

На правах рукописи
УДК 577.32

ЛУНИНА Наталья Леонидовна

**ВЫЧИСЛИТЕЛЬНЫЕ ПОДХОДЫ К РЕШЕНИЮ ФАЗОВОЙ
ПРОБЛЕМЫ КРИСТАЛЛОГРАФИИ БИОПОЛИМЕРОВ ПРИ
НИЗКОМ РАЗРЕШЕНИИ**

03.00.02 - биофизика

автореферат диссертации на соискание ученой степени
кандидата физико-математических наук

Пушино 1998

Работа выполнена в Институте Математических Проблем Биологии РАН (Пушино).

Научный руководитель: кандидат физико-математических наук
А.Г.Уржумцев

Официальные оппоненты : доктор биологических наук
проф. Ю.Н.Чиргадзе.
кандидат физико-математических наук
А.С.Мосунов

Ведущая организация : Институт кристаллографии РАН, г.Москва.

Защита диссертации состоится 27 января 1999 г. в 15.30 часов на заседании диссертационного совета Д 200.22.01 в Институте теоретической и экспериментальной биофизики РАН по адресу: 142292 Пушкино, Московской обл., ИТЭБ РАН

С диссертацией можно ознакомиться в библиотеке ИТЭБ РАН

Автореферат разослан 24 декабря 1998 г.

Ученый секретарь диссертационного
Совета Д-200.22.01
кандидат биологических наук

П.А.Нелипович

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы.

Изучение пространственной структуры биологических макромолекул и их комплексов является важным направлением современной биологии и создает предпосылки для детального понимания механизмов их функционирования. Метод рентгеноструктурного анализа монокристаллов является на сегодняшний день единственным экспериментальным методом, позволяющим получать детальное описание сложных макромолекулярных комплексов при атомном разрешении. Однако спецификой рентгеновского эксперимента является то, что он позволяет получить лишь часть информации (модули структурных факторов), необходимой для восстановления исследуемой структуры. Получение недостающей части информации (значений фаз структурных факторов) представляет собой центральную методологическую проблему рентгеноструктурного анализа. Основные подходы к решению фазовой проблемы, используемые в макромолекулярной кристаллографии, опираются либо на получение химическими методами изоморфных модификаций исследуемого объекта и проведение дополнительных рентгеновских экспериментов, либо на присутствие в объекте аномально рассеивающих атомов, либо на наличие известной атомной структуры гомологичного объекта. Попытки применения этих подходов встречаются с существенными сложностями при работе с макромолекулярными комплексами. В то же время разработаны математические методы, позволяющие решать фазовую проблему для низкомолекулярных соединений, основываясь лишь на данных рентгеновского эксперимента с не модифицированным (нативным) объектом. В связи с этим, в последнее десятилетие одной из актуальных задач макромолекулярной кристаллографии является разработка аналогичных методов решения фазовой проблемы, применимых к макромолекулярным объектам. Такие методы называются обычно прямыми или *ab-initio* методами, чтобы подчеркнуть, что они не требуют наличия ни изоморфных производных, ни известных гомологичных объектов.

Задачи работы.

Задачей исследования являлась разработка процедуры прямого решения фазовой проблемы макромолекулярной кристаллографии при низком разрешении. Решение фазовой проблемы при низком разрешении подразумевает определение значений фаз для нескольких десятков коэффициентов Фурье (структурных

факторов) в разложении в ряд Фурье искомой функции распределения электронной плотности в исследуемом объекте. Наличие такой информации делает возможным расчет соответствующей части ряда Фурье и позволяет получить сведения о локализации объекта в элементарной ячейке кристалла и его внешних очертаниях, что является стартовой точкой для процедур расширения и уточнения набора фаз.

В программу работы входило: исследование степени неоднозначности решения фазовой проблемы при низком разрешении и разработка примененных для этого исследования процедур кластерного анализа; разработка процедуры Монте-Карловского типа для решения фазовой проблемы при низком разрешении; разработка и создание соответствующего программного обеспечения; тестирование метода с использованием объектов с известной структурой; применение разработанных методов к определению структуры рибосомальной частицы T50S из *Thermus thermophilus* при разрешении порядка 40Å.

Научная новизна.

Впервые разработана и опробована методика, позволяющая получить синтезы Фурье низкого разрешения для сложных макромолекулярных комплексов, базируясь лишь на ограниченном наборе данных рентгеновского дифракционного эксперимента с нативным соединением. Для решения этой задачи впервые в кристаллографической практике использованы методы Монте-Карловского сканирования пространства возможных решений фазовой проблемы и применены методы кластерного анализа для анализа результатов сканирования. Полученный для частицы T50S синтез Фурье разрешения 40Å является первым примером *ab-initio* решения фазовой проблемы для столь крупных макромолекулярных объектов (не обладающих высокой внутренней симметрией).

Научная значимость работы.

Разработка методов прямого решения фазовой проблемы рентгеноструктурного анализа для биологических полимеров находится последнее десятилетие в фокусе внимания кристаллографического научного сообщества. Однако основные достигнутые в этой области результаты заключаются в совершенствовании методов низкомолекулярной кристаллографии и ограничены сравнительно небольшими белками; кроме того, они достигнуты при использовании с самого начала работы дифракционного набора очень высокого разрешения (порядка 1.0 Å). Оба эти условия невыполнимы при работе с большими комплексами. Поэтому принципиально важное значение имеет продемонстрированная возможность успешного применения для прямого решения фазовой проблемы альтернативных подходов, не налагающих столь жесткие ограничения на количество экспериментальной информации и применимых к самым большим объектам из изучаемых на сегодняшний день методами рентгеновской кристаллографии.

Самостоятельное значение имеет проведенный анализ степени неоднозначности применяемых в кристаллографии критериев качества набора фаз. Предложенные для анализа процедуры кластерного анализа имеют выходящее за рамки данной методики значение и могут быть применены (и применялись) в других задачах кристаллографии макромолекул.

Структура диссертации.

Диссертация состоит из 3 глав, содержит 3 таблицы и 23 рисунка, а также список литературы (118 названий). Объем диссертации 96 страниц.

Апробация работы и публикации.

Результаты работы докладывались на следующих научных конференциях:

- 15-ая Европейская кристаллографическая конференция, Дрезден, Германия, 1994;
 - Ежегодный конгресс Американской кристаллографической ассоциации, Атланта, США, 1994;
 - Ежегодный конгресс Французской кристаллографической ассоциации, Гренобль, Франция, 1995;
 - 4-ый Европейский семинар по структуре биологических макромолекул, Комо, Италия, 1995;
 - Международная школа по прямым методам решения макромолекулярных структур, Эриче, Италия, 1997;
 - Национальная конференция по применению Рентгеновского, Синхротронного излучений, Нейтронов и Электронов для исследования материалов, Дубна, Россия, 1997;
 - 17-ая Европейская кристаллографическая конференция, Лиссабон, Португалия, 1997;
 - 18-ая Европейская кристаллографическая конференция, Прага, Чехия, 1998,
- а также на ряде семинаров в отечественных и зарубежных лабораториях.

По материалам диссертации опубликовано 18 работ, список которых приведен в конце автореферата.

Работа по развитию изложенных в диссертации методов выполнялась в сотрудничестве с Лабораторией биологических структур Института генетики и молекулярной и клеточной биологии (Страсбург, Франция) и Лабораторией кристаллографии и моделирования минералов и биологических структур Университета Нанси (Франция). Работы по применению разработанной методики к определению структуры рибосомальной частицы T50S велись с использованием экспериментальных данных, полученных в группе А. Yonath (Институт им. Макса Планка, Гамбург, Германия) и в тесном сотрудничестве с А. Yonath и I. Agman.

Работа была поддержана грантами РФФИ 94-04-12844 и 97-04-48319.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

1. Введение. *Ab-initio* решение фазовой проблемы при низком разрешении.

Данная работа посвящена описанию методики *ab-initio* решения фазовой проблемы рентгеновского структурного анализа для больших макромолекулярных комплексов. Математическая суть фазовой проблемы состоит в том, что стандартный эксперимент по дифракции рентгеновских лучей (либо нейтронов или электронов) на кристаллических образцах позволяет определить лишь модули

$\{F_{hkl}\}$ комплексных коэффициентов (структурных факторов) в разложении в ряд Фурье функции распределения электронной плотности в исследуемом кристалле:

$$\rho(x, y, z) = \sum_{hkl} F_{hkl} \exp[i\varphi_{hkl}] \exp[-2\pi i(hx + ky + lz)]. \quad (1)$$

При этом значения фаз $\{\varphi_{hkl}\}$ остаются неизвестными, делая невозможным прямой расчет функции распределения электронной плотности. Здесь (x, y, z) - относительные координаты точки в базисе, образованном периодами кристалла, а

(h, k, l) - целочисленные индексы. Теоретически суммирование в (1) должно быть распространено на все целочисленные индексы, однако на практике используется конечный набор структурных факторов. Размеры использованного при расчете (1) набора структурных факторов принято характеризовать разрешением. При этом под разрешением, соответствующим отдельному структурному фактору, понимается длина d "плоской волны", отвечающей соответствующей гармонике Фурье $\exp[-2\pi i(hx + ky + lz)]$, а под разрешением d_{\min} набора структурных факторов - минимальное из разрешений для структурных факторов, входящих в набор.

Дополнительное предположение о том, что $\rho(x, y, z)$ не является произвольной функцией, а может быть представлена как сумма вкладов отдельных атомов, позволяет получить соотношения (алгебраические и вероятностные) между значениями модулей и фаз структурных факторов. Эти соотношения лежат в основе методов расчета значений фаз для низкомолекулярных соединений. В последние годы эти методы были существенно развиты и позволяют в настоящее время при благоприятных обстоятельствах определять структуры небольших белков, содержащих до 1000 атомов в независимой части ячейки. Однако применимость этих методов к макромолекулярным объектам существенно ограничена размерами исследуемого объекта и необходимостью иметь в наличии набор экспериментальных данных очень высокого разрешения (лучше 1.0 Å, что для белков является скорее исключением, чем правилом).

Основными подходами к решению фазовой проблемы в кристаллографии белка является использование изоморфного замещения, аномального рассеяния или известной атомной модели гомологичного белка. В первом из подходов ключом к решению фазовой проблемы являются различия в модулях структурных факторов, отвечающих исходному объекту и полученному путем его химической модификации изоморфному производному. Во втором подходе оказываются существенными различия в величинах модулей специальных пар структурных факторов, которые возникают при наличии в исследуемом объекте аномальных рассеивателей. Оба эти эффекта убывают по мере роста размеров исследуемых объектов, поэтому их применение начинает встречать существенные сложности при попытках определения структуры больших макромолекулярных комплексов (например, таких, как рибосома и составляющие ее субъединицы). Поэтому существенное значение приобретает создание для макромолекулярной кристаллографии *ab-initio* методов решения фазовой проблемы. К этой группе относятся методы, использующие для расчета фаз лишь общие сведения об исследуемых объектах и ограничивающиеся экспериментальными данными, полученными в эксперименте с нативным объектом.

Имеются две стратегии решения фазовой проблемы. Первая из них восходит к области низкомолекулярных соединений. При таком подходе сразу пытаются использовать данные как можно более высокого разрешения. К белкам такой подход начал применяться не так давно. Этот подход содержит ряд ограничений, которые существенно затрудняют его применение для очень крупных структур. Например, сразу требуются данные очень высокого разрешения, что далеко не всегда можно получить в эксперименте. Кроме того, такие методы оказались чувствительны к наличию в белке атомов металлов, и результаты не всегда стабильны.

В излагаемом подходе используется другая стратегия, традиционная именно для белковой кристаллографии - сначала решить фазовую проблему для низкого разрешения, а потом постепенно расширять набор фаз. Под термином “низкое разрешение” здесь понимается такое разрешение, когда в рассмотрение включены 50-100 малоугловых рефлексов. При таком подходе сама величина разрешения зависит от размеров элементарной ячейки. Для белка обычных размеров (с длиной ребер ячейки порядка 70Å) это разрешение порядка 10Å. Для использованного при тестировании методики комплекса т-РНК-синтетазы с тРНК (параметры ячейки порядка 350Å) - это разрешение 40Å. Для кристаллов рибосомальной частицы T50S (длины ребер ячейки порядка 500Å) - это разрешение порядка 60Å. Карты электронной плотности такого разрешения позволяют определить внешние очертания молекулы и ее местоположение в элементарной ячейке (в частности, это может оказаться полезным и при использовании метода молекулярного замещения при решении проблемы трансляции), а соответствующие фазы структурных факторов являются стартовой точкой для дальнейшей процедуры расширения набора фаз.

2. FAM - метод.

Общая схема предлагаемого FAM (Few Atoms Model) метода приведена на Рис. 1. При анализе отдельных этапов работы в качестве иллюстрации приведены результаты тестирования метода на данных рассеяния нейтронов комплексом аспартил - аминоксил - тРНК синтетазы с тРНК (Moras et al., 1983), далее AspSR-комплекс. Этот объект был кристаллизован в кубической пространственной группе I432 с параметрами ячейки $a=b=c=354\text{\AA}$, $\alpha=\beta=\gamma=90^\circ$. Пространственная структура комплекса была определена ранее (Urzhumtsev et al., 1994) методом молекулярного замещения по данным рентгеновского рассеяния разрешения 8.0Å с использованием атомной модели, отвечающей другой кристаллической форме (P2₁2₁2, Ruff et al., 1991). Значения фаз структурных факторов, рассчитанные по известной модели, привлекались как контрольные значения для оценки результатов тестов, но никак не использовались в процедуре решения фазовой проблемы. Тесты, проведенные с другими объектами, дали аналогичные результаты.

2.1. Основная гипотеза.

В основе предлагаемого метода лежит гипотеза о том, что при низком разрешении распределение электронной плотности можно достаточно хорошо приблизить суммой вкладов небольшого числа очень тяжелых псевдоатомов (глобул). Или, соответственно, структурные факторы при низком разрешении (и, в частности, их фазы) можно приблизить структурными факторами, рассчитанными по модели, состоящей из небольшого числа псевдоатомов. Тесты с известными структурами подтверждают это предположение. Иногда достаточно даже одного псевдоатома, чтобы с удовлетворительной точностью приблизить полтора-два десятка низкоугловых фаз (Andersson & Novmöller, 1996). Так, из рис. 2 и 4 следует, что при удачно выбранных положениях одной-двух глобул можно получить фазы, дающие коэффициент корреляции с точными фазами порядка 80%. Проблема, однако, состоит в том, чтобы найти такие положения центров глобул, которые приводят к правильным фазам.

Популярный подход к решению этой проблемы заключается в том, чтобы перебрать (систематически или случайно) различные варианты взаимного расположения глобул в кристаллической ячейке и выбрать такое, при котором

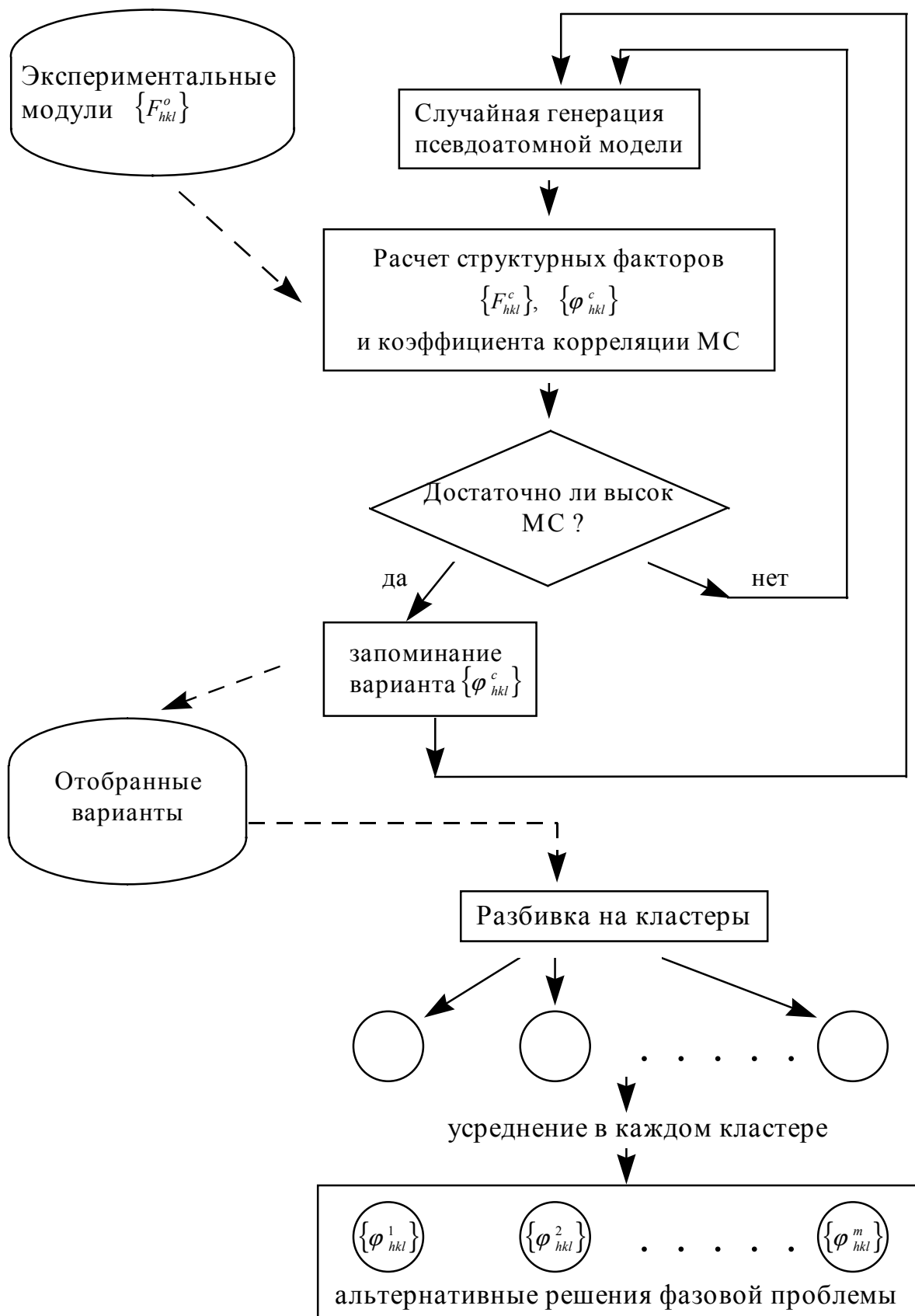


Рис.1. Общая схема FAM-метода.

модули структурных факторов, рассчитанных по такой модели, наиболее близки к экспериментальным. Однако проверка на известных структурах показывает, что столь прямолинейный подход может заканчиваться неудачей.

На рис. 2 показаны результаты 1000 случайных генераций координат центра глобулы, моделирующей AspSR-комплекс. Для каждого положения центра рассчитывались соответствующие величины структурных факторов и вычислялись корреляция модулей (MC) и корреляция фаз (PC) по формулам:

$$MC = \frac{\sum_{hkl} (F_{hkl}^c - \langle F^c \rangle)(F_{hkl}^o - \langle F^o \rangle)}{\sqrt{\sum_{hkl} (F_{hkl}^c - \langle F^c \rangle)^2 \sum_{hkl} (F_{hkl}^o - \langle F^o \rangle)^2}} \quad (2)$$

$$PC = \frac{\sum_{hkl} (F_{hkl}^o)^2 \cos(\varphi_{hkl}^{true} - \varphi_{hkl}^c)}{\sum_{hkl} (F_{hkl}^o)^2} \quad (3)$$

В расчет было включено 49 независимых рефлексов зоны разрешения до 40Å. Каждая точка на диаграмме представляет одну модель. Величина корреляция модулей отложена по горизонтали, фаз - по вертикали. Из рисунка видно, что:

- положение глобулы, при котором достигается наилучшая корреляция модулей, дает совсем плохие фазы;
- положению, приводящему к с самым хорошим фазам, соответствует достаточно низкий коэффициент корреляции модулей.

Ситуация не меняется, если увеличить число атомов в модели. Более того, уточняя координаты этих моделей в обратном пространстве, можно добиться, что коэффициент корреляции модулей станет близким к 1 (а значение стандартного кристаллографического R-фактора - порядка 1%), но точность фаз при этом практически не изменится. Стрелки на рис.3 показывают, как меняется положение точки на диаграмме в процессе уточнения координат атомов.

Тесты показали, что картина существенно не меняется и при использовании вместо коэффициента корреляции других критериев сравнения рассчитанных модулей с экспериментальными. Не меняется картина и при использовании при низком разрешении других критериев отбора, таких, как гистограммы электронной плотности (Lunin et al., 1990), связность или локальная вариация плотности. Более существенным оказался другой параметр - количество атомов в модели. Здесь тестирование показало, что на начальном этапе наилучшая ситуация возникает при использовании 2-4 псевдоатомов (глобул) в модели. Именно при таком их количестве в облаке на MC-PC-диаграмме появляется область, соответствующая высоким значениям как одного критерия (MC), так и другого (PC) (Рис.4), аналогично тому, как это имело место при использовании гистограмм электронной плотности (Lunin et al., 1990).

Проведенное тестирование позволило расширить основную используемую гипотезу. Положение о том, что при низком разрешении структурные факторы могут быть приближенно рассчитаны по модели, состоящей из небольшого числа глобул, было дополнено предположением о характере зависимости точности значений фаз, рассчитанных по такой модели (PC), от степени соответствия

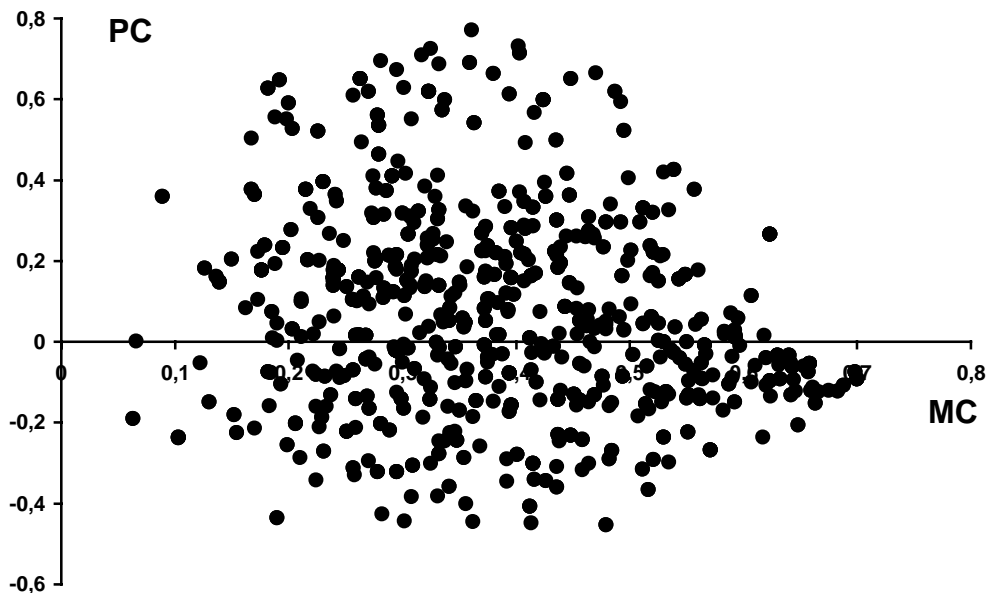


Рис. 2. Значения коэффициентов корреляции модулей (MC) и фаз (PC), рассчитанных по случайно сгенерированным сферическим глобулам по отношению к модулям и фазам AspSR-комплекса. Каждая точка представляет одну модель.

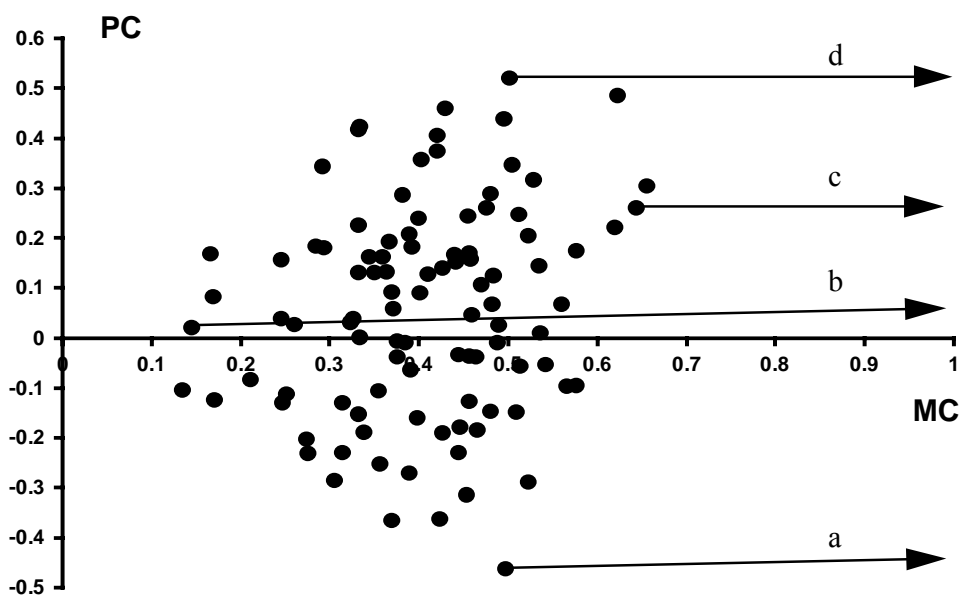


Рис. 3. Значения коэффициентов корреляции модулей (MC) и фаз (PC), рассчитанных по случайно сгенерированным 100-атомным моделям по отношению к модулям и фазам AspSR-комплекса. Каждая точка представляет одну модель. Стрелки показывают результат кристаллографического уточнения в обратном пространстве некоторых моделей.

рассчитанных модулей структурных факторов экспериментально измеренным значениям (МС):

- наилучшее соответствие модулей может достигаться при неправильных фазах;
- модель, воспроизводящая наилучшим образом фазы, может не выделяться среди других по точности воспроизведения модулей;
- локальное уточнение параметров модели (координат центров глобул) существенно повышает качество воспроизведения модулей, не улучшая точность воспроизведения фаз структурных факторов;
- среди моделей, достаточно хорошо воспроизводящих модули структурных факторов, присутствует группа близких между собой моделей, воспроизводящих фазы достаточно хорошо.

Такая ситуация может быть проиллюстрирована схематически на рис.5. Мы предполагаем, что функция, выражающая зависимость корреляции модулей от качества модели, имеет небольшое число широких основных максимумом и много высоких, но узких побочных максимумов. При этом наилучшая модель близка к одному из основных широких максимумов, хотя и не обязательно самому высокому.

2.2. Первый этап. Генерация и отбор фазовых наборов.

Принципиальной особенностью предлагаемой методики является то, что мы отказываемся как от поиска модели, дающей наилучшее соответствие модулей, так и от претензий найти модель, наилучшим образом воспроизводящую фазы. Вместо этого на первом этапе работы случайно генерируется большое число псевдоатомных моделей, и из них отбираются все модели, достаточно хорошо воспроизводящие экспериментально полученные значения модулей структурных факторов. Наборы фаз, рассчитанных по отобранным моделям, запоминаются для дальнейшего анализа. Каждый такой набор значений фаз именуется далее вариантом (решения фазовой проблемы).

Случайная генерация вариантов позволяет отчасти уменьшить риск попадания в один из узких побочных максимумов, однако при таком отборе среди отобранных есть варианты как с хорошими фазами, так и с плохими. Более того, варианты с наилучшими фазами не всегда попадают в отобранное множество. Тем не менее, концентрация вариантов с хорошими значениями фаз в отобранном наборе становится выше, чем среди всех сгенерированных наборов. Происходит некоторое “обогащение набора“. Отбор по наилучшему значению коэффициента корреляции модулей не позволяет однозначно определить правильное решение, но все же увеличивает вероятность встретить хороший вариант среди отобранных.

2.3. Второй этап. Кластерный анализ.

Вторым этапом в применении FAM-метода является анализ множества отобранных на первом этапе вариантов. Целью анализа является выделение групп близких между собой наборов фаз (кластеров) и отсеивание вариантов, отвечающих побочным максимумам критерия отбора. Для осуществления этой задачи привлекаются методы кластерного анализа в метрическом пространстве (Дюран & Оделл, 1977). Эти методы позволяют выделить различные типы расположения точек в пространстве: одна компактная группа; группа и случайные выбросы; две компактные группы; случайные точки и т.п. Эти различные типы расположения точек могут быть проиллюстрированы с помощью кластерного дерева (дендрограммы), которое отражает процесс постепенного “слипания“ точек в кластеры, начиная с самых близких. Для построения такого дерева нужно уметь

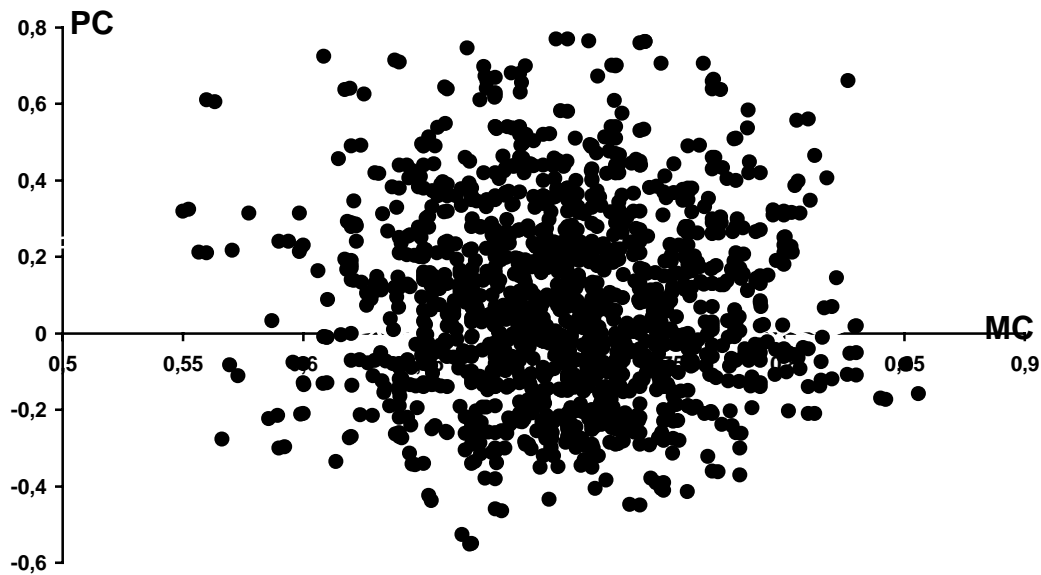


Рис. 4. Значения коэффициентов корреляции модулей (MC) и фаз (PC), рассчитанных по случайно сгенерированным 2-атомным моделям по отношению к модулям и фазам AspSR-комплекса. Каждая точка представляет одну модель.

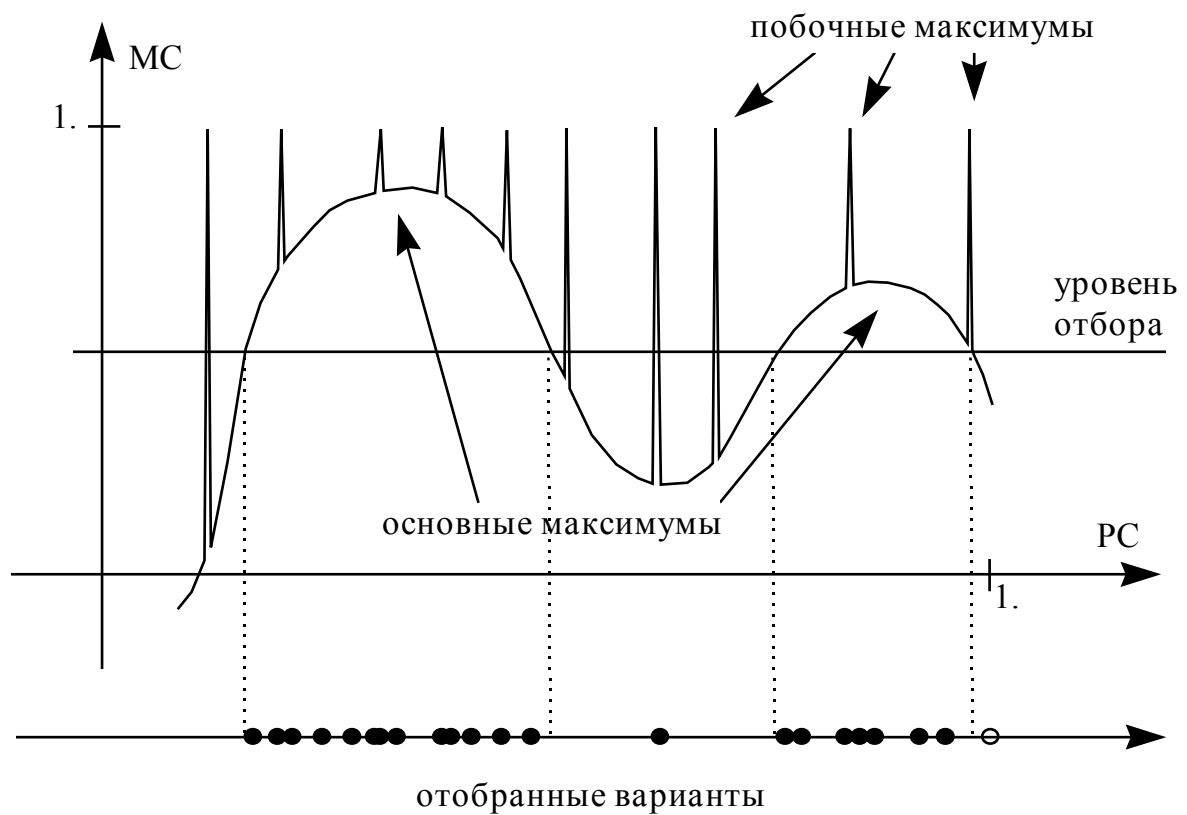


Рис.5. Схема первого этапа FAM - метода.

определять расстояние между двумя точками (в нашем случае - между двумя наборами фаз) и между группами точек. Мы определяем расстояние между двумя наборами фаз (между двумя вариантами $\{\varphi_{hkl}^j\}$ и $\{\varphi_{hkl}^m\}$) следующим образом:

- каждому набору фаз ставится в соответствие распределение электронной плотности $\rho^j = \rho^j(x, y, z)$, вычисленное с использованием данных фаз и экспериментальных модулей по формуле (1);
- два сравниваемых распределения выравниваются посредством допустимых для присущей изучаемому объекту группы кристаллографических симметрий сдвигов начала координат (и выбора энантиоформа, если это допустимо);
- для выровненных распределений вычисляется расстояние

$$\begin{aligned} dist(\{\varphi_{hkl}^j\}, \{\varphi_{hkl}^m\}) &= dist(\rho^j, \rho^m) \\ &= \sqrt{\frac{\int_V [\rho^j(x, y, z) - \rho^m(x, y, z)]^2 dV}{\int_V [\rho^j(x, y, z)]^2 dV \int_V [\rho^m(x, y, z)]^2 dV}} \end{aligned} \quad (4)$$

Расстояние между двумя наборами вариантов может быть определено, например, как среднее значение попарных расстояний при выборе одного варианта из первого набора, а другого из второго. Заметим, что определение этих расстояний и построение кластерного дерева не требует какой-либо информации об истинных значениях фаз структурных факторов исследуемого объекта.

Следует подчеркнуть, что перед вычислением расстояния распределения электронной плотности должны быть выровнены. Это связано с тем, что произвольный сдвиг начала координат не меняет значения модулей, но меняет значения фаз структурных факторов. Поэтому в процессе *ab-initio* решения фазовой проблемы можно получить два различающихся на первый взгляд набора фаз, отвечающих в действительности одному и тому же образу объекта, но по-разному сдвинутому. Наличие кристаллографической симметрии ограничивает набор возможных сдвигов, делая его специфичным для данной пространственной группы симметрий. В рамках FAM-метода была разработана специальная процедура оптимального выравнивания распределений электронной плотности, основанная на алгоритме быстрого преобразования Фурье.

На рис. 6 приведен пример такого дерева, полученного в ходе тестов с AspSR - комплексом. Было сгенерировано 1000 случайных двухатомных моделей. Из них отобрано 100 с наилучшей корреляцией модулей, и для соответствующих 100 фазовых наборов построено кластерное дерево. Каждый узел отвечает слиянию двух кластеров. По вертикальной оси отложено нормированное расстояние между кластерами в момент их слияния.

Для анализа кластерного дерева и выбора перспективных кластеров была создана специальная интерактивная компьютерная программа, позволяющая выделять кластеры для дальнейшего сравнения и отбора.

2.4. Третий этап. Обработка кластера.

Для каждого выделенного кластера фазовых наборов можно рассчитать среднее значение каждой фазы и соответствующего показателя ее достоверности по формуле:

$$m_{hkl} \exp[i\varphi_{hkl}^{best}] = \frac{1}{N} \sum_{m=1}^N \exp[i\varphi_{hkl}^m], \quad (5)$$

где N - номер варианта в кластере; усреднение производится по всем N предварительно выровненным фазовым наборам из рассматриваемого кластера. Это определение аналогично определению центроидных фаз и показателей достоверности в методе изоморфного замещения (Blow & Crick, 1959).

Помимо центроидных фаз, для каждого кластера определяется соответствующая ему маска области молекулы как область наибольших значений взвешенного синтеза Фурье $\{m_{\mathbf{h}}, F_{\mathbf{h}}^{obs}, \varphi_{\mathbf{h}}^{best}\}$.

Таким образом, результатом однократного применения FAM - процедуры является небольшой набор альтернативных решений фазовой проблемы (центроидных фаз, отвечающих выделенным кластерам) или небольшой набор альтернативных масок области молекулы. Следующим шагом на пути определения структуры исследуемого объекта является выбор одной из этих альтернатив.

2.4. Проблема выбора наилучшего кластера.

FAM - метод позволяет продвинуться от полной неопределенности в значениях фаз до проблемы выбора одного из небольшого числа альтернативных вариантов. Для решения проблемы этого выбора могут быть применены разные подходы. При этом, если различные критерии отбора указывают в качестве лучшего один и тот же кластер, это позволяет надеяться, что отбор отражает некоторые объективные характеристики. Если различные критерии отбора указывают на разные кластеры, то целесообразно попытаться сменить параметры FAM - процедуры (способ генерации моделей, количество атомов, их размер и т.п.).

2.4.1. Визуальный анализ.

Небольшое число синтезов Фурье, соответствующих каждому из альтернативных решений, можно проанализировать визуально. Такой анализ позволяет исключить некоторые заведомо непригодные варианты (типа “уранового решения”), но страдает субъективностью при анализе незначительно различающихся правдоподобных решений.

2.4.2. Анализ плотности на осях вращения.

Наличие высокой плотности на кристаллографических осях симметрии в ситуации, когда исследуемый объект сам по себе такой симметрией не обладает, служит обычно признаком наличия существенных ошибок в значениях фаз и позволяет отбрасывать ложные решения (Urzhumtsev *et al.*, 1996).

2.4.3. Топологические свойства (связность).

В основе этого критерия лежит предположение о том, что на синтезе низкого разрешения, рассчитанном с правильными фазами, должно выделяться столько связных компонент, какое число симметрично связанных экземпляров объекта находится в элементарной ячейке кристалла. Этот анализ осуществляется

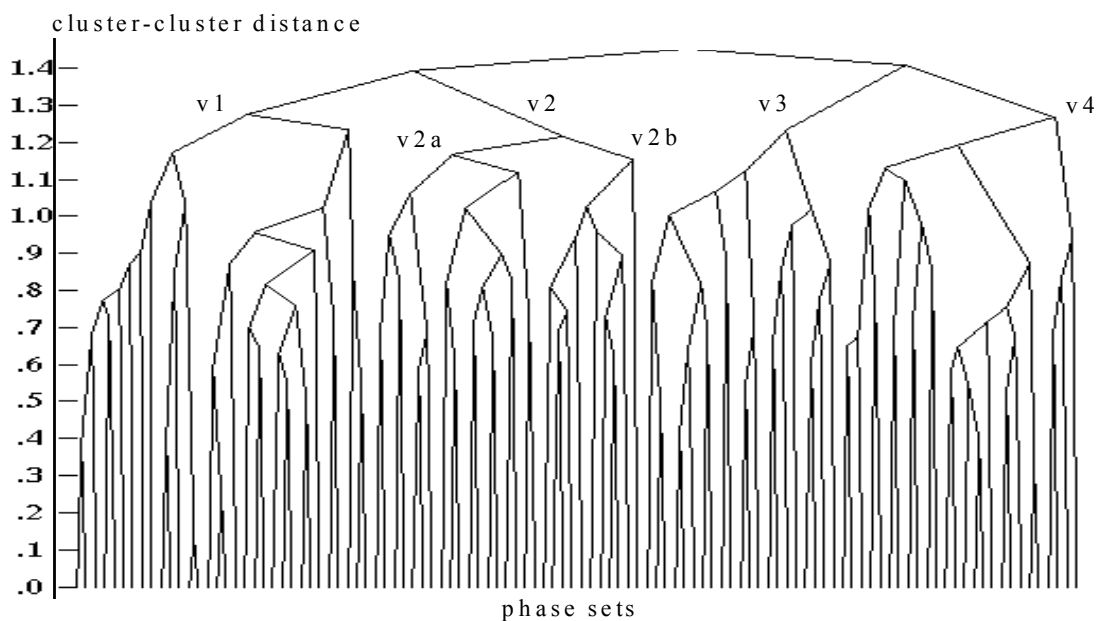


Рис. 6. Пример кластерного дерева, полученного в процессе тестовых расчетов с AspSR-комплексом.

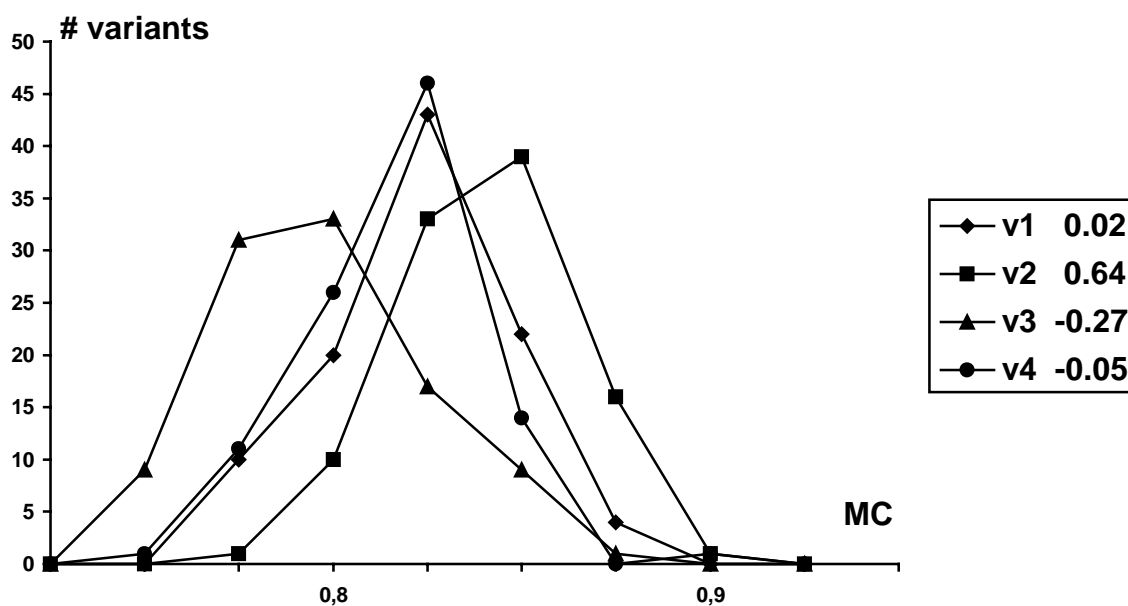


Рис. 7. Распределение псевдоатомных моделей по величине коэффициента корреляции модулей структурных факторов при генерации 100 атомов в альтернативные области элементарной ячейки (v1-v4). Цифры в рамке показывают корреляцию фаз, соответствующих этим областям с точными фазами для AspSR - комплекса.

специально созданной компьютерной программой, производящей подсчет числа компонент связности в синтезах Фурье, рассчитанных с пробными значениями фаз.

2.4.4. Монте-Карловский тест (правдоподобие).

Для решения проблемы выбора из двух (или более) альтернативных масок области был разработан тест, который может рассматриваться как модификация принципа максимального правдоподобия. Основная идея теста: при случайной генерации координат атомов имеется больше шансов воспроизвести экспериментальные значения модулей структурных факторов в том случае, когда атомы бросаются в область, в действительности занятую молекулой, чем при их генерации в произвольно выбранную область. Формально процедура состоит в следующем. Для каждого выделенного кластера:

- рассчитывается синтез Фурье, и на его основе строится маска области молекулы;
- в эту область молекулы производится генерация случайных моделей, состоящих из большого числа псевдоатомов (например, 100). Вычисляется гистограмма распределения коэффициента корреляции рассчитанных модулей с экспериментальными для сгенерированных моделей.

При этом ожидается иметь больший процент высоких значений корреляции при генерации “в правильную маску”.

На рис. 7 показаны гистограммы, полученные при анализе четырех кластеров, отобранных при анализе дендрограммы, приведенной на Рис.6. Каждая кривая отвечает анализу одного кластера. По горизонтальной оси отложены значения корреляции модулей, получающихся при бросании 100-атомных моделей в маску, определенную кластером. По оси ординат - количество вариантов с данной корреляцией модулей. Цифры в легенде показывают корреляцию усредненных фаз с точными фазами для AspSR-комплекса. Видно, что правее всех располагается кривая для кластера v_2 , который обладает и максимальной корреляцией с правильными фазами.

При проведении тестирования кластеров использовалась R-free идеология, предложенная А.Брюнгером (Brunger, 1992): вовлеченные в работу рефлекс делились на две группы - рабочие и контрольные. В процессе генерации и отбора вариантов для вычисления значений критерия отбора использовались лишь модули рабочих рефлексов, в то время как при проведении стандартного теста использовался лишь контрольный набор рефлексов.

2.5. Уточнение значений фаз и повышение разрешения.

На самом первом этапе применения FAM - метода псевдоатомы бросаются случайно во всю элементарную ячейку, После того, как получены некоторые приближенные значения фаз, эту информацию можно вовлечь в процесс генерации моделей, разрешая теперь псевдоатомам находится лишь внутри определенной ранее области молекулы. Разработанное программное обеспечение позволяет вводить при этом дополнительные стереохимические ограничения (ограничить минимально возможное расстояние между псевдоатомами, генерировать их цепочки и т.п.). За процедурой генерации вновь следует отбор вариантов, построение кластерного дерева, выбор кластеров, усреднение вариантов в кластерах и выбор наилучшего из них.

В описанном процессе итеративного уточнения значений фаз число вовлеченных в работу рефлексов может постепенно увеличиваться, позволяя получать более детальные маски области молекулы.

На рис. 8 показан результат применения описанной методики к известной структуре AspSR - комплекса. Здесь приведены сечения независимой части элементарной ячейки и показаны линии уровня для синтеза электронной плотности, рассчитанного с фазами, определенными с помощью FAM - метода. Точки на рисунке показывают действительные положения C_{α} -атомов в молекуле синтазы и атомов фосфора в молекулах тРНК.

3. Определение структуры рибосомной частицы 50S из *Thermus thermophilus* при разрешении 40 Å

Разработанная методика была применена к *ab-initio* определению структуры рибосомальной частицы T50S из *Thermus thermophilus* по данным рентгеновского рассеяния. Знание структуры рибосомы и ее частей играет важную роль в понимании процессов биосинтеза белка (Спирин, 1986), что и объясняет значительные усилия, затраченные в последние десятилетия на решения этой проблемы. Основная структурная информация о рибосоме и ее частях была получена до настоящего времени методами электронной микроскопии (Vasiliev, 1971), трехмерной реконструкции крио- электронно -микроскопических (ЭМ) - изображений (Frank *et al.*, 1995; Stark *et al.*, 1995) либо путем уточнения ЭМ - моделей по данным малоуглового рассеяния нейтронов (Svergun *et al.*, 1997; Фан Ликсин, 1998) или использованием ЭМ - модели совместно с данными рентгеновского рассеяния (Van *et al.*, 1998). В данной работе была предпринята попытка получения структурной информации, независимой от имеющихся электронно - микроскопических моделей, опирающейся лишь на данные рентгеновского рассеяния кристаллами T50S частицы.

В работе использовались экспериментальные данные, полученные в лаборатории А. Yonath (Volkman *et al.*, 1990). Кристаллы T50S имели симметрию пространственной группы $P4_12_12 / P4_32_12$ и размеры ячейки $a=b=498\text{Å}$, $c=198\text{Å}$. Первоначально FAM - процедура позволила получить сильно сглаженный образ частицы при разрешении порядка 70 Å (Urzhumtsev, Vernoslova & Podjarny, 1996). Дальнейшее развитие методики - разработка процедуры уточнения маски области, привлечение для контроля за ходом работы соображений связности образа, разработка Монте-Карловской процедуры тестирования, а также использование методов модификации электронной плотности (Podjarny, Rees & Urzhumtsev, 1996) позволило повысить разрешение и получить более детальный образ исследуемой частицы при разрешении порядка 40 Å.

После того, как были получены указанные результаты, в распоряжение автора поступили компьютерные данные об ЭМ - модели T50S частицы (Stark *et al.*, 1995), что позволило произвести независимое сравнение полученных результатов. Для этих целей ЭМ - модель была размещена в элементарной ячейке кристалла в положении и ориентации, наиболее хорошо отвечающих полученному *ab-initio* синтезу Фурье (совмещение осуществлялось специально созданной для этой цели программой вписывания модели в плотность в прямом пространстве). По размещенной таким образом модели были рассчитаны фазы структурных факторов, которые совместно с экспериментальными величинами модулей были использованы для расчета контрольного синтеза Фурье. На рис.9. представлены в нескольких ракурсах изображения субъединицы T50S, найденные *ab-initio* по рентгеновским данным, и полученные с использованием ЭМ-модели. Следует отметить, что данные рентгеновского рассеяния не позволяют сделать выбор между группами симметрий $P4_12_12$ и $P4_32_12$. Поэтому *ab-initio* решение фазовой проблемы имеет

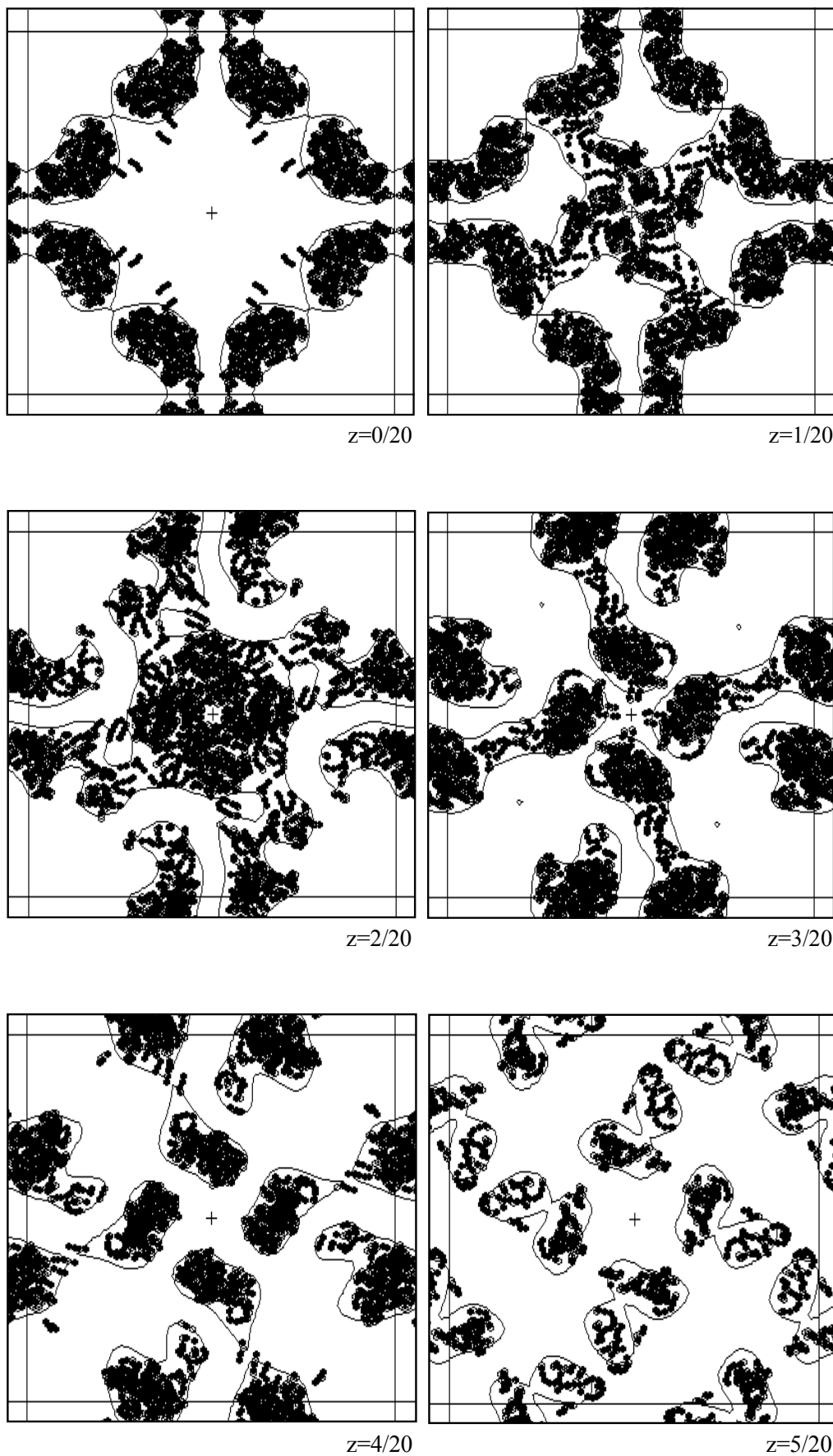


Рис. 8. Границы комплекса аминоксил-тРНК-синтетазы с тРНК определенные FАМ-методом и реальные позиции C_{α} и Р атомов в ячейке кристалла.

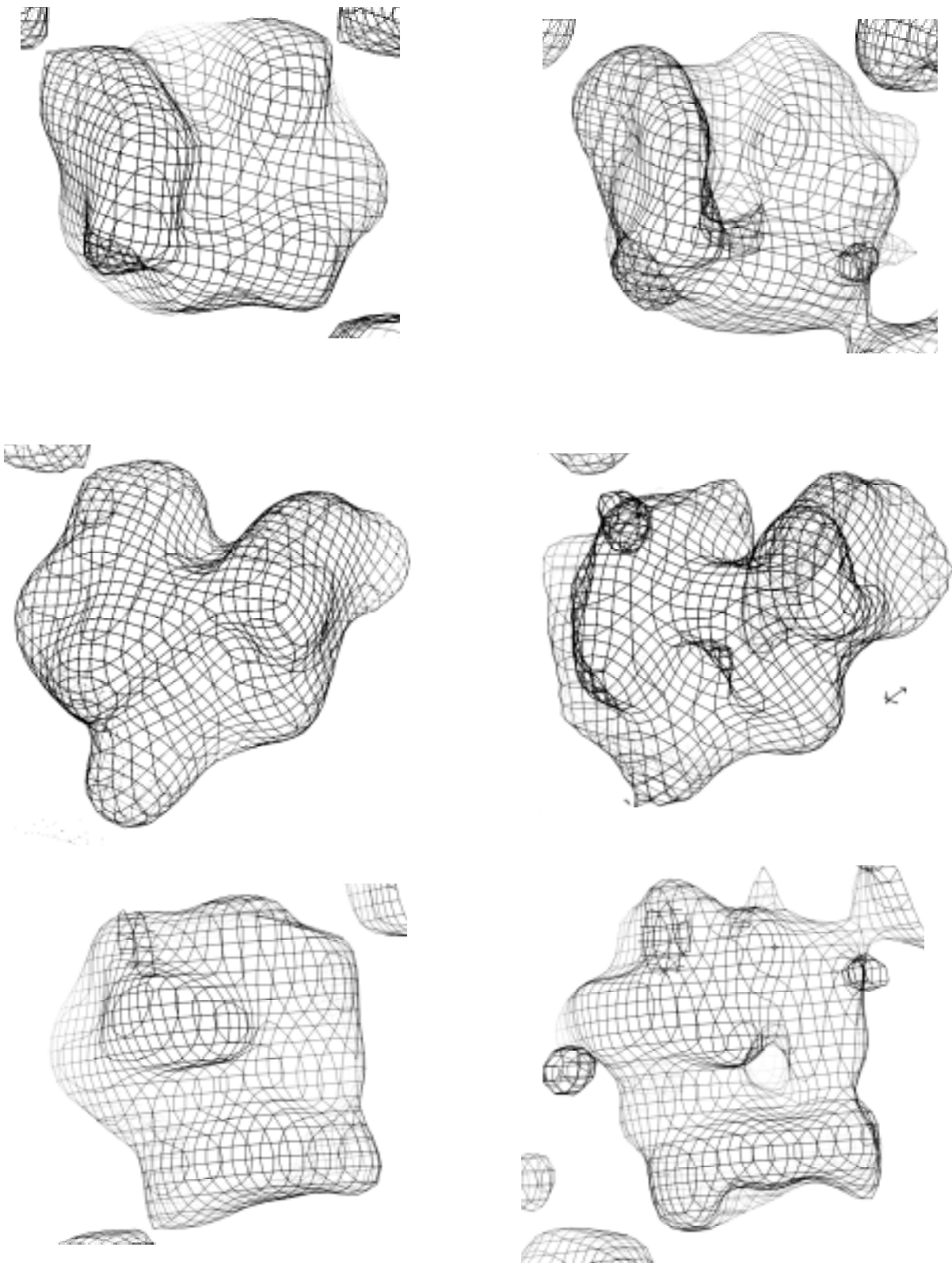


Рис. 9. Синтезы Фурье для T50S субъединицы, рассчитанные с экспериментальными модулями и фазами: слева - определенными FAM - методом; справа - рассчитанными по электронно-микроскопической модели.

принципиальную неоднозначность в выборе энантиомера, которая должна быть позднее разрешена из других соображений (например, стереохимических).

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Продемонстрировано, что применяемые в кристаллографии методы контроля точности определения фаз при низком разрешении имеют статистический характер: наилучшие значения контрольных критериев не соответствуют в точности наилучшим наборам фаз, однако имеется корреляция между значениями рассмотренных критериев качества и действительной точностью значений фаз.
2. Разработана и реализована компьютерная процедура Монте-Карловского типа для просмотра всех потенциально возможных решений фазовой проблемы при низком разрешении и их фильтрации в соответствии с выбранными критериями отбора.
3. Предложена базирующаяся на методах кластерного анализа процедура выделения из отфильтрованных на предыдущем этапе работы вариантов небольшого числа кластеров близких решений, дающих альтернативные решения фазовой проблемы.
4. Предложены и реализованы методы сравнительного тестирования различных вариантов решения фазовой проблемы, основанные на максимизации обобщенного правдоподобия и использовании топологических характеристик изучаемого объекта.
5. Проведено тестирование разработанной методики на данных нейтронного рассеяния для комплекса тРНК синтетазы с тРНК.
6. Разработанная методика применена к определению структуры рибосомальной частицы T50S из *Thermus thermophilus* при разрешении 40Å.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

- Дюран, Б. & Оделл, П. (1977). *Кластерный анализ*. М. Статистика.
- Спирин, А.С. (1986). *Молекулярная биология. Структура рибосомы и биосинтез белка*. М. Высшая школа.
- Фан Ликсин. (1998). Диссертация на соискание ученой степени к.ф.-м.н., ИТЭБ РАН, Пущино.
- Andersson, K.M. & Hovmöller, S. (1996). *Acta Cryst.* D**52**, 1174-1180.
- Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R.A., Sweet, R., Frank, J., Moore, P.B. & Steitz, T.A. (1998). *Cell*, **93**, 1105-1116.
- Blow, D.M. & Crick, F.H.C. (1959). *Acta Cryst.* **12**, 794-802.
- Brünger, A. (1992). *Nature (London)*, **355**, 472-474.
- Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Rademacher, M., Grassucci, R., Lata R.K. & Agarwal, R.K. (1995). *Nature* **376**, 441-444.
- Lunin, V.Yu., Urzhumtsev, A.G. & Skovoroda, T.P. (1990). *Acta Cryst.* A**46**, 540-544.
- Moras, D., Lorber, D., Romby, P., Ebel, J.-P., Giegé, R., Levitt-Bentley, A. & Roth, M. (1983). *J.Biomol. Struct. Dynam.* **1**, 209-223.
- Podjarny, A.D., Rees, B. & Urzhumtsev, A (1996). In *Methods in Molecular Biology* **56**, 205-226.
- Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.-C. & Moras, D. (1991). *Science*, **252**, 1682-1689.

- Stark, H., Mueller, F., Orlova, E.V., Schatz, M., Dube, P., Erdemir, T., Zemlin, F., Brimacombe, R. & van Heel, M. (1995). *Structure* **3**, 815-821.
- Svergun, D.I., Burkhardt, N., Pedersen, J.S., Koch, M.H.J., Volkov, V.V., Kozin, M.B., Meerwink, W., Stuhrmann, H.B., Diedrich, G. & Nierhaus, K.H. (1997). *J.Mol.Biol.* **271**, 602-618.
- Urzhumtsev, A.G., Podjarny, A.D. & Navaza, J. (1994). *Jnt. CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **30**, 29-36.
- Urzhumtsev, A.G., Vernoslova, E.A. & Podjarny, A.D. (1996). *Acta Cryst.* **D52**, 1092-1097.
- Vasiliev, V.D. (1971). *FEBS Lett.* **14**, 203-205.
- Volkman, N., Hottentrotter, S., Hansen, H.A., Zaytsev-Bashan, A., Sharon, R., Yonath, A. & Wittmann, H.G. (1990). *J.Mol.Biol.* **216**, 239-241.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Lunin, V.Yu., **Lunina, N.L.**, Petrova, T.E., Vernoslova, E.A., Urzhumtsev, A.G., Podjarny, A.D. (1994) "On the ab-initio solution of the phase problem for macromolecules at very low resolution". *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **30**, 37-44
2. Urzhumtsev, A.G., Lunin, V.Yu., Vernoslova, E.A., **Lunina, N.L.**, Petrova, T.E., Navaza, J., Podjarny, A.D. (1994) "Very low resolution phasing with simple models". *15th Eur. Cryst. Meeting, Dresden, Germany, 28.8-2.9/1994, Coll.Abstr., R.Oldenbourg Verlag, Munchen*, 217
3. Lunin, V.Yu., **Lunina, N.L.**, Petrova, T.E., Vernoslova, E.A., Urzhumtsev, A.G., Podjarny, A.D. (1994) "A Monte Carlo Approach to low resolution phasing in protein crystallography". *ACA Annual Meeting, June 25 - July 1, 1994, INFORUM, Atlanta, Georgia, Atlanta Convention Center*, 178
4. Lunin, V.Yu., **Lunina, N.L.**, Petrova, T.E., Vernoslova, E.A., Urzhumtsev, A.G., Podjarny, A.D. (1995) "On the ab-initio Solution of the Phase Problem for Macromolecules at Very Low Resolution: the Few Atoms Model Method". *Acta Cryst.*, **D51**, 896-903.
5. Lunin, V.Yu., Urzhumtsev, A.G., Vernoslova, E.A., **Lunina, N.L.**, Petrova, T.E., Podjarny, A.D. (1995) "La methode FAM pour le phasage ab-initio : Theorie.". *AFC95, 24-27 Jan. 1995, Grenoble*, P07a.
6. Lunin, V.Yu., **Lunina, N.L.**, Petrova, T.E., Vernoslova, E.A., Urzhumtsev, A.G., Podjarny, A.D. (1995) "On the ab-initio solution of the phase problem for macromolecules at very low resolution: the Few Atoms Models method". *4th European Workshop on Crystallography of Biological Macromolecules, Como (Italy), May 21-25, O19*.
7. Lunin, V.Yu. & **Lunina, N.L.** (1996) "The Map Correlation Coefficient for Optimally Superposed Maps". *Acta Cryst.* **A52**, 365-368.
8. Lunin V.Yu., **Lunina N.L.**, Petrova T.E., Urzhumtsev A.G. & Podjarny A.D. (1997) "A generalized likelihood and the phase cluster choice problem". In "Direct methods for solving macromolecular structures", Lecture Notes, Erice, Italy, May 22 - June 02 1997, 485.
9. **Лунина Н.Л.**, Лунин В.Ю. & Уржумцев А.Г. (1997) ""Выращивание плотности" - новый метод локального улучшения синтезов Фурье". Национальная конференция по применению Рентгеновского, Синхротронного излучений,

- Нейтронов и Электронов для исследования материалов, Дубна, 25-29 мая 1997, 222.
10. **Lunina, N.L.**, Urzhumtsev, A. (1997) "Density growing : a new approach to the local improvement of electron density maps". *XVIIth European Cryst.Meeting, 24-28 August 1997, Lisbon, Portugal*, 92.
 11. Lunin, V.Yu., **Lunina, N.L.**, Petrova, T.E., Urzhumtsev, A.G., Podjarny, A.D. (1998) "On the *ab initio* solution of the phase problem for macromolecules at very low resolution. II. Generalized Likelihood Based Approach to the Cluster Discrimination.". *Acta Cryst.*, **D53**, 726-734
 12. Lunin, V.Y., **Lunina, N.L.**, Petrova, T.E., Urzhumtsev, A.G., Podjarny, A.D. (1998) "Very low resolution ab-initio phasing. Problems and advances". *ECM-18 Abstracts, XVIIIth European Cryst.Meeting, 16-20 August 1998, Prague, Republic Czech, E5-P7, Bulletin of the Czech and Slovak Crystallographic Association*, 5A, 131-132
 13. Urzhumtsev, A.G., Lunin, V.Y., **Lunina, N.L.** (1998) "Seminvariant density decomposition and connectivity analysis in the low resolution macromolecular phasing". *ECM-18 Abstracts, XVIIIth European Cryst.Meeting, 16-20 August 1998, Prague, Republic Czech, E5-O1, Bulletin of the Czech and Slovak Crystallographic Association*, 5B, 482-483
 14. **Lunina, N.L.** (1998) "Software for very low resolution ab-initio phasing". *ECM-18 Abstracts, XVIIIth European Cryst.Meeting, 16-20 August 1998, Prague, Republic Czech, E5-O1, Bulletin of the Czech and Slovak Crystallographic Association*, 5 B, 480-481
 15. Podjarny, A.D., Urzhumtsev, A.G., Vernoslova, E.A., Petrova, T., **Lunina, N.**, Lunin, V. (1998) "Recent advances on the application of the Few Atom Model method to the T50S ribosomal particle". *AFC98, 24-27 Feb. 1998, Orlean, BSG-A02*
 16. Podjarny, A.D., **Lunina, N.L.**, Urzhumtsev, A.G., Vernoslova, E.A., Petrova, T.E., Lunin, V.Y. (1998) "Ab-initio phasing at 30 Å of the t50S ribosomal particle with the Few Atoms Model method". *ACA Annual Meeting, in press*
 17. Петрова, Т.Е., Лунин, В.Ю., **Лунина, Н.Л.**, Скворода, Т.П. (1998). "Выбор априорного распределения координат атомов для макромолекулярных структур на базе принципа максимального правдоподобия". *Биофизика. Принята к печати.*
 18. Lunin, V.Yu., **Lunina, N.L.**, Urzhumtsev, A.G. (1998) "Seminvariant density decomposition and connectivity analysis in very low resolution macromolecular phasing". *Acta Cryst.*, **A**, *submitted.*

Научное издание

Автореферат Луниной Н.Л.

Налоговая льгота – общероссийский классификатор продукции
ОК-005-93, том 2; 953000 – книги и брошюры.

21.12.98 г. З. 8202Р. Т. 100 экз. Усл.печ.л. 1,25.

Отпечатано с оригинала-макета в Отделе научно-технической информации
Пушкинского научного центра РАН.

142292, г.Пушино Московской обл., проспект Науки, 3. ОНТИ ПНЦ РАН.