra of Mixtures / G.T. Rasmussen, T.L. Isenhour // *Anal. Chim. Acta.* – 1978. – V. 103. – P. 213–221.
69. Gillette, P.C. Computer–assisted Spectral Identification of Unknown Mixtures / P.C. Gillette, J.B. Lando, J.L. Coenig // *Appl. Spectr.* – 1982. – V. 36. – No. 6. – P. 661–665.
70. Malinowski, E.R. Theory of Error Applied to Pure Test Vectors in Target Factor Analysis / E.R. Malinowski // *Anal. Chim. Acta.* – 1981. – V. 133. – P. 99–101.

71. McCue, M. Target Factor Analysis of Infrared Spectra of Multicomponent Mixtures / M. McCue, E.R. Malinowski // *Anal. Chim. Acta.* – 1981. – V. 133. – P. 125–136.
72. McCue, M. Target Factor Analysis of the Ultraviolet Spectra of Unresolved Liquid Chromatographic Fractions / M. McCue, E.R. Malinowski // *Appl. Spectr.* – 1983. – V. 37. – No. 5. – P. 463–469.
73. Lorber, A. Validation of Hypotheses on a Data Matrix by Target Factor Analysis / A. Lorber // *Anal. Chem.* – 1984. – V. 56. – No. 6. – P. 1004–1010.
74. Елисеева, И.И. Эконометрика: Учебник / И.И. Елисеева, С.В. Курышева, Т.В. Костеева, И.В. Бабаева, Б.А. Михайлов, под общ. ред. И. И. Елисеевой. – М: Финансы и статистика, 2002. – 344 с.
75. Johnson, N.L. Statistics and Experimental Design in Engineering and the Physical Sciences. V. 2 (Probability & Mathematical Statistics) / N.L. Johnson, F.C. Leone – 2–nd ed. – New York, John Wiley & Sons, 1977. – 512 p. – ISBN–10: 0471017574; ISBN–13: 978–0471017578.
76. Хеттманспергер, Т. Статистические выводы, основанные на рангах / Т. Хеттманспергер. – М.: Финансы и статистика, 1987. – 334 с.
77. Рунион, Р. Справочник по непараметрической статистике / Р. Рунион. – М. : Финансы и статистика, 1982. – 198 с.
78. Larson, H. J. Statistics: an Introduction / H.J. Larson. – New York: John Wiley, 1975. – P. 328–329. – ISBN–10: 0898746396, ISBN–13: 978–0898746396.
79. Mann, H.B. On a test of whether one of two random variables is stochastically larger than the other / H.B. Mann, D.R. Whitney // *Annals of Mathematical Statistics.* – 1947. – No. 18. – P. 50–60.
80. Wilcoxon, F. Individual comparisons by ranking methods. / F. Wilcoxon // *Biometrics Bulletin.* – 1945. – V. 1. – P. 80–83.
81. Ланкастер, П. Теория матриц / П. Ланкастер. – М. : Наука, 1982. – 280 с.
82. Wald, A. On a test whether two samples are from the same population. / A. Wald, J. Wolfowitz // *The Annals of Mathematical Statistics.* – 1940. – V. 11. – P. 147–162.

83. Большев, Л.Н. Таблицы математической статистики. / Л.Н. Большев, Н.В. Смирнов. – М. : Наука, 1983. – 416 с.
84. Rasmussen, G.T. Identification of Components in Mixtures by a Mathematical Analysis of Mass Spectral Data / G.T. Rasmussen, B.A. Hohne, R.C. Wieboldt, T.L. Isenhour // *Anal. Chim. Acta.* – 1979. – V. 112. – P.151–164.
85. Hammel, M. Structural Flexibility of the N-terminal b-Barrel Domain of 15-Lipoxygenase-1 Probed by Small Angle X-ray Scattering. Functional Consequences for Activity Regulation and Membrane Binding / M. Hammel, M. Walther, R. Prassl, H. Kuhn // *J. Mol. Biol.* – 2004. – V. 343. – P. 917 – 929.
86. Fogarty, M.P. Ratio Method for Fluorescence Spectral Decomposition / M.P. Fogarty, I.M. Warner // *Anal. Chem.* – 1981. – V. 53. – No. 2. – P.259 – 265.
87. Применение масс-спектрометрии при исследовании состава смесей: дисс. канд. физ. мат. наук. / Туров Ю.П. – Кемерово, 1980. – 186 с.
88. Мартиросян, Э.Р. Программа спектрометрического анализа многокомпонентных смесей методом преобразования проектирования / Э.Р. Мартиросян, А.В. Давыдов, А.В. Кузнецов, Л.А. Грибов, Б.Н. Гречушников, В.В. Волков, И.Н. Калинкина // Математические методы и ЭВМ в аналитической химии. Сборник статей / Л. А. Грибов (ред.) – М. : Наука, 1989. – 302 с. – С. 39 – 40. – ISBN 5–02–001366–8.
89. Гилл, Ф. Практическая оптимизация / Ф. Гилл, У. Мюррей, М. Райт. – М.: Мир, 1985. – 512 с. – Перевод изд.: Gill, P.T. *Practical Optimization* / P.E. Gill, W. Murray, M.H. Wright. Elsevier: Academic Press, 1982. – 462 p. – ISBN–10: 0122839528, ISBN–13: 978–0122839528.
90. Дэннис, Дж., мл. Численные методы безусловной минимизации и решения нелинейных уравнений / Дж., мл. Дэннис, Р. Шнабель. – М.: Мир, 1988. – 440 с. ISBN 5–03–001102–1. – Перевод изд.: Dennis, J.E.Jr. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* / J.E.Jr. Dennis, R. B. Schnabel. New Jersey: Prentice Hall, 1983. – 428 p. – ISBN: 978–0–13–627216–8.

91. Моисеев, Н.Н. Методы оптимизации / Н.Н. Моисеев, Ю.П. Иванилов, Е.М. Столярова. – М.: Наука, 1978. – 352 с.
92. Поляк Б.Т. Введение в оптимизацию / Б.Т. Поляк. – М. : Наука, 1983. – 384 с.
93. Dennis, J. E. Jr. An Adaptive Nonlinear Least–Squares Algorithm / J.E.Jr. Dennis, D.M. Gay, R.E. Welsh // ACM Trans. on. Math. Soft. – 1981. – V. 7. – No. 3. – P. 348–368.
94. Dennis, J. E, Jr. Algorithm 573 NL2SOL – An Adaptive Nonlinear Least–Squares Algorithm [E4]. / J.E.Jr. Dennis, D.M. Gay, R.E. Welsh // ACM Trans. on. Math. Soft. – 1981. – V. 7. – No. 3. – P. 369–383.]
95. Ellis J., Sorge B. // J. Chem. Phys. 1934, V. 2. – P. 559. Cabannes J. // J. der Reo Riolo C.R. 1934, V. 198. – P. 30.
96. Степанов, Б.И. Колебательная энергия и люминесценция сложных молекул / Б.И. Степанов, Б.С. Непорент // УФН. – 1951. – Т. 43. – № 3. – С. 47 – 52.
97. Батуев, М.И. Исследование внутримолекулярной водородной связи гваякола методом комбинационного рассеяния света. 3 / М.И. Батуев. // Докл. АН СССР. – 1945. – Т. 47. – № 2. – С. 100 – 103.
98. Ландсберг, Г.С. Комбинационное рассеяние света в воде / Г.С. Ландсберг, С.А. Ухолин, Ф.С. Барышанская // Известия АН СССР. –1946. –№10. – С. 509 – 516.
99. Fox, J.J. Investigations of Infra–red Spectra (2.5–7.5 mkm). Absorption of water / J.J. Fox, A.E. Martin // Proc. Roy. Soc. London. –1940. – V. A174. – No. 959. – P. 234 – 261.
100. Adams, R.M. Vibrational spectra of water / R.M. Adams, I.I. Katy // J. Opt. Soc. America. – 1956, – V. 46. – P. 895 – 900.
101. Mutter, R. Optische Eigenschaften von Wasser / R. Mutter R., R. Mecke R., W. Luttko W. // Z. Phys. Chem. – 1959. – Bd. 19. – P. 83 – 89.
102. Pinkley, L.W. Optical Constants of Water in the Infrared: Influence of Temperature / Pinkley L.W., Sethna P.P., Williams D // J. Opt. Soc. America. – 1977. – V. 67. – P. 494 – 499.

103. Золотарев, В.М. Оптические постоянные природных и технических сред. Справочник / В.М. Золотарев, В.Н. Морозов, Е.В. Смирнова. – Л.: Химия, 1984. – 215 с.
104. Юхневич, Г.В. Инфракрасная спектроскопия воды / Г.В. Юхневич. – М.: Наука, 1973. – 208 с.
105. Okhulkov, A.V. X-ray-scattering in liquid water at pressures of up to 7.7 kbar – test of a fluctuation model / A.V. Okhulkov, Y.N. Demianets, Y.E. Gorbaty // J. Chem. Phys. – 1994. – V. 100. – P. 1578 – 1588.
106. Калинин, И.Н. Анализ спектров равновесных смесей / И.Н. Калинин, Б.Н. Гречушников // ЖПС. – 1979. – Т. 31. – № 1. – С. 168 – 170.
107. Попов, А.П. Использование метода совместной обработки оптических плотностей в спектрофотометрии комплексов переменного состава / А.П. Попов, Г.В. Фокина, О.Д. Кашпарова // ЖПС. – 1978. – Т.28, – № 4. – С. 714 – 718.
108. Sylvestre, E.A. Curve Resolution Using a Postulated Chemical Reaction / E.A. Sylvestre, W.H. Lawton, M.S. Maggio // Technometrix. – 1974. – V. 16. – No. 3. – P. 353 – 368.
109. Бажулина, Н.П. Анализ спектров поглощения равновесных химических смесей на ЭВМ / Н.П. Бажулина, А.Я. Ломакин, Ю.В. Морозов, Ф.А. Савин // ЖПС. – 1971. – Т. 14. – № 4. – С. 682 – 687.
110. Воеводин, В.В. Матрицы и вычисления. / В.В. Воеводин, Ю.А. Кузнецов – М.: Наука, 1984. – 320 с.
111. Стренг, Г. Линейная алгебра и ее приложения. / Г. Стренг. – М.: Мир, 1980. – 456 с.
112. Воеводин, В.В. Вычислительные основы линейной алгебры / В.В. Воеводин. – М.: Наука, 1977. – 304 с.
113. Воеводин, В.В. Линейная алгебра. / В.В. Воеводин. – М.:Наука, 1980. – 400 с.
114. Катковник, В.Я. Непараметрическая идентификация и сглаживание данных / В.Я. Катковник – М.: Наука, 1985. – 336 с.

115. Федоров, В.В. Анализ экспериментов при наличии ошибок в определении контролируемых переменных // В.В.Федоров. Межвуз. лаб. стат. методы, препринт № 2. – М.: изд. МГУ, 1968. – 15 с.
116. Тихонов, А.Н. Регуляризирующие алгоритмы и априорная информация / А.Н. Тихонов, А.В. Гончаровский, В.В. Степанов, А.Г. Ягола. – М.: Наука, 1983. – 200 с.
117. Морозов, В.А. Методы регуляризации неустойчивых задач / Морозов В.А. – М.: Изд-во МГУ, 1987. – 216 с.
118. Тихонов, А.Н. Методы решения некорректных задач / А.Н. Тихонов, В.Я. Арсенин. – М.: Наука, 1986. – 288 с.
119. Морозов, В.А. Регулярные методы решения некорректно поставленных задач / Морозов В.А. – М.: Наука, 1987. – 240 с.
120. Современная кристаллография: в 4 т. Т. 1. Симметрия кристаллов. Методы структурной кристаллографии / Б.К. Вайнштейн / Ред. Б.К. Вайнштейн, А.А. Чернов, Л.А. Шувалов. Т. 1: отв. ред. Л.А. Фейгин. – М., Наука, 1979. – 384 с.
121. Guinier, A. Small-Angle Scattering of X-rays / A. Guinier, G. Fournet. – New York: John Wiley & Sons, 1955. – 269 p.
122. Glatter, O. Small-Angle X-ray Scattering. / O. Glatter, O. Kratky. – London: Academic Press Inc, 1982. – 1982. – 515 p. – ISBN 0-12-286280-5.
123. Рольбин, Ю. А. Строение бактериофага Т7 по данным малоуглового рентгеновского рассеяния / Ю.А. Рольбин, Д.И. Свергун, Л.А. Фейгин, Ш. Гаспар, Д. Ронто // Докл. АН СССР. – 1980. – Т. 255. – С. 1497—1500.
124. Svergun, D. I. Small-angle scattering: direct structure analysis / D. I. Svergun, L. A. Feigin, B. M. Schedrin // Acta Crys. A. 1982. – V. A38. – P. 827 – 835.
125. Agirrezabala, X. Maturation of phage T7 involves structural modification of both shell and inner core components / X. Agirrezabala, J. Martin-Benito, J. R. Caston, R. Miranda, J. M. Valpuesta, J. L. Carrascosa // EMBO Journal. – 2005. – V. 24 – P. 3820 – 3829.

126. Свергун, Д. И. Рентгеновское и нейтронное малоугловое рассеяние / Д. И. Свергун, Л. А. Фейгин – М.: Наука. – 1986. – 280 с.
127. Guinier, A. La diffraction des rayons X aux trs petits angles: Application a l'tude de phnomnes ultramicroscopiques / A. Guinier // Ann. Phys. Paris 11 serie. – 1939. – V. 12. – P. 161 – 237.
128. Glatter, O. A new method for the evaluation of small-angle scattering data / O. Glatter // J. Appl. Cryst. – 1977. – V. 10. – No. 5. – P. 415 – 421.
129. Muller, J. J. Untersuchungen zur Struktur und Symmetrie der Pyruvatdecarboxylase aus Hefe mittels Röntgenkleinwinkelstreuung / J. J. Muller, G. Damaschun, G. Hubner // Acta Biol. Med. Germ. – 1979. – R. 38. – Nr. 1. – S. 1–10.
130. Рольбин, Ю. А. Расчет на ЭВМ интенсивности рентгеновского малоуглового рассеяния моделями произвольной формы с заданным распределением электронной плотности // Ю. А. Рольбин, Л. А. Фейгин, Б. М. Щедрин // Аппаратура и методы рентгеновского анализа. – 1971. – Т. 9. – С. 46 – 50.
131. Рольбин, Ю. А. К расчету на ЭВМ интенсивности рентгеновского малоуглового рассеяния моделями макромолекул / Ю. А. Рольбин, Р. Л. Каюшина, Л. А. Фейгин, Б. М. Щедрин // Кристаллография. – 1973. – Т. 18. – С. 701 – 705.
132. Pilz, I. Small-angle X-ray scattering / I. Pilz, O. Glatter, O. Kratky // Methods Enzymol. – 1980. – V. 61. – P. 148–249.
133. Pilz, I. Roetgenkleinwinklestudien uber die Substruktur von Helix pomatia Намосуанин / I. Pilz, O. Glatter, O. Kratky, I. Morning-Claesson // Z. Naturforsch. – 1972. – B. 27b. – S. 518 – 524.
134. Спири́н, А. С. Четвертичная структура рибосомной 30S субчастицы: модель и ее экспериментальная проверка / А. С. Спири́н, И. Н. Сердюк, И. Л. Шпунгин, В. Д. Васильев // Молек. биол. – 1979. – Т. 13. – С. 1384 – 1395.

135. Glatter, O. Computation of distance distribution functions and scattering functions of models for small-angle scattering experiments / O. Glatter // *Acta Phys. Austriaca*. – 1980. – B.52. – S. 243 – 256.
136. 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 / H. B. Stuhrmann // *Acta Cryst. Sect. A*. — 1970. — V. 26. — No. 3. — P. 297 — 306.
137. Stuhrmann, H. B. Ein neues Verfahren zur Bestimmung der Oberflächenform und der inneren Struktur von gelösten globulären Proteinen aus Röntgenkleinwinkelmessungen / H. B. Stuhrmann // *Z. Phys. Chem. (Frankfurt am Main)*. – 1970. – B. 72. – S. 177 – 184.
138. Свергун, Д. И. Прямой метод интерпретации данных малоуглового рассеяния системами идентичных частиц / Д. И. Свергун, Л. А. Фейгин, Б. М. Щедрин // *Доклады АН СССР*. – 1981. – Т. 261. – № 4. – С. 878 – 882.
139. Свергун, Д. И. К определению структуры биологических макромолекул методом малоуглового рассеяния / Д. И. Свергун, Л. А. Фейгин, Б. М. Щедрин // *Кристаллография*. – 1981. – Т. 26. – № 6. – С. 1163 – 1172.
140. Svergun, D. I. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing / D. I. Svergun // *Biophys. J.* – 1999. – V. 76. – P. 2879 – 2886.
141. Chacon, P. Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm / P. Chacón, F. Morán, J. F. Díaz, E. Pantos, J. M. Andreu // *Biophys. J.* – 1998. – V. 74. – No. 6. – P. 2760 – 2775.
142. Kirkpatrick, S. Optimization by simulated annealing / S. Kirkpatrick, C. D. Jr. Gelatt, M. P. Vecchi // *Science*. – 1983. – V. 220. – P. 671 – 680.
143. Mertens, H. D. Structural characterization of proteins and complexes using small-angle X-ray solution scattering / H.D. Mertens, D.I. Svergun // *J. Struct. Biol.* – 2010. – V. 172(1). – P. 128 – 141.

144. Сердюк, И. Методы в молекулярной биофизике. Структура. Функция. Динамика: в 2 т. / И. Сердюк, Н. Заккаи, Дж. Заккаи. – Изд. Книжный дом "Университет" (КДУ), 2009–2010. – 1304 с.
145. Levitt, M. Growth of Novel Protein Structural Data / M. Levitt // Proc. Natl. Acad. Sci. USA. – 2007. – V. 104. – P. 3183 – 3188.
146. Petoukhov, M. V. Global rigid body modeling of macromolecular complexes against small-angle scattering data / M. V. Petoukhov, D. I. Svergun // Biophys. J. – 2005. – V. 89(2). – P. 1237 – 1250.
147. Svergun, D. I. Mathematical methods in small-angle scattering data analysis. / D. I. Svergun // J. Appl. Crystallogr. – 1991. – V. 24. – P. 485 – 492.
148. Boutin, M. On Reconstructing n-Point Configurations from the Distribution of Distances or Areas / M. Boutin, G. Kemper // Mathematics, abstract math. AC/0304192 April 15, – 2003. – P. 1–21.
149. Котельников, В. А. О пропускной способности «эффира» и проволоки в электросвязи / В. А. Котельников // Успехи физических наук (репринт 1933 г.). – 2006. – Т. 176. – № 7. – С. 762 – 770.
150. Shannon, C. E. A Mathematical Theory of Communication / C. E. Shannon // Bell System Technical Journal. – 1948. – V. 27. – P. 379 – 423, P. 623 – 656.
151. Luzzati, V. Information content and retrieval in solution scattering studies. II. Evaluation of accuracy and resolution. / V. Luzzati, D. Taupin // J. Appl. Cryst. – 1986. – V. 19. – P. 39 – 50.
152. Luzzati, V. Information content and retrieval in solution scattering studies. I. Degrees of freedom and data reduction / D. Taupin, V. Luzzati // J. Appl. Cryst. – 1982. – V. 15. – P. 289 – 300.
153. Luzzati, V. Accuracy and resolution in small-angle crystallographic analyses. A comparison with solution scattering studies / V. Luzzati, D. Taupin // J. Appl. Cryst. – 1986. – V. 19. – P. 51 – 60.
154. Moore, P. B. Small-angle scattering. Information content and error analysis. / P. B. Moore // J. Appl. Cryst. – 1980. – V. 13. – P. 168 – 175.

155. Svergun, D. I. The solution of the one-dimensional sign problem for Fourier transforms / D.I. Svergun, L.A. Feigin, B.M. Scshedrin. // *Acta Cryst. A.* – V. 40 – No. 2. – P. 137 – 142.
156. Svergun, D. I. Direct method of interpretation of small-angle scattering by the solutions of biological macromolecules / D.I. Svergun, L.A. Feigin, B.M. Scshedrin // *Studia Biophysica.* – 1982. – V. 87. – No. 3. – P. 277 – 278.
157. Фейгин, Л. А. Восстановление плотности сферически симметричной частицы по её самосвёртке / Л.А. Фейгин, П.В. Шмидт, Б.М. Щедрин // *Кристаллография.* – 1981. – Т. 26. – № 5. – С. 920 – 924.
158. Brent, R. P. Algorithms for Minimization Without Derivatives. / R. P. Brent. – New York: Englewood Cliffs, 1973. – Chap.5. 131 p.
159. Hamming, R. W. Numerical Methods for Scientists and Engineers (Dover Books on Mathematics) // R. W. Hamming. – New York: Dover Pubns, 1987. – 752 p. – ISBN-10: 0-486-65241-6, ISBN-13: 978-0486652412.
160. Svergun, D. I. New developments in direct shape determination from small-angle scattering. 1. Theory and model calculations / D.I. Svergun, H.B. Stuhrmann // *Acta Cryst.* – 1991. – A47. – P. 736 – 744.
161. Doyle, D. A. The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity / D.A. Doyle, J.M. Cabral, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Chait, R. MacKinnon // *Science.* – 1998. – V. 280. – P. 69 – 77.
162. Wittmann, H. G. Components of bacterial ribosomes / H. G. Wittmann // *Annu. Rev. Biochem.* – 1982. – V. 51. – P. 155 – 183.
163. Frank, J. A model of protein synthesis based on a new cryo-electron microscopy reconstruction of the E. coli ribosome / J. Frank, J. Zhu, P. Penczek, Y. Li, S. Srivastava, A. Verschoor, M. Radermacher, R. Grassucci, R.K. Lata, R.K. Agrawal // *Nature.* – 1995. – V. 376. – P. 441– 444.
164. Brimacombe, R. The structure of ribosomal RNA: a three-dimensional jigsaw puzzle / R. Brimacombe // *Eur. J. Biochem.* – 1995. – V. 230. – P. 365 – 383.

165. Ibel, K. Stuhrmann H.B. Comparison of neutron and X-ray scattering of dilute myoglobin solutions / K. Ibel, H. B. Stuhrmann // *J. Mol. Biol.* – 1975. – V. 93. – P. 255 – 265.
166. Svergun, D.I. Restoring three-dimensional structure of biopolymers from solution scattering / D.I. Svergun, // *J. Appl. Cryst.*, 1997. V. 30. – P. 792 – 797.
167. Svergun, D.I. New developments in direct shape determination from small-angle scattering. 1. Theory and model calculations / D. I. Svergun, H. B. Stuhrmann // *Acta Cryst.* – 1991. – A47. – P. 736 – 744.
168. Svergun, D.I. Solution scattering from biopolymers: advanced contrast variation data analysis / D.I. Svergun // *Acta Cryst.* – 1994. – A50. – P. 391 – 402.
169. Semenyuk, A. V. GNOM: a Program Package for Small-Angle Scattering Data Processing / A.V. Semenyuk, D.I. Svergun // *J. Appl. Cryst.* – 1991. – V. 24. – P. 537 – 540.
170. Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria / D.I. Svergun // *J. Appl. Cryst.* – 1992. – V. 25. – P. 495 – 503.
171. Kozin, M.B. Automated matching of high- and low-resolution structural models / M.B. Kozin, D.I. Svergun // *J. Appl. Cryst.* – 2001. – V. 34. – P. 33 – 41.
172. Гогия, З. В. Структура рибосом *Thermus thermophilus*. 1. Метод выделения и очистки рибосом / З.В. Гогия, М.М. Юсупов, Т.Н. Спирина. // *Молек. биол.* – 1986. – Т. 20. – С. 519 – 526.
173. *Structure and Function of Antibodies* / Ed. by L.E. Glynn, M.W. Steward. – New York: John Wiley and Sons, 1980. – 172 p. – ISBN-10: 047127917X, ISBN-13: 978-0471279174
174. Harris, L.J. The three-dimensional structure of an intact monoclonal antibody for canine lymphoma / L.J. Harris, S.B. Larson, K.W. Hasel, J. Day, A. Greenwood, A. McPherson // *Nature.* – 1992. – V. 360. – P. 369 – 372.
175. Harris, L. Crystallographic structure of an intact IgG1 monoclonal antibody / L.J. Harris, E. Skaletsky, A. McPherson // *J. Mol. Biol.* – 1998. – V. 275. — P. 861 – 872.

176. Harris, L. J. Comparison of the conformations of 2 intact monoclonal–antibodies with hinges / L.J. Harris, S.B.Larson, E. Skaletsky, A. McPherson // Immunological reviews. – 1998. – V. 163. – P. 35 – 43.
177. Harris, L. J. Comparison of intact antibody structures and the implications for effectorfunction / L.J. Harris, S.B. Larson, A. McPherson // Adv. Immunology. – 1998. – V. 72. – P. 191 – 200.
178. Perkins, S. J. Solution Structure of Human and Mouse Immunoglobulin M by Synchrotron X–ray Scattering and Molecular Graphics Modelling. A Possible Mechanism for Complement Activation / S.J. Perkins, A.S. Nealis, B.J. Sutton, A. Feinstein // J. Mol. Biol. – 1991. – V. 221. – P. 1345 – 1366.
179. Ninio, J. Comparative small–angle x–ray scattering studies on unacylated, acylated and cross–linked Escherichia coli transfer RNA I Val / J. Ninio, V. Luzatti, M. Yaniv // J. Mol. Biol. – 1972. – V. 71. – P. 217 – 229.
180. Grossmann, G. X–ray–scattering using synchrotron–radiation shows nitrite reductase from Achromobacter–xylooxidans to be a trimer in solution / G. Grossmann, Z.H.L. Abraham, E.T. Adman, M. Neu; R.R. Eady; B.E. Smith, S.S. Hasnain // Biochemistry. – 1993. – V. 32. – P. 7360 – 7366.
181. Aleshin, A.E. Nonaggregating mutant of recombinant human hexokinase I exhibits wild–type kinetics and rod–like conformations in solution / A.E. Aleshin, M. Malfois, X. Liu, C.S. Kim, H.J. Fromm, R.B. Honzatko, M.H. Koch, D.I. Svergun // Biochemistry. – 1999. – V. 38. – P. 8359 – 8366.
182. Лапук, В.А. Некоторые особенности фрагментации моноклонального ревматоидного иммуноглобулина М "горячим" трипсином / В.А. Лапук, В.Я. Черняк, Н.Н. Магретова // Биохимия. – 1996. – Т. 61, – С. 85 – 88.
183. Тимофеев, В. П. Использование динамического метода спин–метки для изучения комплекса барстар–барназа / В.П. Тимофеев, Т.Г. Баландин, Я.В. Ткачев, В.В. Новиков, В.А. Лапук, С.М. Деев // Биохимия. – 2007. – Т. 72. – № 9. – С. 1220 – 1230.
184. Тимофеев, В. П. Исследование необратимого конформационного перехода в иммуноглобулине М методом спин–метки, введенной в углеводную и

- пептидную части его молекулы / В.П. Тимофеев, В.А. Лапук // Молек. биол. – 1982. – Т. 16. – С. 403 – 410.
185. Lapuk, V.A. Some Peculiarities of the Dynamics of the Immunoglobulin M Structure / V.A. Lapuk, V.P. Timofeev, A.I. Tchukchrova, N.M. Khatiashvili, T.M. Kiseleva // J. Biomol. Struct. Dynamics. – 1984, – V. 2. – P. 63–76.
186. Sutton, B. The structure and origin of rheumatoid factors / B. Sutton, A. Corper, V. Bonagura, M. Taussig // Immunology Today. – 2000. – V. 21. – P. 177 – 182.
187. Timofeev, V. P. Dynamics of macromolecule spin-labeled side chain groups by electron spin resonance spectra simulation / V.P. Timofeev, B.A. Samarianov // J. Chem. Soc. (Perkin Trans). – 1995, – V. 2. – P. 2175 – 2181.
188. Лапук, В. А. Простой метод выделения моноклонального иммуноглобулина М, обладающего ревматоидной активностью / В.А. Лапук, А.И. Чухрова, Е.В. Чернохвостова, В.А. Алешкин, Г.П. Герман, Е.Ю. Варламова, А.М. Пономарева, Н.П. Арбатский, А.О. Желтова // Биохимия. – 1992. – Т. 57. – № 4. – С. 617 – 626.
189. Timofeev, V. P. The role of the fast motion of the spin label in the interpretation of EPR spectra for spin-labeled macromolecules / V.P. Timofeev, D.O. Nikolsky // J. Biomol. Struct. Dyn. – 2003. – Vol. 21. – P. 367 – 378.
190. Ройт, А. Иммунология. / А. Ройт, Дж. Бростофф, Д. М. Мейл. – М.: Мир, 2000. – 592с. – ISBN 5–03–003305–X.
191. Wilhelm, P. Quaternary structure of immunoglobulin–M – model based on small–angle X–ray–scattering data / Wilhelm P., Pilz I., Goral K., Palm W. // Int. J. Biol. Makromol. – 1980. – V. 2. – P. 13 – 16.
192. Harris, L .J. Refined structure of an intact IgG2A monoclonal antibody / L.J. Harris, S.B. Larson, K.W. Hasel, A. McPherson. // Biochem. – 1997. V. 36. – P. 1581 – 1588.
193. Harris, L. J. Crystallographic structure of an intact IgG1 monoclonal antibody / L.J. Harris, E. Skaletsky, A. McPherson // J. Mol. Biol. 1998. V. 275 – P. 861–869.

194. Svergun, D. CRY SOL – a Program to Evaluate X–ray Solution Scattering of Biological Macromolecules from Atomic Coordinates / D. Svergun, C. Barberato, M.H.J. Koch // *J. Appl. Cryst.* – 1995. – V. 28. – P. 768 – 773.
195. Протасевич, И. И. Сравнительное исследование моноклонального иммуноглобулина М и ревматоидного иммуноглобулина М методом дифференциальной сканирующей микрокалориметрии // И.И. Протасевич, В. Ранжбар, Е.Ю. Варламова, И.А. Черкасов, В.А. Лапук // *Биохимия.* – 1997. – Т. 62. – С. 914 – 918.
196. Konarev, P. V. MASSHA a graphics system for rigid–body modelling of macromolecular complexes against solution scattering data / P.V. Konarev, M.V. Petoukhov, D.I. Sverguna // *J. Appl. Cryst.* – 2001. – V. 34. – P. 527 – 532.
197. Livshitz, M. A. Flexibility difference between double–stranded RNA and DNA as revealed by gel electrophoresis / M.A. Livshitz, O.A. Amosova, Yu.L. Lyubchenko // *J. Biomol. Struct. Dyn.* – 1990. – V. 7. – P. 1237 – 1249.
198. Финкельштейн, А. В. Физика белка / Финкельштейн А.В., Птицын О.Б. – М.: Книжный дом, "Университет", 2002. – 131 с.
199. Файнштейн, А. Структура и функции антител / А. Файнштейн, Д. Бил / ред. Л. Глинн, М. Стьюард. – М.: Мир, 1983. – 200 с. – С. 158,169.
200. Lapuk, V.A. Accessibility of tryptophan residues in immunoglobulin M molecule as an indicator of its conformational variability / Lapuk V.A., Chukhrova A.I., Khatiashvili N.M., Shmakova F.V., Kaverzneva E.D., Timofeev V.P. // *Biokhimiia.* – 1989. – V. 54. – No. 12. – P. 1956 – 1964.
201. Protein Data Bank. Human Igm Rheumatoid Factor Fab In Complex With Its Autoantigen IgG Fc. 1ADQ.
202. Percus, J. K. Analysis of classical statistical mechanics by means of collective coordinates / J.K. Percus, G.J. Yevick // *Phys. Rev.* – 1958. – V. 110. – No. 1. – P. 1 – 13.
203. Vrij, A. Mixtures of hard spheres in the Percus–Yevick approximation. Light scattering at finite angles / A. Vrij // *J. Chem. Phys.* – 1979. – V. 71. – No. 8. – P. 3267 – 3270.

204. Schultz, G. V. Z. Uber die Kinetik der Kettenpolymerisationen. V). Der Einfluss verschiedener des Reaktionsprodukt bei Makropolymerisationsvorgangen / G. V. Z. Schultz // Zeit. fur Phys. Chemie (B), Leipzig. – 1939. – B. 43. – P. 25 – 46,
205. Zimm, B. H. The Scattering of Light and the Radial Distribution Function of High Polymer Solutions / B. H. Zimm // J. Phys. Chem. – 1948. – V. 16. – No. 12. – P. 1099 – 1116.
206. Bailey, D. H. A Comparison of Three High–Precision Quadrature Schemes / D. H. Bailey, K. Jeyabalan, X. S. Li. // Experimental Mathematics. – 2005. – V. 14. – No. 3. – P. 317 – 329.
207. Ахиезер, Н. И. Лекции по теории аппроксимации / Н. И. Ахиезер – М.: Наука, 1965. – 384 с.

## Список работ автора

## Публикации в реферируемых журналах

- A1. Гречушников, Б.Н. Об однозначности решения задачи разделения аддитивных спектров смесей на спектры компонентов / Б. Н. Гречушников, И. Н. Калинкина, В. В. Волков, И. П. Петров // Журнал прикладной спектроскопии. – 1984. – Т. 40. – Вып. 1. – С. 89 – 92.
- A2. Волков, В. В. Применение преобразования проектирования в методе варьирования концентраций компонентов при анализе многокомпонентных смесей по спектрам поглощения / В. В. Волков, Б. Н. Гречушников // Журнал прикладной спектроскопии. – 1984. – Т. 40. – Вып. 2. – С. 264 – 267.
- A3. Волков, В. В. К вопросу оценивания спектров компонентов по спектрам родственных смесей / В. В. Волков // Журнал прикладной спектроскопии. – 1988. – Т. 48. – Вып. 3. – С. 436 – 442.
- A4. Волков, В. В. Коррекция решения задачи численного разложения спектров / В. В. Волков, Б. Н. Гречушников, И. Н. Калинкина // Журнал прикладной спектроскопии. – 1988. – Т. 49. – Вып. 2. – С. 320 – 323.
- A5. Волков, В. В. Оценивание спектров компонентов по набору аддитивных спектров смесей / В. В. Волков // Математические методы и ЭВМ в аналитической химии: сборник / Л. А. Грибов (ред.) – М. : Наука, 1989. – 302 с. – С. 42–50. – ISBN 5–02–001366–8.
- A6. Шевчук, Т. С. Изучение комплексообразования в системе  $\text{NiCl}_2\text{--CO}(\text{NH}_2)_2\text{--H}_2\text{O}$  методом электронной спектроскопии путем разложения спектральных контуров на аддитивные составляющие / Т. С. Шевчук, А. Ф. Борина, В. В. Волков // ЖНХ. – 1992. – Т. 37. – Вып. 6. – С. 1355–1361.
- A7. Schmidt, B. Small-angle X-ray solution scattering study on the dimerization of the FKBP25mem from *Legionella pneumophila* / B. Schmidt, S. Koenig, D. Svergun, V. Volkov, G. Fisher, M. H. J. Koch // FEBS Letters. – 1995. – V. 372. – P. 169 – 172.
- A8. Volkov, V. V. Separation of Additive Mixture Spectra by a Self-Modeling

- Method / V. V. Volkov // *Applied Spectroscopy*. – 1996. – V. 50. – No. 3. – P. 320 – 326.
- A9. Svergun, D. I. New Developments in Direct Shape Determination from Small-Angle Scattering. 2. Uniqueness / D. I. Svergun, V. V. Volkov, M. B. Kozin, H. B. Stuhmann // *Acta Cryst. A*. – 1996. – V. 52. – P. 419 – 426.
- A10. Юхневич, Г. В. Полоса валентных колебаний и структура жидкой воды / Г. В. Юхневич, В. В. Волков // *Доклады АН*. – 1997. – Т. 353. – № 4. – С. 465 – 468.
- A11. Svergun, D. I. Solution Scattering Structural Analysis of the 70S *Escherichia coli* Ribosome by Contrast Variation: I. Invariants and Validation of Electron Microscopy Models / D. I. Svergun, N. Burkhardt, J. S. Pedersen, M. H. J. Koch, V. V. Volkov, M. B. Kozin, W. Meerwink, H. B. Stuhmann, G. Diedrich, K. H. Nierhaus // *J. Mol. Biol.* – 1997. – V. 271. – P. 588 – 601.
- A12. Svergun, D. I. Solution Scattering Structural Analysis of the 70S *Escherichia coli* Ribosome by Contrast Variation: II. Structure modelling / D. I. Svergun, N. Burkhardt, J. S. Pedersen, M. H. J. Koch, V. V. Volkov, M. B. Kozin, W. Meerwink, H. B. Stuhmann, G. Diedrich, K. H. Nierhaus // *J. Mol. Biol.* – 1997. – V. 271. – P. 602 – 618.
- A13. Burkhardt, N. Architecture of the 70S *Escherichia coli* Ribosome / N. Burkhardt, G. Diedrich, K. H. Nierhaus, W. Meerwink, H. B. Stuhmann, Skov J. Pedersen, M. H. J. Koch, V. V. Volkov, M. B. Kozin, D. I. Svergun // *Physica B*. – 1997. – V. 234–236. – P. 199 – 201.
- A14. Svergun, D. I. Shape Determination from Solution Scattering of Biopolymers / D. I. Svergun, V. V. Volkov, M. B. Kozin, H. B. Stuhmann, C. Barberato, M. H. J. Koch // *J. Appl. Cryst.* – 1997. – V. 30. – P. 798 – 802.
- A15. Kozin, M. B. ASSA – a Program for Three-Dimensional Rendering in Solution Scattering from Biopolymers / M. B. Kozin, V. V. Volkov, D. I. Svergun // *J. Appl. Cryst.* – 1997. – V. 30. – P. 811 – 815.
- A16. Koenig, S. Small-Angle X-ray Solution-Scattering Studies on Ligand-Induced Subunit Interactions of the Thiamine Diphosphate Dependent Enzyme Pyruvate

- Decarboxylase from Different Organisms / S. Koenig, D. I. Svergun, V. V. Volkov, L. A. Feigin, M. H. J. Koch // *Biochemistry*. – 1998. – V. 37. – P. 5329 – 5334.
- A17. Svergun, D. I. Contrast Variation X-ray and Neutron Structural Studies of the 70S Ribosome from *E. coli* and *Thermus Thermophilus* / D. I. Svergun, V. V. Volkov, J. Skov Pedersen, M. H. J. Koch, M. B. Kozin, H. B. Stuhmann, I. N. Serdyuk, L. Fan, V. L. Lamzin, K. H. Nierhaus // *Structure Research of the Ribosome and its Functional Complexes* / Ed. by R. Willumeit and S. Forthmann. GKSS: Geesthacht. – 1998. – P. 19 – 23.
- A18. Bilgin, N. Solution Structure of the Ternary Complex between Aminoacyl-tRNA, Elongation Factor Tu, and Guanosine Triphosphate / N. Bilgin, M. Ehrenberg, C. Ebel, G. Zaccai, Z. Sayers, M. H. J. Koch, D. I. Svergun, C. Barberato, V. Volkov, P. Nissen, J. Nyborg // *Biochemistry*. – 1998. – V. 37. – No. 22. – P. 8163 – 8172.
- A19. Shilton, B. *Escherichia coli* SecA shape and dimensions / B. Shilton, D. I. Svergun, V. V. Volkov, M. H. J. Koch, S. Cusack, A. Economou // *FEBS Lett.* – 1998. – V. 436. – P. 277 – 282.
- A20. Macheroux P. Evidence for a major structural change in *Escherichia coli* chorismate synthase induced by flavin and substrate binding / P. Macheroux, E. Schoenbrunn, D. I. Svergun, V. V. Volkov, M. H. J. Koch, S. Bornemann, R. N. F. Thorneley // *Biochem. J.* – 1998. – V.335. – P. 319 – 327.
- A21. Svergun, D. I. Quaternary structure of V1 and F1 ATPase: significance of structural homologies and diversities / D. I. Svergun, S. Konrad, M. Huss, M. H. J. Koch, H. Wiczorek, K.-H. Altendorf, V. V. Volkov, G. Grueber // *Biochemistry*. – 1998. – V. 37. – P. 17659 – 17663.
- A22. Свергун, Д. И. Автоматическое определение формы частицы в малоугловом рассеянии монодисперсными системами./ Д. И. Свергун, В. В. Волков, М. Б. Козин // *Поверхность. Рентгеновские, синхротронные и нейтронные исследования*. – 1999. – № 9. – С. 3 – 7.
- A23. Svergun, D. I. A small angle x-ray scattering study of the droplet-cylinder transition in oil-rich sodium bis(2-ethylhexyl) sulfosuccinate microemulsions /

- D. I. Svergun, P. V. Konarev, V. V. Volkov, M. H. J. Koch, W. F. C. Sager, J. Smeets, E. M. Blokhuis // *J. Chem. Phys.* – 2000. – V. 113. – P. 1651 – 1665.
- A24. Svergun, D. I. Small–Angle X–Ray Scattering Study of Platinum–Containing Hydrogel/Surfactant Complexes / D. I. Svergun, E. V. Shtykova, M. B. Kozin, V. V. Volkov, A. T. Dembo, E. V. Jr. Shtykova, L. M. Bronstein, O. A. Platonova, A. N. Yakunin, P. M. Valetsky, A. R. Khokhlov // *J. Phys. Chem. B.* – 2000. – V. 104. – No. 22. – P. 5242 – 5250.
- A25. Svergun, D. I. Formation of metal Nanoparticles in Multilayered Poly(Octadecylsiloxane) as Revealed by Anomalous Small–Angle X–Ray Scattering / D. I. Svergun, M. B. Kozin, P. V. Konarev, E. V. Shtykova, V. V. Volkov, D. M. Chernyshov, P. M. Valetsky, L. M. Bronstein // *Chemistry of Materials.* – 2000. – V. 12. – No. 12. – P. 3552 – 3560.
- A26. Fan, L. Structural studies of the 30S subunit of ribosomes *Thermus thermophilus* by small–angle neutron and X–ray scattering / L. Fan, D. I. Svergun, V. V. Volkov, V. L. Aksenov, C. Ch. Agalarov, O. M. Selivanova, I. N. Shcherbakova, M. H. J. Koch, R. Gilles, A. Wiedenmann, R. May, I. N. Serdyuk // *J. Appl. Crystallogr.* – 2000. – V. 33. – P. 515 – 518.
- A27. Volkov, V. V. Low–resolution structure of immunoglobulins IgG1, IgM and rheumatoid factor IgM–RF from solution X–ray scattering data / V. V. Volkov, V. A. Lapuk, R. L. Kayushina, E. V. Shtykova, E. Yu. Varlamova, M. Malfois, D. I. Svergun. // *J. Appl. Cryst.* – 2003. – V. – 36. – P. 503 – 508.
- A28. Volkov, V. V. Uniqueness of *ab initio* shape determination in small–angle scattering / V. V. Volkov, D. I. Svergun // *J. Appl. Cryst.* – 2003. – V. 36. – P. 860 – 864.
- A29. Shtykova, E. V. Small–angle X–ray scattering reveals hollow nanostructures in iota– and kappa–carrageenan/surfactant complexes / E. V. Shtykova, E. V. Jr. Shtykova, V. V. Volkov, P. V. Konarev, A. T. Dembo, E. E. Makhaeva, I. A. Ronova, A. R. Khokhlov, H. Reynaers, D. I. Svergun // *J. Appl. Cryst.* – 2003. – V. 36. – P. 669 – 673.
- A30. Sokolova, A. V. Prototype of a database for rapid protein classification based on

- solution scattering data / A. V. Sokolova, V. V. Volkov, D. I. Svergun // *J. Appl. Cryst.* – 2003. – V. 36. – P. 865 – 868.
- A31. Волков, В. В. Определение строения иммуноглобулинов человека IgG, IgM и ревматоидного фактора IgM–RF в растворе / В. В. Волков, Р. Л. Каюшина, В. А. Лапук, Э. В. Штыкова, Е. Ю. Варламова, М. Мальфуа, Д. И. Свергун // *Кристаллография.* – 2003. – Т. 48. – № 1. – С. 103 – 110.
- A32. Starodoubtsev, S. G. Poly(acrylamide) gels with embedded magnetite nanoparticles / S. G. Starodoubtsev, E. V. Saenko, A. R. Khokhlov, V. V. Volkov, K. A. Dembo, V. V. Klechkovskaya, E. V. Shtykova, I. S. Znaveskina // *Microelectronic Engineering.* – 2003. – V. 69. – P. 324 – 329.
- A33. Dembo, A. T. Structure of Binary and Ternary Complexes Formed by Sodium Poly(2–acrylamide–2–methyl–1–propanesulfonate) Gel in the Presence of Copper(II) Nitrate and Cetylpyridinium Chloride / A. T. Dembo, K. A. Dembo, V. V. Volkov, A. I. Kokorin, A. A. Lyubimov, E. V. Shtykova, S. G. Starodoubtsev, A. R. Khokhlov // *Langmuir.* – 2003. – V. 19. – P. 7845 – 7851.
- A34. Konarev, P. V. PRIMUS: a Windows PC–based system for small–angle scattering data analysis / P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, D. I. Svergun. // *J Appl. Crystallogr.* – 2003. – V. 36: – P. 1277 – 1282.
- A35. Кононенко, А. В. Молекулярная морфология фактора терминации трансляции 1–го класса эукариот eRF1 в растворе / А. В. Кононенко, К. А. Дембо, Л. Л. Киселев, В. В. Волков // *Молекуляр. биология.* – 2004. – Т. 38, – С. 303 – 311.
- A36. Ozerin, A. N. The spatial structure of dendritic macromolecules / A. N. Ozerin, D. I. Svergun, V. V. Volkov, A. I. Kuklin, V. I. Gordeliy, A. Kh. Islamov, L. A. Ozerina, D. S. Zavoroyniuk // *J. Appl. Crystallogr.* – 2005. – V. 38, – P. 996 – 1003.
- A37. Юрков, Г. Ю. Железосодержащие наночастицы в матрице силоксановых каучуков / Г. Ю. Юрков, Д. А. Астафьев, Л. Н. Никитин, Ю. А. Кокшаров, Н. А. Катаева, Э. В. Штыкова, К. А. Дембо, В. В. Волков, А. Р. Хохлов,

- С. П. Губин // Неорганические материалы. – 2006. – Т. 42, – № 5, – С. 556 – 562.
- A38. Кочервинский, В. В. Роль внутрицепных дипольных взаимодействий при формировании надмолекулярной структуры кристаллизующихся сегнетоэлектрических полимеров / В. В. Кочервинский, В. В. Волков, К. А. Дембо // Физика твердого тела. – 2006. – Т. 48. – Вып. 6. – С. 1019 – 1021.
- A39. Konarev, P. V. ATSAS 2.1, a program package for small-angle scattering data analysis / P. V. Konarev, M. V. Petoukhov, V. V. Volkov, D. I. Svergun // J. Appl. Crystallogr. – 2006. – V. 39. – P. 277 – 286.
- A40. Shtykova, E. V. Autoxidized Phospholipids in Hexane: New Types of Nanoparticles Studied by Synchrotron Small-Angle X-ray Scattering / E. V. Shtykova, V. V. Volkov, H. Wang, T. Fujisawa, Jin-Ye Wang. // Langmuir. – 2006. – V. 22. – No. 19. – P. 7994 – 8000.
- A41. Никитин, Л. Н. Структура композитов, полученных формированием полипиррола в сверхкритическом CO<sub>2</sub> на микропористом полиэтилене / Л. Н. Никитин, М. О. Галлямов, А. Ю. Николаев, Э. Е. Саид-Галиев, А. Р. Хохлов, С. С. Букалов, Г. И. Магдануров, В. В. Волков, Э. В. Штыкова, К. А. Дембо, Г. К. Ельяшевич // Высокомолек. соед. А. – 2006. – Т. 48. – № 8. – С. 1431 – 1447.
- A42. Klechkovskaya, V. V. Network model of Acetobacter Xylinum Cellulose Intercalated by Drug Nanoparticles / V. V. Klechkovskaya, V. V. Volkov, E. V. Shtykova, N. A. Arkharova, Yu. G. Baklagina, A. K. Khripunov, R. Yu. Smyslov, L. N. Borovikova, A. A. Tkachenko // Nanomaterials for Application in Medicine and Biology / Eds. By M. Giersig and G. B. Khomutov. NATO Science for Peace and Security Series B: Physics and Biophysics. Springer Netherlands. – 2008. – P. 165 – 177.
- A43. Yevdokimov, Yu. M. Transition in DNA molecule's spatial ordering due to nano-scale structural changes / Yu. M. Yevdokimov, V. I. Salyanov, E. V. Shtykova, K. A. Dembo, V. V. Volkov, P. V. Spirin, A. S. Slusheva, V. S. Prassolov // The Open Nanoscience Journal. – 2008. – V. 2. – P. 17 – 28.

- A44. Волков, В. В. Особенности структуры Fab-фрагментов ревматоидного фактора IgM-RF в растворе / В. В. Волков, В. А. Лапук, Э. В. Штыкова, Н. Д. Степина, К. А. Дембо, А. В. Соколова, С. В. Амарантов, В. П. Тимофеев, Р. Х. Зиганшин, Е. Ю. Варламова // Кристаллография. – 2008. – Т. 53. – № 3. – С. 476 – 484.
- A45. Aslanov, L. Mesoporous soot-supported palladium as a heterogeneous catalyst for the Heck reaction in ionic liquids / L. A. Aslanov, Y. A. Kabachii, S. Yu. Kochev, B. V. Romanovsky, P. M. Valetsky, V. V. Volkov, A. V. Yatsenko, V. N. Zakharov // Mendeleev Communications. – 2008. – V. 18. – No. 6. – P. 334 – 335.
- A46. Pastukhov, A. V. Magnetic Nanocomposites Based on Hypercrosslinked Polystyrenes / A. V. Pastukhov, V. A. Davankov, V. V. Volkov, K. A. Dembo, Y. V. Zubavichus, A. A. Korlyukov, A. G. Filatova. // Bulletin of the Russian Academy of Sciences: Physics. – 2009. – V. 73. – No. 4. – P. 471 – 473.
- A47. Волков, В. В. Определение размера и фазового состава наночастиц серебра в гель-пленке бактериальной целлюлозы методами малоуглового рентгеновского рассеяния, электронной дифракции и электронной микроскопии / В. В. Волков, В. В. Клечковская, Э. В. Штыкова, К. А. Дембо, Н. А. Архарова, Г. И. Ивакин, Р. Ю. Смыслов // Кристаллография. – 2009. – Т. 54. – №2. – С. 197 – 201.
- A48. Starodubtsev, S. G. Montmorillonite-polycation multilayers incorporated in polyacrylamide / S. G. Starodubtsev, E. K. Lavrentyeva, E. V. Shtykova, K. A. Dembo, V. V. Volkov. // Applied Clay Science. – 2009. – V. 46. – P. 88 – 94.
- A49. Штыкова, Э. В. Формирование наночастиц серебра в матрице ПВП в среде сверхкритического CO<sub>2</sub>: малоугловое рентгеновское рассеяние и моделирование / Э. В. Штыкова, К. А. Дембо, В. В. Волков, Э. Е. Саид-Галиев, А. И. Стаханов, А. Р. Хохлов // Российские нанотехнологии. – 2009. – Т. 4. – № 9–10. – С. 64 – 71.
- A50. Starodubtsev, S. G. Comb-like poly(4-vinylpyridinium) salts with dodecylsulfate, sodium bis(2-ethylhexyl) sulfosuccinate and bromide counter ions. Small-angle X-

- ray scattering and dynamic light scattering study / S. G. Starodubtsev, T. V. Laptinskaya, A. S. Yesakova, A. R. Khokhlov, E. V. Shtykova, K. A. Dembo, V. V. Volkov // *Polymer*. – 2010. – V. 51. – No. 1. – P. 122 – 128.
- A51. Shtykova, E. V. SAXS–data–based structural modeling of DNA–gadolinium complexes fixed in particles of cholesteric liquid–crystalline dispersions / E. V. Shtykova, V. V. Volkov, V. I. Salyanov, Yu. M. Yevdokimov // *Eur. Biophys. J.* – 2010. – V. 39. – No. 9. – P. 1313 – 1322.
- A52. Priola, A. Effect of the swelling degree on the formation of magnetite nanoparticles in hydrogels / A. Priola, A. Di Gianni, R. Bongiovanna, S. G. Starodubtsev, S. S. Abramchuck, S. N. Polyakov, V. V. Volkov, E. V. Shtykova, K. A. Dembo // *European Polymer Journal*. – 2010. – V. 46. – No. 11. – P. 2105 – 2111.
- A53. Babushkina, T. A. Study of the gel films of *Acetobacter Xylinum* cellulose and its modified samples by  $^1\text{H}$  NMR cryoporometry and small–angle X–ray scattering / T. A. Babushkina, T. P. Klimova, E. V. Shtykova, K. A. Dembo, V. V. Volkov, A. K. Khripunov, V. V. Klechkovskaya // *Crystallography Reports*. – 2010. – V. 55. – No. 2. – P. 312 – 317.
- A54. Логинова, Т. П. Синтез и исследование свойств наночастиц феррита кобальта в гибридных мицеллах полистирол –блок–полиэтиленоксида и додецилсульфата натрия / Т. П. Логинова, О. В. Лыхина, Е. А. Юданова, И. А. Хотина, Г. И. Тимофеева, О. Л. Лепендина, В. В. Волков, К. А. Дембо, С. П. Солодовников // *Высокомолек. соед. Серия А*. – 2010. – Т. 52. – № 8. – С. 1458 – 1465.
- A55. Said–Galiev, E. E. Synthesis of Ag and Cu–Chitosan Metal–Polymer Nanocomposites in Supercritical Carbon Dioxide Medium and Study of Their Structure and Antimicrobial Activity / E. E. Said–Galiev, A. I. Gamzazade, T. E. Grigor'ev, A. R. Khokhlov, N. P. Bakuleva, I. G. Lyutova, E. V. Shtykova, K. A. Dembo, V. V. Volkov // *Nanotechnologies in Russia*. – 2011. – V. 6. – No. 5–6. – P. 341 – 352. – ISSN 1995–0780.

- A56. Skuridin, S. G. Retention of gold nanoparticles in the structure of quasinematic layers formed by DNA molecules / S. G. Skuridin, V. A. Dubinskaya, E. V. Shtykova, V. V. Volkov, V. M. Rudoy, O. V. Dement'eva, V. A. Kuzmin, E. S. Lisitsyna, S. T. Zakhidov, I. A. Zelenina, Yu. M. Yevdokimov // *Biochemistry. Supplemental Series A.* – 2011. – V. 5. – No. 2. – P. 191 – 197.
- A57. Божко, Н. Н. Исследование особенностей структурообразования в системах поливиниловый спирт – йодид калия – йод методом измерения малоуглового рентгеновского рассеяния / Н. Н. Божко, В. П. Столяров, Е. Б. Баблюк, В. Г. Назаров, В. В. Волков, С. В. Амарантов, К. А. Дембо // *Высокомолекулярные соединения. Серия А.* – 2011. – Т. 53. – № 9. – С. 1537 – 1541.
- A58. Свергун, Д. И. Рентгеновское малоугловое рассеяние, синхротронное излучение и структура био– и наносистем / Д. И. Свергун, Э. В. Штыкова, В. В. Волков, Л. А. Фейгин // *Кристаллография.* – 2011. – Т. 56. – № 5. – С. 777–804.
- A59. Владимиров, Ю. А. Белково–липидные наночастицы, от которых зависит “быть или не быть” живой клетке / Ю. А. Владимиров, Ю. Ц. Ноль, В. В. Волков // *Кристаллография.* – 2011. – Т. 56. – № 4. – С. 596 – 603.
- A60. Брик, В. В. Механизм образования, структура и адсорбционные характеристики микропористых нанокристаллических тонких пленок композитов (V,Ti)–N–He / В. В. Брик, Р. Л. Василенко, А. В. Гончаров, А. В. Грогорова, А. Г. Гуглия, В. Г. Колобродов, М. Л. Литвиненко, И. Г. Марченко, Е. С. Мельникова, I. V. Sassa, Ю. В. Никитенко, Ю. Н. Хайдуков, В. В. Волков, G. Chaboussant // *Поверхность.* – 2011. – No. 6. – С. 66 – 75.
- A61. Volkov, V. V. A comparative study of immunoglobulin IgM and rheumatoid factor IgMRF in solution by small–angle X–ray scattering / V. V. Volkov, V. A. Lapuk, A. I. Toropova, E. Yu. Varlamova, E. V. Shtykova, K. A. Dembo, V. P. Timofeev // *Mendeleev Communications.* – 2012. – V. 22. – P. 159 – 161.

- A62. Коригодский, А. Р. Короткоблочные полимочевин–полидиметил–силоксановые блок–сополимеры на основе вторичных диаминов / А. Р. Коригодский, С. А. Барулина, В. В. Киреев, Н. П. Бессонова, В. В. Волков, К. А. Дембо, Э. В. Штыкова, М. А. Ванцян, Е. А. Симакина // Высокомолекулярные соединения. – 2013. – Т. 55. – № 4. – С. 488 – 495.
- A63. Volkov, V. V. Difference between conformations of immunoglobulin M and human rheumatoid factor based on small–angle X–ray solution scattering data / V. V. Volkov, V. A. Lapuk, D. I. Chekrygina, E. Yu. Varlamova, A. V. Chekushin // Mendeleev Comm. – 2013. – Т. 23. – Р. 319 – 320.

### **Стандарты Российской Федерации**

ГОСТ Р 8.698 – 2010. Размерные параметры наночастиц и тонких пленок. Методика выполнения измерений с помощью малоуглового рентгеновского дифрактометра. – Введ. 2010–02–10. – М.: Стандартинформ, 2010. – 41 с. (Государственная система обеспечения единства измерений). Введен впервые.

### **Авторские свидетельства**

Дедков Ю.М., Волков В.В. Способ количественного определения третичных аминов в водных растворах. – Авторское свидетельство No 792117 от 1 сент. 1980г.

Клечковская В.В., Архарова Н.А., Волков В.В.. Высокотемпературная приставка к электронографу. – Дата подачи 29 сентября 2006 г. рег. номер 2006134542, получено решение о выдаче патента.

**Тезисы докладов на конференциях**

1. Петров, И.П. Идентификация органических соединений, содержащихся в сточных водах, по хроматомасс–спектрометрическим данным с применением ЭВМ / И.П.Петров, Б.Г.Дерендяев, В.В.Волков, В.А.Немцев, В.А.Коптюг / Тез. докладов 4 Всесоюзной конференции по аналитической химии органических соединений. – Москва, 1980 – с. 122.
2. Гончаров, А.И. Применение метода варьирования концентраций компонентов для анализа сложных смесей органических соединений / А.И. Гончаров, В.В. Волков, И.П. Петров, Б.Н. Гречушников, И.Н. Калинкина, В.А.Немцев / Тез. докладов 5 Всесоюзной конференции по использованию вычислительных машин в спектроскопии молекул и химических исследованиях. – Новосибирск, 1980. –с. 176–177.
3. Гречушников, Б.Н. Применение преобразования проектирования в спектрофотометрическом анализе сложных смесей / Б.Н. Гречушников, В.В. Волков, И.Н. Калинкина, И.П. Петров / Тез. докладов 6 Всесоюзной конференции по использованию вычислительных машин в спектроскопии молекул. – Новосибирск, 1983. – с. 165.
4. Волков, В.В. Оценка надежности решения при хемометрическом анализе сложных смесей по их спектрам / В.В. Волков /Тез. докладов 6 Всесоюзной конференции по использованию вычислительных машин в спектроскопии молекул. – Новосибирск, 1983. – с. 166–167.
5. Гречушников, Б.Н. Уточнение концентраций и определение спектров индивидуальных компонентов по спектрам их смесей / Б.Н. Гречушников, В.В. Волков, И.Н. Калинкина / Тез. докладов Всесоюзной конференции "Математические методы и ЭВМ в аналитической химии." – Москва, 1986. – с. 117.
6. Волков, В.В. Оценивание спектров компонентолв по спектрам родственных смесей / В.В. Волков, Б.Н. Гречушников, И.Н. Калинкина / Тез. докладов

Всесоюзной конференции "Математические методы и ЭВМ в аналитической химии." – Москва, 1986 – с. 118.

7. Волков, В.В. Программная реализация методов разложения спектров родственных смесей на спектры компонентов / В.В. Волков / Тез. докладов Всесоюзной конференции "Математические методы и ЭВМ в аналитической химии." – Москва, 1986. – с. 119.
8. Akhmanova, M.V. Laser optoacoustic determination of atmospheric pollutants and their computer identification / M.V. Akhmanova, A.V. Davydov, A.V. Kuznetsov, A.E. Platonov, V.V. Volkov / 4th International Conference on Laser Applications in Life Sciences. – Jyvaskyla, Finland, 7–11 September, 1992. – P. 113.
9. Volkov, V.V. Shape Determination in Small–Angle Scattering: Theory and Practice / V.V. Volkov, M.B. Kozin, H.B. Stuhrmann, D.I. Svergun. / X International Conference on Small–Angle Scattering. July 21–25, 1996. – Campinas Sao Paulo Brazil. – P. 118.
10. Kozin, M.B. Real–Time Shape Determination from Solution Scattering Data / M.B. Kozin, V.V. Volkov, D.I. Svergun / X International Conference on Small–Angle Scattering. July 21–25, 1996. – Campinas SP Brazil. – P. 105.
11. Svergun, D.I. Contrast Variation Study of the 70S E.Coli Ribosome / D.I. Svergun, M.H.J. Koch, V.V. Volkov, M.B. Kozin, W. Meerwink, H.B. Stuhrmann, J.Skov Pedersen, N. Burkhardt, G. Diedrich, K.H. Nierhaus. / X International Conference on Small–Angle Scattering. July 21–25, 1996, Campinas SP Brazil. – P. 226.
12. Svergun, D.I. Shape determination of dimeric particles / D.I. Svergun, M.H.J. Koch, V.V. Volkov, C. Barberato / X International Conference on Small–Angle Scattering. July 21–25, 1996 – Campinas SP Brazil. – P. 105.
13. Козин, М.Б. Автоматическое определение формы частицы в малоугловом рассеянии монодисперсными системами / М.Б. Козин, В.В. Волков, Д.И. Свергун / Тезисы докл. Национальной конференции по применению рентгеновского, синхротронного излучений, нейтронов и электронов для

исследования материалов РСНЭ'97. – Москва–Дубна, 25–29 мая 1997. – с. 231.

14. Свергун, Д.И. Исследование структуры 70S E.coli рибосомы методом малоуглового рассеяния с вариацией контраста / Д.И. Свергун, N. Burkhardt, J. Skov Pedersen, M.H.J. Koch, В.В. Волков, М.Б. Козин, W.Meerwink, Н.В. Stuhrmann, G.Dietrich, К.Н. Nierhaus / Тезисы докл. Национальной конференции по применению рентгеновского, синхротронного излучений, нейтронов и электронов для исследования материалов РСНЭ'97. – Москва–Дубна, 25–29 мая 1997. – с. 242.
15. Svergun, D.I. Structure of the 70S E.coli Ribosome in Solution from Contrast Variation on Selectively Deuterated Particles / D.I. Svergun, V.V. Volkov, N. Burkhardt, J. Skov Pedersen, M.H.J. Koch, M.B. Kozin, W. Meerwink, Н.В. Stuhrmann, G. Dietrich, К.Н. Nierhaus / International workshop "Deuteration of Biological Molecules for Structural and Dynamic Studies. Applications to Neutron Scattering and NMR (DBM–98)" (1998). – Dubna (Russia), May 19–25. – P. 119.
16. Конарев, П.В. Исследование структурных изменений в трехкомпонентных микроэмульсиях типа вода–АОТ–изооктан методом малоуглового рассеяния / П.В. Конарев, Д.И. Свергун, В.В. Волков, М.Н.К. Koch, J. Smeets, W.F.C. Sager / II Национальная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования материалов (РСНЭ–99), – Москва, 1999, 23–27 мая, – Материалы конференции. – с. 280.
17. Свергун, Д.И. О единственности определения формы молекул биополимеров по данным малоуглового рассеяния / Д.И. Свергун, В.В. Волков / II Национальная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования материалов (РСНЭ–99), – Москва, 1999, 23–27 мая, – Материалы конференции. – с. 292.
18. Лисин, Ф. Исследование структуры субъединицы 30S рибосомы Thermus Thermophilus методами нейтронного и рентгеновского малоуглового рассеяния / Ф. Лисин, Д.И. Свергун, В.В. Волков, И.Н. Сердюк / Национальная

- конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования материалов (РСНЭ–99). – Москва, 1999, 23–27 мая. – Материалы конференции. – с. 320.
19. Volkov, V.V. *Ab initio* Determination of Low Resolution Particle Shape from Small–Angle Scattering Data. / V.V. Volkov, M.V. Petoukhov, D.I. Svergun / Second International Conference on Modern Trends in Computational Physics. July 24–29, 2000. – Dubna, Russia. – Book of abstracts, – p. 142.
20. Volkov, V. Basic numerical and statistical methods / V. Volkov / EMBO Practical Course on "Solution Scattering from Biological Macromolecules", European Molecular Biological Laboratory (EMBL), – Hamburg, Germany, 5–14 September 2000. – P. 8.
21. Штыкова, Э.В. Сравнительное малоугловое рентгеновское рассеяние и моделирование структуры комплексов к– и i–каррагинанов с ПАВ. / Э.В. Штыкова, Д.И. Свергун, Е.В. Штыкова, В.В. Волков, П.В. Конарев, А.Т. Дембо, Е.Е. Махаева, А.Р. Хохлов, Г.А. Евмененко, Х. Райнерс / – РСНЭ–2001. – Москва, 2001. – с. 180
22. Волков, В.В. Однозначность определения 3D структуры низкого разрешения частиц в растворе по данным малоуглового рассеяния / В.В. Волков, Д.И. Свергун / – РСНЭ–2001. – Москва, 2001. – с.306.
23. Конарев, П.В. Massha – новый программный пакет для моделирования трехмерных структур по данным малоуглового рассеяния. / П.В. Конарев, М.В. Петухов, В.В. Волков, Д.И. Свергун / РСНЭ–2001. – Москва, 2001. – с. 319
24. Волков, В.В. Свергун Определение строения молекулы иммуноглобулина М человека по данным малоуглового рентгеновского рассеяния от раствора и атомно–силовой микроскопии. / В.В. Волков, В.А. Лапук, Р.Л. Каюшина, А.Л. Толстихина, Р.В. Гайнутдинов, Д.И./ Российская конференция по электронной микроскопии РКЭМ–2002. – Черногоровка, 2002. – с. 201.
25. Volkov, V.V. Determination of Low Resolution Structure of Human

- Immunoglobulin M and Rheumatoid Factor IgM–RF in Solution. / V.V. Volkov, R.L. Kayushina, V.A. Lapuk, A.L. Tolstikhina, R.V. Gaynutdinov, E.V. Shtykova, D.I. Svergun / International Workshop on Scanning Probe Microscopy. – Nizhny Novgorod, March 3–6, 2002. – p. 220–222.
26. Volkov, V.V. Low Resolution Structure of Immunoglobulins IgG, IgM and Rheumatoid Factor IgM–RF from Solution X–ray Scattering Data / V.V. Volkov, V.A. Lapuk, R.L. Kayushina, E.V. Shtykova, E.Yu. Varlamova, D.I. Svergun / XIII International Conference On Small–Angle Scattering. – Venice, Italy, August 25–29, 2002. – p. 28–29.
27. Volkov, V.V. Uniqueness of *ab initio* shape determination in small angle scattering. / V.V. Volkov, D.I. Svergun / XIII International Conference On Small–Angle Scattering. – Venice, Italy, August 25–29, 2002. – p. 177.
28. Konarev, P.V. WINTEL–based menu–driven program packages for small–angle scattering data analysis. / P.V. Konarev, M.V. Petoukhov, V.V. Volkov, D.I. Svergun / XIII International Conference On Small–Angle Scattering. – Venice, Italy, August 25–29, 2002. – p. 167.
29. Sokolova, A.V. Prototype of Database for Rapid Protein's Classification Based on Solution Small–Angle Scattering Data. / A.V. Sokolova, D.I. Svergun, V.V. Volkov / XIII International Conference On Small–Angle Scattering. – Venice, Italy, August 25–29, 2002. – p. 176.
30. Соколова, А.В. Быстрый поиск белков с похожей формой и аналогов по доменной структуре: новый подход к анализу данных рентгеновского малоуглового рассеяния. / А.В. Соколова, В.В. Волков, Д.И. Свергун / IV Национальная конференция по применению рентгеновского, синхротронного излучений, нейтронов и электронов для исследования материалов РСНЭ–2003. – Москва, 17–22 ноября 2003г. – с. 322.
31. Соколова, А.В. Новый метод классификации белков с неизвестной внутренней структурой по данным рентгеновского малоуглового рассеяния. / А.В. Соколова, В.В. Волков, Д.И. Свергун / III Съезд биофизиков. – Воронеж, 24–29

июня 2004. – Сб. тез. т.1. – с. 99–100.

32. Klechkovskaya, V.V. Structure characterization of nanoscale organic–inorganic composites, partially ordered and disordered systems. / V.V. Klechkovskaya, V.V. Volkov, E.V. Shtykova, L.A. Feigin / I France–Russia Seminar “New Achievements in Materials Science”. – Nancy, France, 27–29 October, 2004. – Abstracts. – p. 156–159.
33. Said–Galiyev, E. Structure and morphology of metal–polymeric nanocomposites of noble metals made by means of critical technology / E. Said–Galiyev, L. Nikitin, A. Khokhlov, E. Stykova, V. Volkov, K. Dembo, O. Lebedeva, A. Aliev / 2–nd International Conference “Supercritical Fluid Technology: innovation potential of Russia”. – Rostov–na–Donu, Russia, 2005. – Abstracts. – P. 86–88.
34. Shtykova, E.V. Autoxidized phospholipid nanoparticles in hexane studied by synchrotron small–single X–ray scattering. / E.V. Shtykova, V.V. Volkov, H. Wang, T. Fujisawa, Jin–Ye Wang / European Polymer Congress EPF 2005. – M.V. Lomonosov Moscow State University. – Moscow, Russia, June 27 – July 1 2005. – Abstracts, ref. 5283. – P. 149.
35. Shtykova, E.V. Self–assembled polymer matrices as media for metal nanoparticles formation: anomalous scattering and electron diffraction study / E.V. Shtykova, V.V. Volkov, V.V. Klechkovskaya / II France–Russia Seminar “New Achievements in Materials Science”. – Moscow, 10–12 November, 2005. – Abstracts. – P. 67–69.
36. Волков, В.В. О выборе параметров поиска структурных моделей наночастиц по данным малоуглового рассеяния. / В.В. Волков, Г.В. Парохонский, Д.И. Свєргун / V Национальная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования наноматериалов и наносистем РСНЭ НАНО–2005. – Москва, ИК РАН, 14–19 ноября 2005. – Сб.тезисов. – с. 315.
37. Волков, В.В. Исследование формы фрагментов Fab, Fab–RF, Fc и Rc–RF молекул иммуноглобулина М и ревматоидного фактора методом малоуглового рентгеновского рассеяния от раствора и атомно–силовой микроскопии. / В.В.

- Волков, В.А. Лапук, М.В. Занавескин, А.Л. Толстихина, Э.В. Штыкова, Д.И. Свергун / V Национальная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования наноматериалов и наносистем РСНЭ НАНО–2005. – Москва, ИК РАН, 14–19 ноября 2005. – Сб.тезисов. – с. 108.
38. Волков, В.В. Определение формы белка регулятора Hfq *E.coli* в растворе по данным малоуглового рентгеновского рассеяния. / В.В. Волков, С.В. Кравченко, К.А. Дембо., А.В. Соколова / V Национальная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования наноматериалов и наносистем РСНЭ НАНО–2005. – Москва, ИК РАН, 14–19 ноября 2005. – Сб.тезисов. – с. 105.
39. Штыкова, Э.В. Исследование наночастиц фосфолипидов в гексане методом малоуглового рентгеновского рассеяния. / Э.В. Штыкова, В.В. Волков, С.В. Амарантов, Н. Wang, Т. Fujisawa, Jin–Ye Wang. / V Национальная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования наноматериалов и наносистем РСНЭ НАНО–2005. – Москва, ИК РАН, 14–19 ноября 2005. – Сб.тезисов. – с. 88.
40. Volkov, V.V. Investigation of Isolated Fab Fragments of Immunoglobulin IgM and Rheumatoid Factor by SAXS, ESR, and Mass Spectrometry. / V.V. Volkov, V.A. Lapuk, E.V. Shtykova, K.A. Dembo, V.P. Timofeev, R.H. Ziganshin, A.V. Sokolova / XIII International Conference on Small–angle Scattering (SAS2006). – Japan, Kyoto, July 9 – 13, 2006. – Abstracts. – P. 48.
41. Volkov, V.V. Determination of shape of Fab and Fc5 fragments separated from immunoglobulin IgM and rheumatoid factor from SAXS solution data by computer modeling / V.V. Volkov, V.A. Lapuk, K.A. Dembo, A.V. Sokolova, E.V. Shtykova, S.V. Amarantov, E.Yu. Varlamova / Electron Microscopy and Multiscale Modeling EMMM–2007. – Moscow, 3–7 September 2007. – Abstracts. – P. 96.
42. Volkov, V.V. Investigation of uniqueness of shape determination of nanoparticles from X–Ray small–angle scattering data by computer modeling / V.V. Volkov, S.V.

Amarantov, A.V. Sokolova, D.I. Svergun / Electron Microscopy and Multiscale Modeling EMMM–2007. – Moscow, 3–7 September 2007. – Abstracts. – P. 97.

43. Крупянский, Ю.Ф. Эффекты взаимодействия химических шаперонов с ферментами: изменение структуры, динамики и функциональной активности / Ю.Ф. Крупянский, О.А. Коротина, С.В. Есин, М.Г. Михайлюк, Д.С. Родин., К.А. Дембо, В.В. Волков, Ю.А. Николаев, Г.И. Эль–Регистан / VI Международная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования материалов (РСНЭ–2007). – Москва, 12–17 ноября, 2007. – Сб. тезисов, – с. 46.
44. Амарантов, С.В. Аналитический расчет интенсивности малоуглового рассеяния и расчет функции Паттерсона для единичного потенциала, заданного на торе / С.В. Амарантов, В.В. Волков / VI Международная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования материалов (РСНЭ–2007). – Москва, 12–17 ноября, 2007. – Сб. тезисов. – с. 400.
45. Волков, В.В. Комплексное сравнительное исследование фрагментов структур иммуноглобулина М и ревматоидного фактора методами малоуглового рентгеновского рассеяния, ЭПР и масс–спектропии / В.В. Волков, В.А. Лапук, К.А. Дембо, Н.Д. Степина, Э.В. Штыкова, М.Л. Занавескин, С.В. Амарантов, А.В. Соколова, Е.Ю. Варламова / VI Международная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования материалов (РСНЭ–2007). – Москва, 12–17 ноября, 2007. – Сб. тезисов. – с. 409.
46. Лапук, В.А. Применение малоуглового рентгеновского рассеяния для уточнения процесса фрагментации моноклональных IgM и ревматоидного IgM (IgM–RF) «горячим трипсином» / В.А. Лапук, В.В. Волков, К.А. Дембо, Э.В. Штыкова, С.В. Амарантов, Р.В. Гайнутдинов, М.Л. Занавескин, Е.Ю. Варламова / VI Международная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования

- материалов (РСНЭ–2007). – Москва, 12–17 ноября, 2007. – Сб. тезисов. – с. 127.
47. Валькова, Л.А. Моделирование слоевой упаковки в ЛБ пленках аминзамещенного фталоцианина кобальта / Л.А. Валькова, Л.Г. Янусова, В.В. Волков, С.Б. Астафьев / VI Международная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования материалов (РСНЭ–2007). – Москва, 12–17 ноября, 2007. – Сб. тезисов. – с. 474.
48. Хомутов, Г.Б. Получение и исследование структуры новых организованных органико–неорганических наноматериалов / Г.Б. Хомутов, В.В. Волков, В.В. Ключевская / VI Международная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования материалов (РСНЭ–2007). – Москва, 12–17 ноября, 2007. – Сб. тезисов. – с. 218.
49. Штыкова, Э.В. Малоугловое рентгеновское рассеяние в исследовании наноструктурированных материалов / Э.В. Штыкова, В.В. Волков, Л.М.Бронштейн, Д.И. Свєргун / VI Международная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования материалов (РСНЭ–2007). – Москва, 12–17 ноября, 2007. – Сб. тезисов. – с. 222.
50. Волков, В.В. Исследование однозначности и устойчивости определения формы частиц в монодисперсных системах по данным малоуглового рассеяния с помощью компьютерного моделирования / В.В. Волков, Д.И. Свєргун, С.В. Амарантов, К.А. Дембо / VI Международная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования материалов (РСНЭ–2007). – Москва, 12–17 ноября, 2007. – Сб. тезисов. – с. 384.
51. Запорожец, М.А. Полупроводниковые наночастицы – основа уникальных оптических материалов. / М.А. Запорожец, Д.А. Баранов, О.М. Жигалина,

- В.В. Волков, С.Н. Сульянов, В.И. Николайчик, И.И. Ходос, С.П. Губин, А.С. Авилов / XI Симпозиум "Современная химическая физика". – Туапсе, 22 сентября – 3 октября 2007. – Сб. тезисов, – с. 73.
52. Запорожец, М.А. Структура полупроводниковых наночастиц CdS. / М.А. Запорожец, Д.А. Баранов, О.М. Жигалина, В.В. Волков, В.И. Николайчик, И.И. Ходос, С.П. Губин, А.С. Авилов / XXII Российская конференция по электронной микроскопии. – Черногоровка 2–6 июня 2008. – Сб. тезисов. – с. 169.
53. Запорожец, М.А. Монодисперсные наночастицы CdSe. / М.А. Запорожец, Д.А. Баранов, О.М. Жигалина, В.В. Волков, В.И. Николайчик, И.И. Ходос, С.П. Губин, А.С. Авилов / XIII Национальная конференция по росту кристаллов. – Москва, 17–21 ноября 2008. – Сб. тезисов. – с. 427.
54. Volkov, V.V. Uniqueness and consistency of particle shape determination from small-angle scattering data: computer modeling / V.V. Volkov, S.V. Amarantov, E.V. Shtykova, D.I. Svergun / XXI Congress of the International Union of Crystallography (IUCr2008), August 23–31. – Osaka, Japan, 2008. – Book of Abstracts & Acta Crys. – 2008. A64. – p. 226–227.
55. Shtykova, E.V. Polymer matrices as templates for metal nanoparticles / E.V. Shtykova, V.V. Volkov, L.M. Bronstein, D.I. Svergun / 7th International Conference on X-Ray investigations of Polymer Structure (XIPS'2007), 5 – 7 December 2007. – Krakow, Poland, 2007. – Abstracts. – p. 16.
56. Волков, В.В. Малоугловое рентгеновское рассеяние и структура биологических, полимерных и композитных наноматериалов. / В.В. Волков / Международный форум по нанотехнологиям 3–15.12.2008. – Москва, 2008. – Сб. тезисов. – Т.1. – с. 192.
57. Babushkina, T.A. Interactions in the Cellulose Acetobacter Xylinum – Water System: X – Ray Small–Angle Scattering and Pulsed Nuclear Magnetic Resonance Porometry Study / T.A. Babushkina, T.P. Klimova, A.K. Khripunov, E.V. Shtykova, K.A. Dembo, V.V. Volkov, V.V. / 6th International Symposium “Molecular Order

and Mobility in Polymer Systems”, (2008). – St.Peterburg, June 2–6, 2008. – Abstracts. – p. 222.

58. Штыкова, Э.В. Исследование комплексов гадолиния с линейной и жидкокристаллическими формами ДНК методом малоуглового рентгеновского рассеяния / Э.В. Штыкова, К.А. Дембо, В.В. Волков, Ю.М. Евдокимов / XIII Национальная конференция по росту кристаллов (НКРК–2008). – Москва, 17–21 ноября 2008. – Сб. тезисов. – с. 488.
59. Марченко, И.В. Исследование наночастиц серебра в оболочке полиэлектролитных капсул, сформированных на ядрах из карбоната кальция / И.В. Марченко, Б.В. Парахонский, Т.В. Букреева, Ю.В. Григорьев, В.В. Волков, Г.С. Плотников, Л.А. Фейгин / Конференция–школа для молодых ученых «Дифракционные методы исследования вещества: от молекул к кристаллам и наноматериалам». – Черноголовка, 30 июня – 3 июля 2008. – Сб.тезисов, – с. 31.
60. Крупянский, Ю.Ф. Влияние химических шаперонинов на структуру, динамику и функцию белков. / Ю.Ф. Крупянский, Э.Г. Абдулнасыров, Н.Г. Лойко, П.П. Нокс, А.С. Степанов, К.Б. Терешкина, В.В. Волков, Г.И. Эль–Регистан, А.Б. Рубин / VIII Национальная конференция «Рентгеновское, Синхротронное излучения, Нейтроны и Электроны для исследования наносистем и наноматериалов. Нано–Био–Инфо–Когнитивные технологии» РСНЭ–НБИК 2011. Москва, 2011. Тезисы докладов. С.28. (Устный)
61. Родина, Е.В. Структурно–функциональная характеристика фрыктозо–1,6–бисфосфатальдозазы класса I из E.coli / Е.В. Родина, А.В. Сазонов, Н.Н. Воробьева, С.А. Курилова, В.Р. Самыгина, В.В. Волков, Э.В. Штыкова, К.А. Дембо, Л.А. Дадинова, Ю.В. Кобыльская, М. Groves, V. Lamzin, Т.И. Назарова / VIII Национальная конференция «Рентгеновское, Синхротронное излучения, Нейтроны и Электроны для исследования наносистем и наноматериалов. Нано–Био–Инфо–Когнитивные технологии» РСНЭ–НБИК 2011. – Москва, 2011. – Сб. тезисов. – с. 34.

62. Саид–Галиев, Э.Е. Связь структурной организации металлокомпозитов, полученных с участием сверхкритического диоксида углерода, со свойствами флюида и матрицы / Э.Е. Саид–Галиев, А.Ю. Николаев, А.Р. Хохлов, Э.В. Штыкова, В.В. Волков, К.А. Дембо / VIII Национальная конференция «Рентгеновское, Синхротронное излучения, Нейтроны и Электроны для исследования наносистем и наноматериалов. Нано–Био–Инфо–Когнитивные технологии» РСНЭ–НБИК 2011. – Москва, 2011. – Сб. тезисов. – с. 36.
63. Скуридин, С.Г. Изучение структурных характеристик холестерических жидкокристаллических дисперсий ДНК с инкорпорированными наночастицами золота методом малоуглового рентгеновского рассеяния / С.Г. Скуридин, Ю.М. Евдокимов, В.В. Волков, К.А. Дембо, Э.В. Штыкова. / VIII Национальная конференция «Рентгеновское, Синхротронное излучения, Нейтроны и Электроны для исследования наносистем и наноматериалов. Нано–Био–Инфо–Когнитивные технологии» РСНЭ–НБИК 2011. – Москва, 2011. – Сб. тезисов. – с. 79.
64. Штыкова, Э.В. Матриксный белок вируса гриппа М1: малоугловое рентгеновское рассеяние и моделирование / Э.В. Штыкова, Л.А. Баратова, В.В. Волков, К.А. Дембо, Д.И. Свергун / VIII Национальная конференция «Рентгеновское, Синхротронное излучения, Нейтроны и Электроны для исследования наносистем и наноматериалов. Нано–Био–Инфо–Когнитивные технологии» РСНЭ–НБИК 2011. – Москва, 2011. – Сб. тезисов. – с. 92.
65. Казак, А.В. Надмолекулярная организация мезо–замещенных производных тетрафенилпорфиринов в тонких пленках / А.В. Казак, Н.В. Усольцева, С.Г. Юдин, К.А. Дембо, В.В. Волков / VIII Национальная конференция «Рентгеновское, Синхротронное излучения, Нейтроны и Электроны для исследования наносистем и наноматериалов. Нано–Био–Инфо–Когнитивные технологии» РСНЭ–НБИК 2011. – Москва, 2011. – Сб. тезисов. – с. 161.
66. Волков, В.В. Адаптивный алгоритм сглаживания экспериментальных данных, содержащих резкие перепады интенсивности с большим динамическим

- диапазоном / В.В. Волков / VIII Национальная конференция «Рентгеновское, Синхротронное излучения, Нейтроны и Электроны для исследования наносистем и наноматериалов. Нано–Био–Инфо–Когнитивные технологии» РСНЭ–НБИК 2011. – Москва, 2011. – Сб. тезисов. – с. 390.
67. Лапук, В.А. Моделирование формы молекул ревматоидного фактора человека по данным малоуглового рентгеновского рассеяния / В.А. Лапук, В.В. Волков, Э.В. Штыкова, А.И. Торопова, К.А. Дембо, Е.Ю. Варламова, В.П. Тимофеев / VIII Национальная конференция «Рентгеновское, Синхротронное излучения, Нейтроны и Электроны для исследования наносистем и наноматериалов. Нано–Био–Инфо–Когнитивные технологии» РСНЭ–НБИК 2011. – Москва, 2011. – Сб. тезисов. – с. 400.
68. Амарантов, С.В. Исследование методом малоуглового рентгеновского рассеяния белков – шаперонинов в различных конформациях / С.В. Амарантов, И.Н. Налетова, Л.П. Курочкина / VIII Национальная конференция «Рентгеновское, Синхротронное излучения, Нейтроны и Электроны для исследования наносистем и наноматериалов. Нано–Био–Инфо–Когнитивные технологии» РСНЭ–НБИК 2011. – Москва, 2011. – Сб. тезисов. – с. 425.
69. Чекушин, А.В. Модификация алгоритма расчета распределений по размерам наночастиц из данных малоуглового рентгеновского рассеяния / А.В. Чекушин, В.В. Волков / VIII Национальная конференция «Рентгеновское, Синхротронное излучения, Нейтроны и Электроны для исследования наносистем и наноматериалов. Нано–Био–Инфо–Когнитивные технологии» РСНЭ–НБИК 2011. – Москва, 2011. – Сб. тезисов. – с. 484.
70. Vladimirov, Yu.A. Cytochrome C – Cardiolipin Complex: A Lipoprotein Nanostructure that Turns on the Cascade of Reactions Terminated by Self-destruction of Living Cells / Yu.A. Vladimirov, A.V. Alekseev, V.V. Volkov, B.V. Nabatov / 3rd Russian–Hellenic Symposium with International Participation and Young Scientists School "BIONANOTOX 2012" "Biomaterials and Bionanomaterials: Recent Problems and Safety Issues". – Heraclion, Crete, Greece.

6–13 May 2012. – Abstracts. – P. 21.

71. Shtykova, E. Solution Structure of Virus Matrix Protein M1 Revealed by SAXS. / E. Shtykova, L. Baratova, N. Fedorova, V. Volkov, V. Radyukhin, A. Ksenofontov, A. Shishkov, A. Dolgov, L. Shilova, A. Indenbom / 15th Int. Small Angle Scattering Conf. 2012. – Sydney, Australia, 2012. – Abstracts. – P. 514.
72. Chekrygina, D.I. Comparative morphology of macromolecules of Immunoglobulin-M and human Rheumatoid Factor from SAXS data / D.I. Chekrygina, V. Volkov, V. Lapuk, E. Varlamova / RACIRI Summer school 2013. – Петергоф, 17–25 июня 2013. – Сб. тезисов. – с. 24–26.
73. Amarantov, S. Size Distribution and Shape of Alpha-crystallin in Solution from Analytical Ultracentrifugation and SAXS Data. / S. Amarantov, N. Chebotareva, N. Polyanskij, K. Muranov / 11th International Conference Biology and Synchrotron Radiation 2013 (BSR–2013). – Hamburg, 8–11th September 2013. – Abstracts. – p. 67.
74. Dadinova, L. SAXS-derived 3D models of a novel bacterial fructose-1,6-bisphosphate aldolase / L. Dadinova, E. Rodina, N. Vorobyeva, V. Volkov, E. Shtykova / 11th International Conference Biology and Synchrotron Radiation 2013 (BSR–2013). – Hamburg, 8–11th September 2013. – Abstracts. – p. 75.
75. Konarev, P. POLYSAS: an interactive graphical system for small-angle scattering analysis of polydisperse systems and multiple data sets / P. Konarev, V. Volkov, D. Svergun / 11th International Conference Biology and Synchrotron Radiation 2013 (BSR–2013). – Hamburg, 8–11th September 2013. – Abstracts. – p. 91.
76. Shtykova, E. Shape Determination of Polydisperse and Polymorphic Objects in SAXS: Possibilities and Limitations / E. Shtykova, V. Volkov, D. Svergun / 11th International Conference Biology and Synchrotron Radiation 2013 (BSR–2013). – Hamburg, 8–11th September 2013. – Abstracts. – p. 116.
77. Shtykova, E. Probing a Complex of Cytochrome C and Cardiolipin by SAXS in Solution and in Precipitate / E. Shtykova, V. Volkov, V. Remenschikov, Yu.

Vladimirov / 11th International Conference Biology and Synchrotron Radiation 2013 (BSR–2013). – Hamburg, 8–11th September 2013. – Abstracts. – p. 117.

78. Shtykova, E. Linear clusters of gold nanoparticles in quasinematic layers of DNA liquid–crystalline dispersion particles at different temperatures: morphology and structural transition / E. Shtykova, L. Dadinova, V. Volkov, S. Skuridin, Yu. Evdokimov / 11th International Conference Biology and Synchrotron Radiation 2013 (BSR–2013). – Hamburg, 8–11th September 2013. – Abstracts. – p. 119.
79. Volkov, V. Modeling of shape of rheumatoid factor from SAXS data: ambiguity of solution and difference from immunoglobulin M / V. Volkov, V. Lapuk, D. Chekrygina, E. Varlamova, A. Chekushin / 11th International Conference Biology and Synchrotron Radiation 2013 (BSR–2013). – Hamburg, 8–11th September 2013. – Abstracts. – p. 125.
80. Volkov, V. Adaptive data denoising in SAXS and diffraction: algorithms based on statistical analysis of residuals / V. Volkov / 11th International Conference Biology and Synchrotron Radiation 2013 (BSR–2013). – Hamburg, 8–11th September 2013. – Abstracts. – p. 126.