ИНСТИТУТ МАТЕМАТИЧЕСКИХ ПРОБЛЕМ БИОЛОГИИ РОССИЙСКАЯ АКАДЕМИЯ НАУК

На правах рукописи УДК 577.32

ЛУНИНА Наталья Леонидовна

ВЫЧИСЛИТЕЛЬНЫЕ ПОДХОДЫ К РЕШЕНИЮ ФАЗОВОЙ ПРОБЛЕМЫ КРИСТАЛЛОГРАФИИ БИОПОЛИМЕРОВ ПРИ НИЗКОМ РАЗРЕШЕНИИ

03.00.2 - Биофизика

Диссертация на соискание ученой степени кандидата физико-математических наук

> Научный руководитель кандидат физико-математических наук, А.Г.Уржумцев

Пущино 1998

СОДЕРЖАНИЕ

	стр
ВВЕДЕНИЕ	5
ГЛАВА І. ФАЗОВАЯ ПРОБЛЕМА ПРИ НИЗКОМ	
РАЗРЕШЕНИИ. (ЛИТЕРАТУРНЫЙ ОБЗОР)	10
1. Зачем нужны фазы рефлексов низкого разрешения	10
2. Восстановление структурных факторов низкого	
разрешения	12
2.1. Определение фаз методом максимального	
детерминанта	12
2.2. Двухшаговая процедура восстановления	
структурных факторов	13
2.3. Восстановление структурных факторов с	
использованием гистограмм электронной плотности	15
3. Классические прямые методы	15
4. Экспериментальные методы определения фаз	
структурных факторов низкого разрешения	16
4.1. Изоморфное замещение	16
4.2. Молекулярное замещение	17
4.3. Метод вариации контраста	17
4.4. Многоволновая дифракция	18
5. Ab-initio определение фаз при низком разрешении	19
5.1. Поиск в пространстве фаз	19
5.2. Критерии, основанные на свойствах функции	
распределения электронной плотности.	20
5.2.1. Гистограмма электронной плотности	20
5.2.2. Топологические свойства (связность)	21
5.2.3. Локальная вариация плотности	22
5.3. Поиск в пространстве параметров	22
5.3.1. Сферически симметричная модель	23

5.3.2. Сферически несимметричная модель	23
5.3.3. Параметризация при помощи нескольких сфер	24
5.3.4. Параметризация при помощи большого числа	
сфер	24
5.3.5. Использование сферической оболочки	24
5.3.6. Пиксельное моделирование	25
5.4. Критерии поиска в пространстве параметров	25
5.4.1. Критерий типа R-фактора	25
5.4.2. Коэффициент корреляции модулей	26
5.4.3. Анализ упаковки	27
5.5. Вероятностные подходы	27
5.5.1. Априорные распределения с максимальной	
энтропией	28
5.5.2. Маска области как априорное распределение	28
5.5.3. Выбор из нескольких альтернативных масок	29
5.5.4. Максимум правдоподобия	29
5.5.5. Аппроксимация априорного распределения	
градиентом функции правдоподобия	31
ГЛАВА II. FAM - МЕТОД РЕШЕНИЯ ФАЗОВОЙ	
ПРОБЛЕМЫ ПРИ НИЗКОМ РАЗРЕШЕНИИ	32
1. Предварительные сведения	33
1.1.Тестовый объект	33
1.2. Контрольный критерий: фазовая корреляция	33
1.3. Критерий отбора: корреляция модулей	35
1.4. FAM-модели	36
2. Связь между качеством модулей и фаз, рассчитанных	
по FAM - модели	37
2.1. Точность фаз, получаемых при помощи FAM	
моделей	37
2.2. Поиск оптимальной FAM-модели	40
2.3. Переформулировка задачи определения фаз	43

3. FAM - метод	45
4. 3.1. Генерация и отбор вариантов	45
3.2. Кластерный анализ	47
3.3. Обработка кластера	53
3.4. Уточнение значений фаз и повышение разрешения	53
4. Проблема выбора наилучшего кластера	54
4.1. Визуальный анализ	54
4.2. Анализ плотности на осях вращения	55
4.3. Топологические свойства (связность)	55
4.4. Монте-Карловский тест (правдоподобие)	55
ГЛАВА III. ОПРЕДЕЛЕНИЕ СТРУКТУРЫ РИБОСОМНОЙ	
ЧАСТИЦЫ 50S ИЗ Thermus thermophilus ПРИ	
РАЗРЕШЕНИИ ПОРЯДКА 40Å	61
1. Определение структуры T50S	62
2. Сравнение с электронно-микроскопической моделью	64
ПРИЛОЖЕНИЕ. Оптимальное выравнивание наборов фаз	74
1. Введение	74
2. Выравнивание карт	75
2.1. Группа Р1	75
2.2. Нетривиальные пространственные группы	76
3. Использование асимметричной части множества структурных	
факторов	78
4. "Перевернутые" карты	80
5. Разрешенные сдвиги начала координат	81
ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ	84
СПИСОК ЛИТЕРАТУРЫ	85

ВВЕДЕНИЕ

Изучение пространственной структуры биологических макромолекул и их комплексов является важным направлением современной биологии и создает предпосылки для летального 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 необходимой (модули структурных факторов), ЛЛЯ восстановления Получение недостающей исследуемой структуры. 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 объекта исследуемого 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 В макромолекулярными разработаны комплексами. то же время математические методы, позволяющие решать фазовую проблему для низкомолекулярных соединений, основываясь лишь на ланных 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, требуют наличия изоморфных чтобы подчеркнуть, что ОНИ не ΗИ производных, ни известных гомологичных объектов.

5

Задачей данного исследования являлась разработка процедуры прямого решения фазовой проблемы макромолекулярной кристаллографии при низком Решение фазовой разрешении. проблемы при низком разрешении значений подразумевает определение фаз для нескольких лесятков коэффициентов Фурье (структурных факторов) в разложении в ряд Фурье искомой функции распределения электронной плотности в исследуемом Наличие информации объекте. такой 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 рибосомной частицы T50S из *Thermus thermophilus* при разрешении порядка 40Å.

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

6

(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 $\exp\left[-2\pi i(hx+ky+lz)\right]$, 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 расчета значений соотношения лежат В основе методов 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 макромолекулярным объектам существенно ограничена размерами к И исследуемого объекта 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 разности 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 (параметры ячейки порядка 350Å) - это разрешение

8

40Å. Для кристаллов рибосомной 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.

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 рибосомной частицы T50S велись с использованием экспериментальных данных, полученных в группе A.Yonath (Институт им.Макса Планка, Гамбург, Германия) и в тесном сотрудничестве с A.Yonath и I.Agman.

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

I. ФАЗОВАЯ ПРОБЛЕМА ПРИ НИЗКОМ РАЗРЕШЕНИИ. (ЛИТЕРАТУРНЫЙ ОБЗОР).

1. Зачем нужны фазы рефлексов низкого разрешения.

Необходимость решения фазовой проблемы является главным препятствием на пути от полученных экспериментально значений модулей структурных факторов к картам распределения электронной плотности. Обычно для макромолекулярных объектов эта проблема решается методами изоморфного замещения (Perutz, 1965; Blow & Crick, 1959; Blow & Rossman, 1961), аномального рассеяния или молекулярного замещения (Rossmamm, 1972). При этом достигается разрешение карт, близкое к атомному, что позволяет интерпретировать эти карты в терминах координат атомов. При этом структурные факторы низкого разрешения (разрешение ниже 10Å) традиционно игнорируются в силу ряда экспериментальных и методических сложностей. Тем не менее знание фаз рефлексов низкого разрешения является весьма важным в ряде случаев, например:

- в случае, когда знание маски для области молекулы необходимо для решения фазовой проблемы при более высоком разрешении; Это включает в себя методы модификации электронной плотности (Podjarny, Rees & Urzhuntsev, 1996) и молекулярное замещение, где необходимо знание положения и ориентации объекта (Rossmann, 1990);
- в случае, когда знание фаз для базовых рефлексов низкого разрешения позволяет повысить эффективность классических прямых методов (Мо et al., 1996; Mathiesen & Mo, 1998);
- в случае исследования больших макромолекулярных комплексов, когда изображения даже при низком разрешении могут нести важную структурную информацию;
- когда измеряемое дифракционное поле кристаллов ограничено низким разрешением;

• в случаях, когда исключение рефлексов низкого разрешения вызывает большие искажения в изображении объекта (Podjarny, Schevitz & Sigler, 1981; Urzhumtsev, 1991; Stenkamp, 1994).

Определение термина "низкое разрешение" зависит от традиций различных лабораторий и типа исследуемых объектов. Этот термин часто используется просто для обозначения рефлексов центральной зоны обратного пространства, не измеренных в эксперименте. Альтернативный путь использовать термины "очень низкое разрешение" и "низкое разрешение" в соответствии с тем, как влияет рассеяние рентгеновских лучей растворителем на рефлексы этой зоны (Urzhumtsev & Podjarny, 1995; Podjarny & Urzhumtsev, 1997):

- Очень низкое разрешение (d > 25Å). В этой зоне разрешения близко расположенные атомы объекта рассеивают рентгеновские лучи почти в фазе, что приводит к очень большим величинам модулей структурных факторов для некоторых рефлексов. С другой стороны, при этом уровне разрешения может использоваться модель однородного растворителя. В этом случае комплексные структурные факторы, соответствующие рассеянию растворителем, коллинеарны структурным факторам, отвечающими изучаемому объекту. Проблема учета влияния растворителя сводится поэтому к переопределению коэффициента шкалирования экспериментальных данных.
- Низкое разрешение (25Å > d > 8Å). В этой зоне разрешения границы области молекулы становятся более сложными, и плотность растворителя не может более рассматриваться как постоянная во всей области, занятой растворителем (Badger & Caspar, 1991; Urzhumtsev & Podjarny, 1995). Проблема моделирования влияния растворителя становится более сложной.
- *Среднее и высокое разрешения* (8Å > d). В этой зоне разрешения влиянием растворителя на интенсивность рассеянных рентгеновских лучей можно в первом приближении пренебречь.
- 2. Восстановление структурных факторов низкого разрешения.

Рефлексы низкого (и очень низкого) разрешения чрезвычайно важны для формирования правильных внешних очертаний изучаемого объекта. Отсутствие даже небольшого числа структурных факторов низкого разрешения может приводить к существенным искажениям изображения и создавать значительные сложности при попытках изучать его топологические свойства (например, связность; Baker, Krukowski & Agard, 1993; Urzhumtsev, Lunin & Lunina, 1998; Lunin, Lunina & Urzhumtsev, 1998) или применять процедуры автоматической интерпретации карт электронной плотности (Lamzin & Wilson, 1993). Тем не менее эта информация часто игнорируется. Причины кроются как в технических сложностях измерения интенсивностей рассеянных лучей при малых углах рассеяния, как и в сложности определения фаз таких рефлексов. До некоторой степени отсутствие малоугловых рефлексов может быть компенсировано кооперативным эффектом сдвигов в фазах большого числа рефлексов среднего разрешения (Brazhnikov et al. 1993). Такой сдвиг, в частности, достигается при использовании широко распространенной процедуры модификации электронной плотности: плотность вне области молекулы принудительно зануляется, и фазы, Фурье преобразованием модифицированной полученные плотности, используются вместе с экспериментальными модулями для расчета нового синтеза Фурье (Bricogne, 1974; Hendrickson, Klippenstein & Ward, 1975). Эта процедура позволяет получить более отчетливые границы молекулы в элементарной ячейке даже при отсутствии малоугловых рефлексов, но ценой введения дополнительных ошибок в фазы структурных факторов. В этом случае рефлексам среднего разрешения "навязывается" несвойственная им роль формирования внешних очертаний молекулы.

2.1. Определение фаз методом максимального детерминанта.

В случае, когда интенсивности рефлексов низкого разрешения измерены, но соответствующие фазы не определены, существенное улучшение карт электронной плотности может быть достигнуто за счет определения этих фаз и включения соответствующих рефлексов в синтез

12

Фурье. При определении структуры тРНК было продемонстрировано (Podjarny, Schevitz & Sigler, 1981), как добавление в синтез 28 первоначально отсутствующих рефлексов низкого разрешения с фазами, определенными методом максимального детерминанта (Tsoucaris, 1970), существенно улучшило интерпретируемость карт.

2.2. Двухшаговая процедура восстановления структурных факторов.

Когда и модули, и фазы для рефлексов низкого разрешения неизвестны, но имеются достаточно точные фазы для рефлексов среднего разрешения, для некоторого числа отсутствующих малоугловых структурных факторов могут быть восстановлены значения как модулей, так и фаз. Такое восстановление в широко применяемой неявном виде присутствует В процедуре автоматического определения границ молекулы, предложенной Вонгом (Wang, 1985). В явном виде процесс восстановления значений структурных факторов может быть описан как двухшаговая процедура (Urzhumtsev, 1985; Urzhumtsev, Lunin & Luzyanina, 1989). На первом шаге (нелинейная фильтрация в прямом пространстве) восстанавливаются фазы рефлексов низкого разрешения. На втором шаге (усреднение в прямом пространстве или линейная фильтрация в обратном) восстанавливаются модули (Urzhumtsev, 1991). Первый шаг состоит в том, что каждой точке в элементарной ячейке приписывается некоторый вес, отражающий вероятность того, что эта точка принадлежит области молекулы. Определенная таким образом функция надежности используется на втором шаге процедуры для построения маски области. обладающей максимальной интегральной надежностью. Это достигается путем свертки функции надежности с некоторым гладким ядром. При благоприятных условиях в результате свертки возникает сглаженный образ области молекулы.

Существуют различные стратегии задания функции надежности. В процедуре Вонга значение надежности в точке совпадает со значением на синтезе Фурье, если это значение выше среднего, и равно нулю в противном случае. Очевидно, что все разнообразие функций, применяемых для модификации электронной плотности низкомолекулярных (Simonov, 1976) и макромолекулярных объектов (Podjarny, Rees & Urzhumtsev, 1996), может быть использовано аналогичным образом на этом этапе. В частности, при работе с синтезами среднего разрешения не только самые высокие значения электронной плотности, но и самые низкие (отрицательные) могут быть использованы как индикатор того, что точка принадлежит области молекулы. Более сложные подход может состоять в приписывании большей надежности точкам, обладающим максимальной локальной флюктуацией электронной плотности (Jones, Walker & Stuart, 1991). Было показано, что фазы, полученные обратным преобразованием Фурье функции надежности, могут являться хорошим приближением к правильным значениям фаз малоугловых структурных факторов. Таким образом, нелинейная процедура фильтрации в прямом пространстве может рассматриваться как метод определения фаз для рефлексов низкого разрешения. Этот метод определения фаз был использован на ранних стадиях определения фактора элонгации G (Чиргадзе и др., 1991; Urzhumtsev, 1991). Здесь определенные методом изоморфного замещения фазы разрешения 30-8Å были использованы для определения фаз 29 рефлексов в зоне разрешения 120-30Å.

Второй шаг процедуры восстановления малоугловых структурных факторов эквивалентен (Urzhumtsev, 1985; Leslie, 1988) умножению модулей обратного преобразования Фурье функции надежности на некоторую убывающую функцию в обратном пространстве (например, гауссову функцию). Для этой цели были предложены разные сглаживающие функции (Shevyrev & Simonov, 1981; Number & Stubbs, 1985; Wang, 1985; Urzhumtsev, Lunin & Luzyanina, 1989). Было показано (Urzhumtsev, 1991), что сглаженные таким образом структурные факторы функции надежности являются разумным приближением структурных факторов низкого разрешения и могут быть добавлены в синтез Фурье, если соответствующие модули не были экспериментально определены.

2.3. Восстановление структурных факторов с использованием гистограмм электронной плотности.

Альтернативная процедура определения фаз или полного восстановления отсутствующих малоугловых структурных факторов может быть основана на свойствах гистограмм электронной плотности (Lunin, 1988; 1993). Гистограммы электронной плотности (Lunin, 1988; Harrison, 1988; Luzzati, Mariani & Delacroix, 1988, Zhang & Main, 1990) могут быть предсказаны для синтезов среднего разрешения до того, как соответствующая атомная структура определена (Lunin & Skovoroda, 1991). Они чувствительны к ошибкам в фазах, и неизвестные фазы (и при необходимости модули) могут быть определены из требования, чтобы при использовании их вместе с известными структурными факторами гистограмма соответствующего синтеза Фурье была максимально близка к предсказанной. Этот подход применялся (Lunin & Skovoroda, 1991) для восстановления амплитуд и фаз примерно 30% рефлексов в зоне разрешения до 4Å для "сухой" формы у-кристаллина IIIb (Chirgadze et al., 1991), модули которых не были определены в эксперименте. В качестве стартовой точки при минимизации расхождения гистограмм (Lunin, 1988) были использованы фазы ранее определенной "влажной" формы у-кристаллина IIIb (Чиргадзе и др., 1986). Возможна и альтернативная техника использования гистограмм для уточнения значений фаз (Harrison, 1988; Zhang & Main, 1990), которая более близка к обычным методам модификации электронной плотности. Эта процедура более проста в реализации, чем минимизация расхождения гистограмм, но вовлекает в работу только половину ограничений, налагаемых известной гистограммой (Lunin & Vernoslova, 1991).

3. Классические прямые методы.

Развитие классических прямых методов привело недавно к вдохновляющим успехам в *ab-initio* определении фаз для небольших белков (Miller & Weeks, 1998; Sheldrick, 1998). Однако для успешного применения этих методов требуется наличие экспериментальных данных очень высокого 1.0Å), разрешения (порядка И ОНИ ограничены пока структурами, содержащими не более 1000 атомов в независимой части элементарной ячейки. Первые попытки перенесения прямых методов с низкомолекулярных соединений на макромолекулы были основаны на надежде, что при изменении размеров (примерно в 10 раз) большие молекулы при низком разрешении станут похожими на малые молекулы при высоком разрешении. Эта идея была проверена на данных тРНК (Podjarny & Yonath, 1977), где для определения фаз были использованы матричные методы (Tsoucaris, 1970), которые могут рассматриваться как расширение классических прямых методов.

4. Экспериментальные методы определения фаз структурных факторов низкого разрешения.

Мы обсудим ниже методы определения фаз, использующие информацию, полученную в дополнительных экспериментах. Следует заметить, что стандартные методы определения фаз для рефлексов среднего и высокого разрешения могут давать совсем плохие значения фаз для рефлексов низкого разрешения.

4.1. Изоморфное замещение.

Метод изоморфного замещения обычно не позволяет надежно определить фазы структурных факторов низкого разрешения. Это связано с тем, что относительный вклад тяжелых атомов в очень сильные рефлексы малоугловой зоны становится очень мал, что ведет к большим ошибкам в определяемых фазах. Чтобы обойти эту сложность, делаются попытки использовать в качестве изоморфных добавок достаточно большие кластеры Олнако встречаются большие тяжелых атомов. на ЭТОМ ПУТИ экспериментальные трудности.

4.2. Молекулярное замещение.

При наличии приближенной модели исследуемого объекта для определения фаз может быть использован метод молекулярного замещения (Rosmann, 1972). Однако этот метод, естественно, должен быть

модифицирован для работы при низком разрешении, поскольку в качестве приближенной модели здесь более уместно использовать некоторую оболочку (например, полученную методами электронной микроскопии), чем атомную модель структуры. С технической точки зрения это не вносит существенных изменений в метод, поскольку поиск положения и ориентации модели в ячейке осуществляется путем сравнения экспериментальных и рассчитанных модулей структурных факторов, и при этом не имеет значения, как рассчитаны структурные факторы - по атомной модели или по некоторой оболочке. Тем не менее, некоторые изменения в методику должны быть внесены. В частности, было показано (Urzhumtsev & Podjarny, 1995), что при низком разрешении используемая в качестве критерия функция корреляции дает более надежные результаты, чем функция перекрытия синтезов Паттерсона. С этими изменениями стандартные пакеты программ, такие как AMoRe (Navaza, 1994; Navaza & Vernoslova, 1995), могут быть успешно применены для правильного размещения оболочки в элементарной ячейке кристалла.

Отметим, что фазы, рассчитанные по размещенной таким образом оболочке, могут быть использованы, например, для определения мест расположения кластеров из тяжелых атомов (Ban *et al.*, 1998).

4.3. Метод вариации контраста.

Рассеивающая сила растворителя может быть изменена либо изменением концентрации солей (Bragg and Perutz, 1952) при использовании рентгеновского излучения, либо изменением соотношения D_2O и H_2O (Roth, 1987; Bentley et al.; 1987) при использовании рассеяния нейтронов. Это создает возможность определять значения фаз структурных факторов за счет использования разностей в интенсивностях рассеяния при разном уровне контраста (Carter *et al.*, 1990).

4.4. Многоволновая дифракция.

В последнее десятилетие был разработан практический подход к экспериментальному определению значений трехфазных фазовых инвариантов ($T_{\mathbf{hk}} = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}$). Он основывается (Weckert & Hümmer, 1997) на исследовании характера незначительных изменений в интенсивности рассеянной волны в процессе пересечения дополнительной точкой обратной решетки сферы Эвальда. Анализ формы графика изменения интенсивности (ускана) позволяет приближенно определить значение соответствующего трехфазного инварианта. На примере кристаллов лизоцима тетрагональной формы было продемонстрировано (Weckert et al., 1997; Weckert et al., 1998), что этот метод может быть успешно применен к небольшим белкам. В этом эксперименте были определены значения более 800 фазовых триплетов, из которых были получены значения более 600 индивидуальных фаз со средней ошибкой менее 20°). Еще два примера экспериментального определения значений фаз представляют гексагональная форма лизоцима и целлюлаза (Mo et al., 1998), где точность определения фаз составила 17.9° и 15.9° соответственно.

Следует заметить, что не все индивидуальные фазы могут быть при этом определены однозначно, и для расширения набора фаз необходимо использовать некоторые дополнительные критерии (например, максимум правдоподобия или максимум энтропии) (Weckert *et al.*, 1998). Тем не менее, знание даже неполного набора трехфазных инвариантов может существенно расширить возможности прямых методов (Mo *et al.*, 1996; Mathiesen & Mo, 1998).

5. Ab-initio определение фаз при низком разрешении.

В этом разделе рассматриваются методы определения фаз, которые из экспериментальной информации используют только величины модулей структурных факторов, полученные в рентгеновском эксперименте с не модифицированным (нативным) объектом. Кроме того, для решения фазовой проблемы привлекаются некоторые сведения общего характера, не связанные со структурой конкретного объекта. Большинство этих методов являются многовариантными: делается попытка рассмотреть все возможные наборы фаз (либо их представительную выборку) и выбрать наиболее надежные из них.

Методы определения фаз для рефлексов низкого разрешения могут быть условно разделены на три группы в соответствии с тем, в каком конфигурационном пространстве производится перебор вариантов решения. Первую группу составляют методы, в которых различные наборы фаз перебираются непосредственно. Во второй группе методов для расчета фаз используются некоторые модели (например, состоящие из псевдоатомов), и в процессе поиска решения перебираются различные модели. В третьей группе объектом является наиболее методов поиска адекватное априорное распределение координат атомов. Знание этого априорного распределения позволяет рассчитать ожидаемые значения фаз структурных факторов. Центральным вопросом во всех этих подходах является вопрос о том, какие существуют критерии отбора, позволяющие отделить "плохие" варианты от "хороших". Отсутствие достаточно надежных критериев отбора является основной проблемой при *ab-initio* определении фаз. Для разных групп методов могут быть предложены различные критерии, которые будут рассмотрены ниже.

5.1. Поиск в пространстве фаз.

Наиболее прямолинейный подход - проверить все возможные комбинации фаз. Очевидно, что это можно сделать для центросимметричных рефлексов. Для нецентросимметричных рефлексов перебираемые значения фаз должны быть ограничены небольшим число дискретных значений рефлексов (Bricogne, 1993). Как правило, при этом берутся значения $\pi/4$, $3\pi/4$, 5π/4, 7π/4. Число возможных фазовых наборов возрастает экспоненциально при увеличении числа рассматриваемых рефлексов. Поэтому даже такой перебор небольшого числа возможен только ДЛЯ очень рефлексов (Urzhumtsev, Lunin & Lunina, 1998; Lunin, Lunina & Urzhumtsev, 1998). В том случае, когда число возможных вариантов слишком велико, возможным

19

выходом является анализ реалистичного (с точки зрения возможностей компьютера) числа случайно сгенерированных наборов фаз (Lunin, & Skovoroda. 1990: Woolfson & Yao Urzhumtsev Jia-xing. 1990). Альтернативный путь - использование регулярных сеток в конфигурационном пространстве всех фазовых наборов. Для построения таких сеток могут использоваться, например, магические числа (White & Woolfson, 1973) или коды, исправляющие ошибки (Woolfson, 1954; Bricogne, 1993).

5.2. Критерии, основанные на свойствах функции распределения электронной плотности.

В случае, когда анализируются различные наборы фаз, естественно использовать критерии, основанные на свойствах синтезов Фурье, рассчитанных с экспериментальными модулями и пробными значениями фаз. Некоторые из таких критериев рассмотрены ниже.

5.2.1. Гистограмма электронной плотности.

Гистограммы распределения электронной плотности (Lunin, 1988; Harrison, 1988; Luzzati, Mariani & Delacroix, 1988, Zhang & Main, 1990; Lunin, 1993) содержат информацию о том, какие значения электронной плотности и с какой частотой могут быть обнаружены в элементарной ячейке кристалла. Такие химической гистограммы зависят от природы содержимого элементарной ячейки, от количества растворителя и от разрешения, при котором рассчитываются синтезы Фурье. Однако они не зависят (в первом приближении) от способа укладки полипептидной цепи и местонахождения молекулы в ячейке. Поэтому такие гистограммы могут быть предсказаны без знания конкретной молекулярной структуры либо для всей элементарной ячейки (Lunin & Skovoroda, 1991), либо только для области молекулы (Main, 1990). Для решения фазовой проблемы *ab-initio* был предложен подход Urzhumtsev,& Skovoroda, 1990), (Lunin, при котором гистограммы вычислялись синтезов случайно сгенерированными ДЛЯ co фазами.

"Хорошими" наборами фаз считались те, гистограммы которых лучше соответствовали некоторой эталонной (предсказанной) гистограмме.

В качестве упрощенных подходов к использованию специфики гистограмм электронной плотности могут рассматриваться подходы, в которых использовались ограничения на отдельные статистические моменты распределения электронной плотности (Lunin, 1993). Примерами этого является максимизация интеграла Кокрена (Cochran, 1952) или, что то же самое, максимизация ассиметрии синтеза Фурье (Podjarny & Yonath, 1977). Другой пример - введение ограничений на допустимую величину четвертого центрального момента (коэффициента эксцесса) (Luzzati, Mariani & Delacroix, 1988).

5.2.2. Топологические свойства (связность).

Первоначально идея использования характеристик связности области высоких значений на синтезе Фурье как показателя качества фаз была предложена для синтезов среднего и высокого разрешения (Baker, Krukowski, & Agard, 1993). Она основывалась на наблюдении, что на синтезах Фурье, рассчитанных с правильными значениями фаз, обнаруживаются протяженные непрерывные области высоких значений плотности, соответствующие полипептидной цепи. Наоборот, наличие большого числа небольших изолированных "клякс" свидетельствует о значительных ошибках в фазах. В связи с этим был предложен принцип, согласно которому качество набора фаз считалось тем выше, чем меньшее количество связных компонент присутствовало на синтезе.

Анализ свойств связности синтезов Фурье низкого разрешения (Urzhumtsev, Lunin & Lunina, 1998; Lunin, Lunina & Urzhumtsev, 1998) показал, что к ним такой принцип не вполне применим. При низком разрешении плохо определенные фазы приводят скорее к появлению одной общей связной области, чем к появлению множества мелких. Поэтому в качестве признака правильности фаз было предложено свойство синтеза иметь столько связных компонент, сколько молекул содержится в элементарной ячейке. Преобладание в наборе рефлексов некоторых специальных типов (например, таких, у которых фазы являются однофазными структурными инвариантами) может искажать карту электронной плотности, делая невозможным использование этого признака. Удаление из синтеза компоненты, не меняющейся при допустимых сдвигах начала координат, приводит к некоторому "обострению" синтеза и позволяет применить этот критерий (Urzhumtsev, Lunin & Lunina, 1998; Lunin, Lunina & Urzhumtsev, 1998).

5.2.3. Локальная вариация плотности.

Наблюдение, что значительные локальные флюктуации электронной плотности присущи области молекулы и не характерны для области растворителя (Urzhumtsev, 1991; Jones, Walker & Stuart, 1991), было использовано для построения нового критерия качества фаз, основанного на среднем значении локальной вариации плотности (Mishnev & Woolfson, 1994).

5.3. Поиск в пространстве параметров.

Поскольку поиск в конфигурационном пространстве всех фазовых наборов чрезвычайно трудоемок, то возможным выходом является замена его поиском в некотором специально выбранном подпространстве. Это может быть достигнуто при использовании для расчета значений фаз некоторых моделей. описываемых меньшим числом параметров, чем число определяемых фаз. Каждая такая модель порождает некоторый набор фаз, но не каждый набор фаз может быть получен из некоторой модели. Выбор подходящей параметризации является ключевым моментом в этом подходе. С одной стороны, чем меньше число параметров модели, тем проще С осуществляется анализ всевозможных моделей. другой стороны, перебираемых моделей должно быть достаточно много, чтобы среди них присутствовали модели, порождающие фазы, близкие к правильным.

5.3.1. Сферически симметричная модель.

Простейшим примером параметризации пространства фазовых наборов является аппроксимация молекулы одной сферой. Электронная плотность может предполагаться либо постоянной внутри этой сферы, либо задаваться некоторой сферически симметричной функцией, например, гауссовой. Параметрами, описывающими такую модель, являются координаты центра сферы. Процедура поиска в этом случае состоит в поочередном помещении центра сферы во все точки некоторой элементарной ячейки. Для каждого положения центра рассчитываются модули структурных факторов и сравниваются с экспериментальными значениями. Этот подход применялся во многих лабораториях с разным успехом (Podjarny et al., 1987; Volkmann et al., 1995). Более подробный анализ (Andersson & Hovmöller 1996; Harris1995; 1998^a, 1998^b). показал, Petrova, Lunin & Podjarny, ЧТО наилучшее (обеспечивающее рассчитанных максимальное соответствие И экспериментальных модулей) положение центра сферы может оказываться иногда внутри области молекулы, иногда внутри области растворителя, а иногда на границе их раздела.

5.3.2. Сферически несимметричная модель.

В ряде случаев сферическая аппроксимация исследуемого объекта перестает быть адекватной, и требуется более аккуратная модель. Хорошей аппроксимацией может являться модель гомологичного объекта, если такая известна. В этом случае фазовые наборы параметризованы шестью переменными (3 угла Эйлера и 3 координаты центра тяжести), и мы имеем широко распространенный метод молекулярного замещения (Rossmann, 1972). **5.3.3. Параметризация при помощи нескольких сфер.**

В работе Поджарни с соавторами (Podjarny *et al.*,1987) результаты, полученные с использованием сферической модели, были уточнены путем использования более сложной модели, образованной несколькими большими сферами. Положение этих сфер в элементарной ячейке были определены из разностных карт. Возможности использования нескольких сфер для

параметризации пространства всевозможных фазовых наборов будут далее проанализированы подробно в основной части диссертации.

5.3.4. Параметризация при помощи большого числа сфер.

Попытка использовать модели, состоящие из большого числа псевдоатомов (более 100), была сделана Субайи (Subbiah, 1991, 1993; David and Subbiah, 1994). В его подходе стартовая модель получалась случайным размещением псевдоатомов в элементарной ячейке. Далее координаты их центров уточнялись в обратном пространстве. Автор анонсировал, что при этом возникает тенденция к "конденсации" псевдоатомов либо в области молекулы, либо в области растворителя. Однако, как показано далее в данной работе, интерпретация результатов такого уточнения требует значительной осторожности.

5.3.5. Использование сферической оболочки.

Определение структуры кристаллических вирусов имеет ряд особенностей. Во-первых, в этом случае положение центра объекта и его ориентация бывают известны из геометрических соображений. Поскольку внутреннее ядро вируса (содержащее ДНК и РНК) обычно существенно разупорядочено, структура может быть аппроксимирована сферической или цилиндрической оболочкой. Неизвестными параметрами при этом являются геометрические характеристики этой оболочки. Фазы, рассчитанные с такой упрощенной модели, могут использоваться как первое приближение к фазам структурных факторов низкого разрешения.

Во-вторых, наличие одновременно нескольких экземпляров молекулы в независимой части ячейки приводит к определенным соотношениям между структурными факторами. Эти ограничения могут быть очень сильными, если число идентичных экземпляров велико. Использование усреднения по некристаллографичекой симметрии (Rossmann and Blow, 1963; Jack, 1973; Harrison & Jack, 1975; Rayment et al., 1978; Rayment et al., 1982; Chapman et al., 1992; Filman et al., 1994) позволяет решать фазовую проблему для вирусов и при отсутствии изоморфных производных. Следует заметить, однако, что присутствие модулей низкого разрешения может быть очень существенным для успеха этой процедуры (Urzhumtsev, 1991; Rayment, 1983).

5.3.6. Пиксельное моделирование.

Еще один путь параметризации пространства всевозможных фазовых наборов - задание непосредственно функции электронной плотности во всех точках некоторой грубой сетки элементарной ячейки. Естественно, в этом случае число параметров может быть очень велико, и должны быть введены какие-то сильные дополнительные ограничения на значения этих параметров, чтобы сделать поиск в пространстве параметров возможным. Эта проблема была успешно решена при исследовании структуры сферических вирусов, где число независимых узлов сетки было существенно ограничено симметрией вируса, а возможные значения электронной плотности были ограничены значениями 0 или 1 (Miller *et al.*, 1996).

5.4. Критерии поиска в пространстве параметров.

5.4.1. Критерий типа R-фактора.

Поскольку при использовании моделей для генерации значений фаз структурных факторов одновременно генерируются и значения их модулей, то естественно пытаться оценивать качество модели, сравнивая величины рассчитанных и экспериментальных модулей. Обычной способом оценки степени этого соответствия является, например, вычисление кристаллографического R-фактора.

$$R = \frac{\frac{\sum_{\mathbf{h}} \left| F_{\mathbf{h}}^{obs} - F_{\mathbf{h}}^{calc} \right|}{\sum_{\mathbf{h}} \left| F_{\mathbf{h}}^{obs} \right|} \qquad (2)$$

Эта характеристика имеет 2 недостатка. Во-первых, для ее вычисления необходимо приведение в единую шкалу рассчитанных и экспериментальных модулей структурных факторов, что может вызывать существенные

затруднения при использовании данных только низкого разрешения. Вовторых, эта характеристика (как функция параметров модели) не имеет производной во многих точках, что делает невозможным применение к ней градиентных методов минимизации. Второе затруднение может быть преодолено путем использования критериев типа

$$R_2 = \sum_{\mathbf{h}} \left(F_{\mathbf{h}}^{obs} - F_{\mathbf{h}}^{calc} \right)^2 \quad , \tag{3}$$

однако проблема шкалирования в этом случае все равно остается.

5.4.2. Коэффициент корреляции модулей.

Проблема выбора коэффициента шкалирования может быть снята путем введения в рассмотрение коэффициентов корреляции модулей. Коэффициент корреляции модулей может быть определен либо по аналогии со статистическим коэффициентом корреляции как

$$MC_{1} = \frac{\sum_{\mathbf{h}} \left(F_{\mathbf{h}}^{obs} - \left\langle F^{obs} \right\rangle \right) \left(F_{\mathbf{h}}^{calc} - \left\langle F^{calc} \right\rangle \right)}{\sqrt{\sum_{\mathbf{h}} \left(F_{\mathbf{h}}^{obs} - \left\langle F^{obs} \right\rangle \right)^{2} \sum_{\mathbf{h}} \left(F_{\mathbf{h}}^{calc} - \left\langle F^{calc} \right\rangle \right)^{2}}}$$
(4)

либо просто как косинус угла меду двумя наборами модулей, рассматриваемых как вектора многомерного пространства

$$MC_{2} = \frac{\sum_{\mathbf{h}} F_{\mathbf{h}}^{obs} F_{\mathbf{h}}^{calc}}{\sqrt{\sum_{\mathbf{h}} \left(F_{\mathbf{h}}^{obs}\right)^{2} \sum_{\mathbf{h}} \left(F_{\mathbf{h}}^{calc}\right)^{2}}} \qquad (5)$$

5.4.3. Анализ упаковки.

Поскольку геометрические размеры изучаемого объекта обычно приблизительно известны, анализ упаковки молекул в элементарной ячейке кристалла позволяет сразу отсечь значительное число вариантов выбора центра молекулы (Ward, Hendrickson & Klippenstein, 1975). Однако специальные усилия могут быть необходимы для вычисления функции упаковки на достаточно подробной сетке в элементарной ячейке (Shapiro *et al.*, 1992). Следует также заметить, что, сокращая число допустимых вариантов, этот критерий все же не дает однозначного ответа и оставляет для дальнейшего анализа значительное число возможностей.

5.5. Вероятностные подходы.

В рамках вероятностного подхода изучаемая структура рассматривается как результат случайной и независимой генерации необходимого числа атомов при условии, что задано некоторое априорное распределение вероятностей для координат атомов. Такая математическая модель позволяет предсказывать величины фазовых инвариантов и в благоприятных случаях В отдельных фаз. качестве таких предсказаний могут, например, использоваться математические ожидания соответствующих структурных факторов. В классических прямых методах априорное распределение вероятностей берется равномерным, то есть предполагается, что атом с равной вероятностью может находиться в любой точке ячейки. Эта гипотеза не очень хороша для макромолекулярных кристаллов, поскольку здесь наблюдается существенная разница между областями молекулы и растворителя.

Известное априорное распределение вероятностей до некоторой степени решает фазовую проблему, поскольку оно позволяет рассчитать ожидаемые значения фаз, и карты этого априорного распределения сами по себе могут давать возможность определить границы области молекулы. В связи с этим проблема определения фаз при низком разрешении может быть переформулирована как задача поиска наиболее подходящего априорного распределения в рамках вероятностного подхода (Bricogne & Gilmore, 1990). Как и выше, два основных вопроса, возникающих на этом пути:

- какое множество допустимых априорных распределений рассматривается;
- какие существуют критерии для оценки адекватности рассматриваемого априорного распределения.

5.5.1. Априорные распределения с максимальной энтропией.

В ситуации, когда для набора *S* структурных факторов известны как модули, так и фазы, было предложено (Bricogne, 1984) использовать априорное распределение, обладающее двумя следующими свойствами:

- математические ожидания (по отношению к рассматриваемому априорному распределению) структурных факторов из набора S имеют предписанные значения;
- среди всех априорных распределений, удовлетворяющих первому условию, выбирается то, для которого энтропия максимальна.

Если предположить, что модули структурных факторов из набора *S* определены в эксперименте, то каждое такое априорное распределение задается значениями фаз структурных факторов из этого набора. И поэтому проблема перебора всех таких априорных распределений эквивалентна проблеме перебора всех возможных наборов фаз (см. раздел 5.1.).

5.5.2. Маска области как априорное распределение.

В качестве альтернативного подхода можно рассмотреть семейство априорных распределений, каждое из которых задает маску области молекулы (принимает значение 0 или 1). Например, в качестве простейшего семейства таких масок можно рассмотреть семейство сферических областей в элементарной ячейке с разными положениями центра сферы (Petrova, Lunin & Podjarny, 1998^a, 1998^b).

5.5.3. Выбор из нескольких альтернативных масок.

При решении фазовой проблемы при низком разрешении обычной является ситуация, когда применяемые критерии отбора не позволяют однозначно решить фазовую проблему, но сводят ее к небольшому числу альтернативных решений. Каждому из этих решений соответствует маска области и соответствующее априорное распределение вероятностей (Lunin *et al.*, 1998).

5.5.4. Максимум правдоподобия.

Состоятельность гипотезы о том, что изучаемая структура могла бы быть получена как результат случайной генерации координат атомов с данным априорным распределением, может быть оценена величиной ее правдоподобия (Cox & Hinkley, 1974).. Величина правдоподобия отражает вероятность воспроизвести экспериментально полученные значения модулей в процессе случайной генерации атомных моделей, контролируемом этим априорным распределением. Поэтому рассматриваемые априорные распределения могут сравниваться на основании величины их правдоподобия.

Главное препятствие на пути практического применения этой идеи - то, что вычисление величины правдоподобия является достаточно сложным. Аналитическая процедура вычисления правдоподобия включает в себя вывод вероятностей для множества структурных совместного распределения факторов и интегрирование этого распределения по величинам фаз при условии, что значения модулей совпадают с экспериментально полученными величинами. Поскольку выражение ДЛЯ совместного распределения вероятностей структурных факторов не удается получить в аналитическом виде, обычно применяются некоторые асимптотические разложения,. Наиболее широко используемая форма такого разложения (Klug, 1958) дает хорошую точность только для структурных факторов, незначительно отклоняющихся от своих математических ожиданий. Для случая больших отклонений более точным является асимптотическое разложение, полученное Бриконем (Bricogne, 1984) с использованием метода перевала. Однако эти формулы включают в себя неявную функцию, определяемую как решение большой нелинейных системы уравнений, которая не поддается аналитическому решению. Интегрирование по фазам также представляет существенный математические сложности даже в случае, когда совместное распределение вероятностей аппроксимируется многомерным гауссовым распределением. Единственным проблему исключением, когда интегрирования удается решить аналитически, является так называемая диагональная аппроксимация (Bricogne & Gilmore, 1990), в которой

29

внедиагональные элементы матрицы ковариаций полагаются равными нулю. Это приводит к тем же результатам, как если бы структурные факторы являлись независимыми случайными величинами. Такое упрощение может приводить к потере существенной части фазовой информации. То же самое упрощение функции правдоподобия используется для оценки ошибок в фазах при расчете фаз по несовершенным атомным моделям (Lunin & Urzhumtsev, 1984; Read, 1986; Lunin & Skovoroda, 1995; Urzhumtsev, Skovoroda & Lunin, 1996).

Как альтернатива диагональной аппроксимации для вычисления величины правдоподобия может быть предложена Монте-Карловская компьютерная процедура (Lunin *et al.*, 1998). В этом подходе правдоподобие определяется в несколько более слабой форме. Для процесса генерации атомных моделей с тестируемым априорным распределением вычисляется вероятность воспроизвести значения модулей структурных факторов, достаточно близкие к экспериментальным (а не в точности совпадающие, как в обычном определении правдоподобия). Такая величина может быть вычислена путем использования компьютерной процедуры.

Однако следует подчеркнуть, что возможность эффективного вычисления величины правдоподобия не снимает вопроса, является ли априорное распределение, обладающее максимальным правдоподобием, искомым распределением. В каждом конкретном случае здесь необходим тщательный анализ (Lunin, 1998; Petrova, Lunin, Podjarny, 1998^a, 1998^b; Petrova *et al.*, 1998).

5.5.5. Аппроксимация априорного распределения градиентом функции правдоподобия.

В случае, когда в качестве стартового априорного распределения вероятностей берется равномерное распределение, в качестве первого приближения к неравномерному распределению может использоваться градиент функции правдоподобия по отношению к априорному распределению (Lunin, 1998; Petrova *et al.*, 1998). (Более точно, должно рассматриваться условное правдоподобие при условии, что определены фазы, фиксирующие начало координат и энантиомер). Карта градиента функции правдоподобия может быть использована для определения вероятных и маловероятных областей для нахождения атомов в элементарной ячейке.

ГЛАВА II. FAM - МЕТОД РЕШЕНИЯ ФАЗОВОЙ ПРОБЛЕМЫ ПРИ НИЗКОМ РАЗРЕШЕНИИ

В этой главе излагаются основы FAM (Few Atoms Model) -метода и приводятся результаты его тестирования на объектах с известной структурой. Первоначально разрабатываемый метод основывался на следующих гипотезах:

- при низком разрешении распределение электронной плотности можно достаточно хорошо приблизить суммой вкладов небольшого числа очень тяжелых псевдоатомов (глобул); соответственно, структурные факторы низкого разрешения (и, в частности, их фазы) можно приблизить структурными факторами, рассчитанными по модели, состоящей из небольшого числа псевдоатомов;
- поскольку число параметров, описывающих такую модель, сравнительно невелико, то при случайной генерации большого числа моделей будут появляться и модели, достаточно хорошо воспроизводящие искомые фазы;
- такие модели могут быть идентифицированы по близости рассчитанных по модели модулей структурных факторов к модулям, полученным экспериментально.

Однако исследование показало, что эти гипотезы выполняются лишь частично, и в изложенную схему нужно внести существенное добавление. Эти вопросы будут подробно обсуждены ниже.

Предлагаемая методика тестировалась на нескольких объектах с известной структурой: комплекс аспартил-аминоацил тРНК синтетазы с тРНК, а также синтетазы и тРНК в отдельности, РНКазе *Sa*, факторе элонгации G, модельными данными для рибосомной частицы 50S и применялась к определению структуры рибосомной частицы 50S из *Thermus thermophilus* и нескольких небольших белков. В данной работе излагаются результаты тестирования метода на данных AspRS-комплекса и применения его к определению структуры T50S, полученные при личном участии автора.

1. Предварительные сведения.

1.1. Тестовый объект.

Ниже при анализе отдельных этапов работы в качестве иллюстрации приведены результаты тестирования метода на данных рассеяния нейтронов комплексом аспартил - аминоацил - 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=354Å, α=β=γ=90°. Пространственная структура комплекса была определена ранее (Urzhumtsev et al., 1994) методом молекулярного замещения при помощи пакета AMoRe (Navaza, 1994) по данным рентгеновского рассеяния разрешения 8.0Å с использованием атомной модели, отвечающей другой кристаллической форме. 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. Данная форма кристаллов комплекса оказалась особенно удобна для работы разрешении вследствие большой элементарной ячейки, при низком значительной доли растворителя (78%) и компактной формы комплекса. Были собраны 3 набора данных нейтронного рассеяния, основанных на разнице в контрасте H_2O/D_2O (Moras et al., 1983), соответствующих полному комплексу, и синтетазной и тРНКовой компонентам. Данные нейтронного рассеяния с высокой точностью коррелировали co структурными факторами, рассчитанными по модели, использованной в методе молекулярного замещения (corr(F_{mod},F_{obs})>92% при 20 Å).

1.2. Контрольный критерий: фазовая корреляция.

При тестирования методов решения фазовой проблемы на известных структурах обычно используются различные контрольные критерии, отражающие близость рассматриваемых фаз $\{\varphi_h\}$ к их истинным значениям $\{\varphi_h^{ex}\}$. При этом под точными значениями фаз понимаются фазы,

рассчитанные по известной атомной модели объекта. В этой работе мы будем использовать, главным образом, один тип контрольного критерия - фазовую корреляцию (Carbo et al., 1980; Read & Moult, 1992; Lunin & Woolfson, 1993):

$$PC(\{\varphi_{\mathbf{h}}\},\{\varphi_{\mathbf{h}}^{ex}\}) = \sum_{\mathbf{h}} (F_{\mathbf{h}}^{o})^{2} \cos(\varphi_{\mathbf{h}}^{ex} - \varphi_{\mathbf{h}}) / \sum_{\mathbf{h}} (F_{\mathbf{h}}^{o})^{2}$$
$$= \frac{\int \rho^{exact}(\mathbf{r})\rho(\mathbf{r})dV_{\mathbf{r}}}{\sqrt{\int \rho^{exact}(\mathbf{r})^{2}dV_{\mathbf{r}}\int \rho(\mathbf{r})^{2}dV_{\mathbf{r}}}}, \qquad (6)$$

где $\{F_{\mathbf{h}}^{o}\}$ - экспериментальные модули. Эта величина совпадает с коэффициентом корреляции двух рассчитанных с экспериментальными модулями синтезов Фурье, один из которых рассчитан с тестируемыми, а другой с точными фазами, и может быть вычислена только в тестовом случае, когда известны точные фазы. Нужно отметить, что экспериментальные модули сами по себе не фиксируют однозначно начало координат и энантиоморф. Поэтому при определении фаз ab initio мы должны либо фиксировать некоторым образом начало координат и энантиоморф, либо выполнить процедуру выравнивания фаз (см. Приложение) перед тем, как вычислять корреляцию приведенной формуле. В дальнейшем по предполагается, что перед сравнением фаз всегда будет проводиться такое выравнивание.

Коэффициент корреляции фаз - это не единственный способ сравнения текущих фаз с правильными. Например, среднее значение абсолютной фазовой ошибки или некоторая взвешенная фазовая ошибка также могут быть использованы для этой же цели. Мы будем поэтому понимать под "фазовой корреляцией" любой контрольный критерий, основанный на сравнении пробных и точных фаз. Примером контрольного критерия, который не связан с фазовой корреляцией, является степень вписывания модели в синтез, определяемая как число атомов правильной (уточненной) модели внутри маски области, которая рассчитывается c пробными фазами И экспериментальными модулями.

1.3. Критерий отбора: корреляция модулей

Мы называем критерием отбора некоторую косвенную оценку качества пробного набора фаз (сгенерированного непосредственно либо рассчитанного по какой-то модели), которая может быть вычислена либо на основании либо экспериментальных модулей, с использованием некоторой предварительной информации общего вида. Имеется в виду предварительная информация, такая как гистограммы распределения электронной плотности, связность, атомность и так далее, не требующая знания точных фаз. Единственное требование - должна существовать возможность вычислить величину этого критерия для каждого сгенерированного набора фаз, не зная правильных значений фаз. Подходящий выбор этого критерия является ключевой точкой для успешной работы метода. Необходимо отметить, что до сих пор не найдено критерия, который бы безошибочно указывал на правильное решение.

В этой работе будет рассмотрен частный случай, когда пробные фазы будут рассчитываться по псевдоатомной модели. В этом случае модули, рассчитанные по модели, могут быть использованы для определения критерия отбора, а именно

$$MC_{1}\left(\left\{F_{\mathbf{h}}^{c}\right\},\left\{F_{\mathbf{h}}^{o}\right\}\right) = \frac{\sum_{\mathbf{h}} \left(F_{\mathbf{h}}^{o} - \left\langle F^{o}\right\rangle\right) \left(F_{\mathbf{h}}^{c} - \left\langle F^{c}\right\rangle\right)}{\sqrt{\sum_{\mathbf{h}} \left(F_{\mathbf{h}}^{o} - \left\langle F^{o}\right\rangle\right)^{2} \sum_{\mathbf{h}} \left(F_{\mathbf{h}}^{c} - \left\langle F^{c}\right\rangle\right)^{2}}}$$
(7)

или

$$MC_{2}\left(\left\{F_{\mathbf{h}}^{c}\right\},\left\{F_{\mathbf{h}}^{o}\right\}\right) = \frac{\sum_{\mathbf{h}} F_{\mathbf{h}}^{o} F_{\mathbf{h}}^{c}}{\sqrt{\sum_{\mathbf{h}} \left(F_{\mathbf{h}}^{o}\right)^{2} \sum_{\mathbf{h}} \left(F_{\mathbf{h}}^{c}\right)^{2}}} \quad .$$

$$(8)$$

Мы будем называть корреляцией модулей (МС) любой тип критерия отбора, основанный на сравнении рассчитанных и экспериментальных модулей. Очевидно, что возможны и другие критерии, кроме указанных только что, например, обычный R-фактор. Примером критерия отбора, который мы не называем корреляцией модулей, может служить близость стандартной гистограммы электронной плотности и гистограммы, которая получается из синтеза Фурье, рассчитанного с пробными фазами и экспериментальными модулями.

Мы будем говорить, что структурные факторы, рассчитанные по пробной атомной модели, являются ω -коррелированными, если соответствующая величина корреляции модулей МС выше, чем некоторый предписанный порог ω . Мы будем называть структурные факторы хорошо коррелированными (WMC), если они являются ω -коррелированными с достаточно высоким уровнем ω , который будет определен в каждом отдельном случае.

1.4. FAM-модели

Мы называем FAM-моделью модель, построенную из сравнительно небольшого числа псевдоатомов (глобул). Электронная плотность этих глобул моделируются либо с помощью Гауссовой функции (обычно с искусственно высоким значением В)

$$\rho(\mathbf{r}) = \left(\frac{4\pi}{B}\right)^{3/2} \exp\left(-\frac{4\pi^2 r^2}{B}\right) \qquad , \tag{9}$$

либо считается постоянной ненулевой плотностью внутри сферы заданного радиуса и равна нулю за ее пределами. Мы будем дальше ссылаться на число таких глобул в независимой части ячейки как на число атомов в FAM-модели, предполагая, что все симметрично-связанные глобулы также включены в модель. Простейший пример FAM-модели - модель из одного очень большого псевдоатома. Следует отметить, что величина параметра В не влияет на значения фаз, вычисляемые по FAM - модели, но влияет на значения модулей и, следовательно, на величины критериев отбора. Подходящее значение В может быть определено из предварительной оценки размеров объекта или подобрано таким образом, чтобы обеспечивать максимально возможное среднее значение этого критерия по результатам многих генераций.
2. Связь между качеством модулей и фаз, рассчитанных по FAM - модели.

2.1. Точность фаз, получаемых при помощи FAM - моделей.

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 разрешении (и, в частности, их фазы) можно приблизить низком факторами, рассчитанными по модели, состоящей структурными ИЗ Тесты небольшого числа псевдоатомов. с известными структурами подтверждают ЭТО предположение. Иногда остаточно даже одного псевдоатома, чтобы с удовлетворительной точностью приблизить полторадва десятка малоугловых фаз.

Ниже приведены результаты тестов, проведенных с использованием данных по нейтронному рассеянию для AspRS-комплекса. Эти тесты были проведены при разрешении 50Å (31 рефлекс), где экспериментальные данные практически полны. Тест заключался в том, что:

- при заданном числе атомов в FAM (это число варьировалось в разных тестах от одного до десяти) случайно генерировалось большое число таких моделей; координаты центров псевдоатомов при этом считались распределенными равномерно в элементарной ячейке;
- для каждой сгенерированной модели рассчитывались соответствующие ей структурные факторы и вычислялся коэффициент корреляции (6) рассчитанных фаз с точными;
- были рассчитаны гистограммы распределения значений этого коэффициента корреляции.

Для сравнения была также рассчитана гистограмма распределения коэффициента корреляции фаз для случая, когда случайно генерировались фазы структурных факторов (естественно, с учетом ограничений, налагаемых на значения фаз симметрией структуры).

На Рис. 1 приведены гистограммы для моделей, состоящих из 1, 2 и 10 псевдоатомов, а также для случая случайной генерация независимых фаз. Для моделей, состоящих из 10 атомов, распределение фаз по степени их близости к правильным фазам почти совпадает с распределением для случайно сгенерированных фаз. Т.е. использование FAM, состоящих из большого числа глобул, не дает выигрыша по сравнению с непосредственной генерацией фаз с точки зрения частоты генерации хороших наборов фаз. Однако использование моделей, состоящих из 1-2 глобул, существенно повышает (по сравнению со фазами) вероятность получения набора фаз, случайными близких к правильным (с фазовой корреляцией выше 75%). Заметим, что этот эффект ослабевает по мере того, как число глобул в модели растет; он больше всего для одноатомной модели, остается заметным для двухатомной, но почти исчезает для 10-атомной, так как в этом случае число переменных (30 координат атомов) практически совпадает с числом фаз (31 фаза). Таким образом, уменьшение числа степеней свободы модели при переходе от случайной генерации фаз (31 параметр) к, например, FAM из двух псевдоатомов (6 степеней свободы) позволяет повысить вероятность случайного получения хороших значений фаз.



Рис.1. Распределение величины коэффициента корреляции фаз для случайно сгенерированных моделей и случайно сгенерированных фаз.



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

2.2. Поиск оптимальной FAM-модели.

Таким образом, случайная генерация моделей, состоящих из небольшого числа псевдоатомов, позволяет, потенциально, получать достаточно хорошие наборы фаз. Проблема, однако, состоит в том, как распознать те модели, что 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 модули структурных факторов, рассчитанных по такой модели, наиболее близки к экспериментальным. Однако проверка на известных структурах показывает, что столь прямолинейных 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 (6-7). Здесь 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.

Из этой диаграммы, в частности, следует, что помещение центра глобулы в центр тяжести комплекса, хотя и приводит к разумным значениям фаз, но не обеспечивает высокой корреляции модулей. Максимальная же корреляция модулей обнаруживалась при размещении центра глобулы в специальных позициях на осях симметрии.

На Рис.3 показано, как меняется распределение сгенерированных наборов фаз по степени их близости к правильным фазам при введении отбора



Рис.3. Распределение коэффициента корреляции фаз для псевдомоделей, отобранных при разном пороговом значении критерия отбора.



Рис. 4. Значения коэффициентов корреляции модулей (МС) и фаз (РС), 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.

FAM-моделей. Из рисунка видно, что даже при достаточно жестком отборе моделей (при пороге, превышающем среднее значение величины корреляции модулей на два стандартных отклонения) среди отобранных вариантов остаются фазовые наборы с очень низкой корреляцией фаз.

Ситуация не меняется, если увеличить число атомов в модели (Рис.4). 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. В таблице 1 приведены результаты такого уточнения. Стрелки на Рис.3 показывают, как меняется положение точки на диаграмме в процессе уточнения координат атомов.

вариант	перед уточнением		после уточнения	
	МС	РС	МС	РС
a	0.50	0.53	1.0	0.54
b	0.50	-0.48	0.99	-0.44
с	0.17	0.04	1.00	0.06
d	0.64	0.24	0.98	0.24
e	0.54	-0.20	1.00	-0.22
f	0.40	-0.04	1.00	-0.05

Таблица 1. Результат уточнения в обратном пространстве случайно сгенерированных 100-атомных моделей по сравнению с экспериментальными данными при разрешении 40Å (49 рефлексов).

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.

Эти пессимистических вывода могут быть частично лва компенсированы наблюдением о том, что, по-видимому, наборы фаз (варианты), соответствующие хорошим значениям критерия отбора, не заполняют "конфигурационное пространство" случайно, а формируют небольшое число компактных областей в конфигурационном пространстве, одна из которых достаточно близка к правильному решению. Так, из Рис.3 видно, что отобранные по степени корреляции модулей варианты не распределены равномерно по величине фазовой корреляции, а распадаются на несколько групп. Это также может быть видно из Рис.5, где при отборе вариантов по корреляции модулей образуются две группы моделей: одна с корреляциями фаз в районе 0., и другая с корреляциями фаз около 0.5. Из этого следует, что может быть сформулирована задача изучения и описания всех главных областей в многомерном конфигурационном пространстве, значениям отбора. соответствующих хорошим критерия Подходящим инструментом для этого является многомерный кластерный анализ.

2.3. Переформулировка задачи определения фаз.

Проведенное тестирование заставило модифицировать 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 рассчитанных модулей структурных факторов экспериментально измеренным значениям (MC):

- наилучшее соответствие модулей может достигаться при неправильных фазах;
- модель, воспроизводящая наилучшим образом фазы, может не выделяться среди других по точности воспроизведения модулей;



Рис. 4. Значения коэффициентов корреляции модулей (МС) и фаз (РС), рассчитанных по случайно сгенерированным 2-атомным моделям. Каждая точка представляет одну модель.



Рис.6. Схематическое представление характера связи фазовой корреляции с корреляцией 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.6. 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.

3. FAM - метод.

Общая схема метода приведена на Рис.7.

3.1.. 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



Рис.7. Общая схема FAM - метода.

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.

Следует заметить, что на практике для отбора моделей могут применяться одновременно несколько разных критериев отбора, рассчитанных к тому же, может быть, по разным зонам разрешения. Также надо отметить, что в процессе генерации координат псевдоатомов на эти координаты могут налагаться дополнительные ограничения, например, препятствующие чрезмерному сближению центров атомов.

3. 2. 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 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:

47

- 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 энантиоморфа, если это допустимо);

• для выровненных распределений вычисляется расстояние

$$dist\left(\left\{\varphi_{hkl}^{j}\right\},\left\{\varphi_{hkl}^{m}\right\}\right) = dist\left(\rho^{j},\rho^{m}\right)$$
$$= \sqrt{\frac{\int_{V} \left[\rho^{j}(x,y,z) - \rho^{m}(x,y,z)\right]^{2} dV}{\int_{V} \left[\rho^{j}(x,y,z)\right]^{2} dV \int_{V} \left[\rho^{m}(x,y,z)\right]^{2} dV}}$$
(10)

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 сдвинутому. Наличие кристаллографической объекта, но по-разному 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 (см. Приложение).

На Рис.8 показано дерево, полученное в тесте с AspRS-комплексом при разрешении 50Å, а на Рис.9 возможный вариант его разбивки на кластеры. Цифры, присутствующие на рисунке у вершин кластеров, показывают среднюю корреляцию фаз из этих кластеров с точными. Следует подчеркнуть, что построение дерева осуществляется программой автоматически, в то время как анализ полученного дерева и принятиерешения о том, сколько и какие кластеры следует выделить, остается за исследователем. Для облегчения процедуры выбора перспективных кластеров была создана специальная интерактивная компьютерная программа, позволяющая рассчитывать и рисовать дерево и выделять кластеры для дальнейшего сравнения и отбора.

На Рис.10 показано распределение вариантов по величине корреляции фаз внутри каждого из выделенных кластеров и среди всех анализируемых вариантов вместе (как и на Рис.3). Из рисунка видно, что резкие пики, проявившиеся на Рис.3, действительно соответствуют различным кластерам близких между собой наборов фаз.

На Рис.8 показано необработанное кластерное дерево, а помеченные и проинтерпретированные варианты - на Рис.9. Здесь выделены 3 главных кластера, названных С11, С12 и С13 (Рис.9). Изучение соответствующих синтезов Фурье позволяет выбрать один из них (С12), основываясь на том критерии, что на синтезе не должно быть плотности на элементах симметрии (здесь на осях х=у, z=0). С12 - самый большой кластер, и он может быть разделен далее на 2 новых, С21 и С22. Выбор из этих кластеров может быть снова сделан на основании того же критерия. Таким образом, правильный отбор может быть проведен без всякого сравнения с правильными фазами. Когда процесс закончился, то синтезы, соответствующие "кластерным" наборам фаз, были сравнены с правильными фазами при разрешении 50Å. При продвижении вниз по кластерному дереву с первого уровня на второй



Рис. 8. Кластерное дерево, полученное в тестах с RS - комплексом.



Рис.9. Схематическое представление кластерного дерева, приведенного на Рис.8.



Рис.10. Распределение вариантов по величине корреляции модулей в отдельных кластерах. Обозначения кластеров те же, что на Рис.9.



Рис.11. Кластерное дерево, полученное в тестах с RS - комплексом при разрешении 40Å.

корреляция синтезов возросла с 77 до 93%; однако дальнейшее продвижение уже не давало никакого увеличения корреляции.

Как показано на Рис.10, эти кластеры были уже видны перед процедурой кластеризации на Рис.3 после отбора по корреляции модулей, где они появились как отдельные пики на разном расстоянии от правильного решения. Рис.10 демонстрирует разложение гистограммы (без нормализации) на отдельные части, соответствующие разным кластерам. Прослеживается явная корреляция между тремя главными пиками и тремя главными кластерами С11, С12 и С13. Далее, кластер С12 дает более тонкую структуру пика, соответствующую его вкладу в более маленькие; также показан наилучший кластер С22.

Процесс кластеризации является. тем самым. полезным R распознавании правильного решения. Интересно посмотреть, как слияние и взвешивание фазовых наборов повышает их качество. Это происходит при помощи двух механизмов: приписывания маленьких весов тем структурным факторам, чьи фазы неколлинеарны у разных моделей, и погашение ошибок (ошибки усреднения отдельных измерений меньше, чем среднее абсолютных величин ошибок каждого измерения). Это наиболее хорошо видно для кластера СОО. Фазовая корреляция кластера - 77%, в то время как средняя фазовая корреляция отдельных наборов фаз - только 26%. Также фазовая корреляция кластера С22 - 93%, в то время как средняя фазовая корреляция входящих в него наборов - 77%.

На Рис. 11 приведен еще один пример кластерного 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.

3.3. Обработка кластера.

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

$$m_{hkl} \exp\left[i\varphi_{hkl}^{best}\right] = \frac{1}{N} \sum_{m=1}^{N} \exp\left[i\varphi_{hkl}^{n}\right], \qquad (11)$$

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 может быть определена соответствующая ему маска молекулы как область наибольших значений взвешенного синтеза Фурье $\{m_h, F_h^{obs}, \varphi_h^{best}\}$.

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.

3.4. 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.

При проведении тестирования кластеров использовалась R-free идеология, предложенная А.Брюнгером: вовлеченные в работу рефлексы 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.

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.).

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.

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).

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

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, кода 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. 12 показаны гистограммы, полученные при анализе четырех кластеров, отобранных при анализе дендрограммы, приведенной на Рис.11. 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 v2, который обладает и максимальной корреляцией с правильными фазами.

На Рис.13 представлена попытка повторить тот же тест для того, чтобы расщепить кластер v2 на два более мелких. Здесь мы не видим такой отчетливой картины, как на предыдущем рисунке. Более того, гистограмма для худшего кластера v2b выглядит лучше, чем для лучшего кластера v2a. Таким образом, этот тест различает кластеры, резко различающиеся по качеству, но не является достаточно чувствительным, чтобы разрешать неоднозначность в случае исследования близких кластеров.

Маска, отвечающая кластеру v2, была использована на следующем шаге для генерации 1000 FAM - моделей, состоящих из 5 глобул. Пять отобранных для анализа кластера (выделенные из 100 вариантов, отобранных по корреляции модулей) отмечены на Рис.14. Результаты тестирования этих представлены на Рис.15. Здесь опять лучшему кластеров кластеру соответствует лучшая гистограмма, но следует отметить, что остальные кластеры могут быть ранжированы однозначно сравнения не ИЗ соответствующих им гистограмм.

На завершающей стадии теста маска, отвечающая кластеру w1a, была использована для генерации в нее 1000 FAM - моделей, состоящих из 100 глобул каждая, и отобранные по корреляции модулей 100 лучших вариантов

были усреднены. На Рис.16 показаны сечения элементарной ячейки с нанесенными контурами области молекулы, определенной FAM- методом, и реальные позиции С_α- атомов синтетазы и атомов Р в тРНК.



Рис.12. Распределение 100-атомных псевдомоделей по величине корреляции модулей при генерации атомов в альтернативные области элементарной ячейки. Цифры в рамке показывают корреляцию центроидных фаз для каждого из кластеров с точными фазами SR - комплекса.



Рис.13. то же, что на предыдущем рисунке, но при другом выборе кластеров.



Рис.14.Кластерное дерево, полученное в результате генерации 5-атомных псевдомоделей в маску, отвечающую кластеру v2 на Рис.11.



Рис.15. Распределение 100-атомных псевдомоделей по величине корреляции модулей при генерации атомов в альтернативные области элементарной ячейки. Цифры в рамке показывают корреляцию центроидных фаз для каждого из кластеров с точными фазами SR - комплекса.



Рис. 16. Границы комплекса аминоацил-тРНК-синтетазы с тРНК определенные FAM- методом и реальные позиции C_{α} и P атомов в ячейке кристалла.

ГЛАВА III. ОПРЕДЕЛЕНИЕ СТРУКТУРЫ РИБОСОМНОЙ ЧАСТИЦЫ 508 ИЗ Thermus thermophilus ПРИ РАЗРЕШЕНИИ ПОРЯДКА 40Å

Разработанная методика была применена к ab-initio определению структуры рибосомной 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; Васильев, 1974), трехмерной реконструкции крио- электронно -микроскопических (ЭМ) - изображений (Frank et al., 1995; Stark et al., 1995; Malhotra et al., 1998) 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 (Ban et al., 1998). В данной работе была предпринята попытка получения структурной информации, независимой от имеющихся электронномикроскопических моделей, опирающейся лишь на данные рентгеновского рассеяния кристаллами T50S частицы.

В работе использовались экспериментальные данные, полученные в лаборатории A.Yonath (Volkmann *et al.*, 1990). Кристаллы T50S имели симметрию пространственной группы $P4_12_12$ / $P4_32_12$ и размеры ячейки a=b=498Å, c=198Å. 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 принципиальную неоднозначность в выборе энантиомера, которая должна быть позднее разрешена из других соображений (например, стереохимических). Для данной работы была выбрана группа $P4_12_12$.

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 40Å.

1. Определение структуры Т508.

Работа была начата с генерации двух псевдоатомов BO всю элементарную ячейку. Значения параметра В, задающего размер глобулы, было определено как 60 000Å². Для работы имеющийся набор модулей структурных факторов разрешения до 30Å был разбит на оболочки в обратном пространстве (см. табл. 2). При этом половина рефлексов была объявлена контрольными, и их модули не участвовали в вычислении величин критериев отбора. Отбор на первом шаге велся по коэффициенту корреляции, вычисляемому в зоне 80-9999Å по рабочим рефлексам. В результате генерации 1000 моделей было отобрано 150 из них, которые и были подвергнуты кластерному анализу. Кластерный анализ позволил выявить два кластера. Для сравнения кластеров здесь и далее применялся анализ связности соответствующих синтезов Фурье и стандартный Монте Карловский тест (см. раздел II.4.4.). При проведении стандартного теста производилась генерация 100-атомных моделей в отвечающую кластеру маску, занимающую 30% объема элементарной ячейки. Анализировались гистограммы распределения величин критериев отбора по каждой зоне разрешения (в отдельности по рабочим и контрольным рефлексам). Кластер b2 по всем показателям превосходил альтернативный.

В дальнейшем на каждом следующем шаге применения FAMпроцедуры генерация глобул велась в область, определяемую как 30% точек, имеющих максимальные значения электронной плотности на синтезах Фурье, построенных с фазами, определенными на предыдущем шаге. Количество глобул в модели и используемая зона разрешения постепенно увеличивались по мере продвижения. Некоторые характеристики первых этапов работы приведены в таблице 3.

Зона разрешения	Количество рефлексов		
	всего	рабочих	контрольных
80-9999	41	24	17
60-9999	87	52	35
50-9999	141	74	67
40-9999	266	131	135
60-80	46	28	18
50-60	54	22	32
40-50	125	57	68

Таблица 2. Распределение рефлексов по зонам.

К-во глобул	В	Разрешение
2	60 000	80-9999
4	30 000	80-9999
6	30 000	60-9999
8	22 000	60-9999
8	22 000	50-9999
10	22 000	50-9999
10	22 000	40-9999
20	20 000	40-9999
20	20 000	40-9999
1		

Таблица 3. Характеристики первых шагов по применению FAM- процедуры.

Процедура шла таким образом до тех пор, пока соображения связности и все гистограммы стандартного теста указывали на один и тот же кластер в качестве наилучшего. Когда кластеры сблизились настолько, что показатели стали отличаться очень мало и, как следствие, стали указывать на разные кластеры в качестве наилучшего, процедура выбора маски для генерации глобул была изменена:

- по фазам, определенным на предыдущем шаге, рассчитывался синтез Фурье (в качестве модулей брались экспериментальные);
- к синтезу применялась двухшаговая процедура фильтрации (Urzhumtsev, Lunin & Luzianina, 1989), позволяющая автоматически строить маску области.

Кроме того, для уточнения значений полученных фаз применялась обычная процедура модификации электронной плотности (Podjarny, Rees & Urzhumtsev, 1996) с кубической и кусочно-линейной модифицирующей функциями. Для оценки точности получаемых фаз и расчета весов для взвешенных синтезов применялась процедура, использующая максимизацию маргинальной (рассчитанной только по контрольным рефлексам) функции правдоподобия (Lunin, Urzhumtsev, 1984; Read, 1986; Lunin, Skovoroda, 1995; Urzhumtsev, Skovoroda, Lunin, 1996).

Было проведено около 200 шагов процедуры, в результате чего был определен набор фаз разрешения 40Å и рассчитаны соответствующие синтезы Фурье. Далее мы будем ссылаться на этот набор фаз как на FAM-фазы. Некоторые изображения, дающие наглядное представление о форме частицы, приведены на Рис.17-19.

2. Сравнение с электронно-микроскопической моделью.

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 сравнение независимо полученных результатов. Данные поступили в виде файла, содержащего модули и фазы



Рис. 17. Синтез Фурье разрешения 40Å для рибосомной частицы T50S, рассчитанный с экспериментальными модулями структурных факторов и определенными FAM - методом фазами



Рис. 18. Синтез Фурье разрешения 40Å для рибосомной частицы T50S, рассчитанный с экспериментальными модулями структурных факторов и определенными FAM - методом фазами



Рис. 19. Синтез Фурье разрешения 40Å для рибосомной частицы T50S, рассчитанный с экспериментальными модулями структурных факторов и определенными FAM - методом фазами

структурных факторов, позволяющих рассчитать бинарную функцию, задающую маску частицы согласно данным трехмерной электронномикроскопической реконструкции. Для того, чтобы привести данные к виду, удобному для визуального сравнения, была проделана следующая работа:

- рассчитана указанная выше бинарная функция, задающая маску частицы; при этом разрешение было понижено до 40Å, что дало несколько сглаженную модель частицы;
- эта маска была смоделирована посредством большого числа атомов, равномерно заполняющих приведенную оболочку;
- полученная псевдомодель была размещена в элементарной ячейке кристалла таким образом, чтобы обеспечить наилучшее соответствие синтезу Фурье электронной плотности, рассчитанному с FAM-фазами;
- по размещенной таким образом псевдомодели были рассчитаны модули и фазы структурных факторов, которые далее именуются как ЕМ-модули и ЕМ-фазы.

Поиск наилучшего местоположения и ориентации псевдомодели осуществлялся двумя путями. В первом случае работа была выполнена вручную на графической станции. Для более объективного анализа была создана специальная программа вписывания модели в плотность в прямом пространстве, применение которой дало те же результаты. В процессе вписывания рассматривались всевозможные способы расположения и ориентации модели (на некоторой сетке, измельчающейся по ходу работы). Для каждого способа определялся процент атомов псевдомодели, не попадающих область молекулы, определенную FAМ-методом. В Оптимальными считались такое положение центра псевдомодели и ее ориентация, при которых этот процент был минимальным.

После того, как таким образом были определены ЕМ-фазы, был рассчитан и комбинированный синтез с экспериментальными модулями структурных факторов и ЕМ-фазами. Этот синтез оказался весьма похожим на синтез, рассчитанный с экспериментальными модулями и FAM-фазами.



Рис. 20. Синтезы Фурье для T50S субъединицы, рассчитанные с экспериментальными модулями и фазами: слева - определенными FAM - методом; справа - рассчитанными по электронно-микроскопической модели.

Несколько различных ракурсов этих синтезов приведены на Рис.20.

На Рис.21-23 приведено сравнение трех синтезов, показанных в одинаковой ориентации:

- рассчитанного с экспериментальными модулями и FAM-фазами;
- рассчитанного с ЕМ-модулями и ЕМ-фазами, то есть воспроизводящего электронно-микроскопическую модель, но в измененной ориентации и сглаженную понижением разрешения до 40Å;
- комбинированный синтез с экспериментальными модулями и ЕМ-фазами.
 Из этих рисунков видно, что замена ЕМ-модулей на экспериментальные приводит к сближению синтезов, рассчитанных с FAM- и ЕМ-фазами.



Рис. 21. Сравнение сглаженной до разрешения 40Å маски ЭМ - модели с синтезом Фурье, рассчитанным с экспериментальными модулями и фазами, полученными FAM -методом, и комбинированным синтезом (экспериментальные модули и ЭМ - фазы).



Рис. 22. Сравнение сглаженной до разрешения 40Å маски ЭМ - модели с синтезом Фурье, рассчитанным с экспериментальными модулями и фазами, полученными FAM -методом, и комбинированным синтезом (экспериментальные модули и ЭМ - фазы).


Рис. 23. Сравнение сглаженной до разрешения 40Å маски ЭМ - модели с синтезом Фурье, рассчитанным с экспериментальными модулями и фазами, полученными FAM -методом, и комбинированным синтезом (экспериментальные модули и ЭМ - фазы).

ПРИЛОЖЕНИЕ. Оптимальное выравнивание наборов фаз.

1. Введение

При оценке близости двух наборов фаз обычно рассчитывается их среднее отклонение. Чтобы учесть роль сильных рефлексов, обычно используют метрику, где среднее расхождение фаз взвешено величинами модулей, или коэффициент корреляции карт.

$$C = \int_{V} \rho_1(\mathbf{r}) \rho_2(\mathbf{r}) dV_{\mathbf{r}} \left/ \left(\int_{V} \rho_1^2(\mathbf{r}) dV_r \int_{V} \rho_2^2(\mathbf{r}) dV_r \right)^{1/2}$$
(12)

Коэффициент корреляции фаз может быть также выражен в терминах структурных факторов как

$$C = \alpha \sum_{\mathbf{s}} F_1(\mathbf{s}) F_2(\mathbf{s}) \exp\{i[\varphi_2(\mathbf{s}) - \varphi_1(\mathbf{s})]\}$$

$$= \alpha \sum_{\mathbf{s}} F_1(\mathbf{s}) F_2(\mathbf{s}) \cos[\varphi_2(\mathbf{s}) - \varphi_1(\mathbf{s})]$$

$$\alpha = \left(\sum_{\mathbf{s}} F_1^2(\mathbf{s}) \sum_{\mathbf{s}} F_2^2(\mathbf{s})\right)^{-1/2}$$
(14)

Здесь и ниже мы предполагаем, что все карты и суммы структурных факторов вычислены без члена F_{000} .

Решая проблему фаз *ab-initio*, например, путем генерации случайного набора фаз, необходимо осознавать, что полученные наборы фаз могут отвечать разному положению начала координат и выбору энантиоморфа, поэтому перед вычислением разности фаз должно быть проведено предварительное выравнивание карт (или соответствующих наборов фаз). В качестве критерия наилучшего выравнивания может быть использован коэффициент корреляции карт (12).

Если симметрия пространственной группы допускает только конечное число различных положений начала координат, можно легко просмотреть все варианты начала координат и выбора энантиоморфа, чтобы выбрать лучший из них. Когда же допускается произвольный сдвиг вдоль одной или трех осей (группа P1), поиск лучшего выравнивания несколько более труден, и это заставляет иногда искать другие, менее прямые способы определения близости фаз.

Однако, как будет показано ниже (и как происходит при расчете функции трансляции в методе молекулярного замещения), коэффициент корреляции карт, рассматриваемый как функция сдвига начала координат, имеет структуру ряда Фурье. В этом случает может быть применен алгоритм быстрого преобразования Фурье, чтобы вычислить массив коэффициентов корреляции, отвечающим конечному числу сдвигов начала координат, и выбрать лучший из них. Эти сдвиги задаются узлами некоторой сетки в прямом пространстве. При необходимости найденные значения могут быть затем уменьшены методами локальной минимизации, однако представляется, что для практических приложений поиск на достаточно подробной сетке дает приемлемую точность.

2. Выравнивание карт.

2.1. Группа Р1.

Предположим, что карта $\rho_1(\mathbf{r})$ фиксирована и что мы ищем такое положение другой карты $\rho_2(\mathbf{r})$, при котором она наилучшим образом совпадает с первой, используя в качестве критерия коэффициент корреляции карт (13). В качестве базовых возможных преобразований карты $\rho_2(\mathbf{r})$ мы рассмотрим:

• - сдвиг начала координат

$$\rho_2'(\mathbf{r}) = \rho_2(\mathbf{r} + \mathbf{u}) \tag{15}$$

• - смену энантиоморфа

$$\rho_2'(\mathbf{r}) = \rho_2(-\mathbf{u}) \tag{16}$$

Соответствующие преобразования фаз имеют вид:

$$\varphi_2'(\mathbf{s}) = \varphi_2(\mathbf{s}) - 2\pi(\mathbf{s}, \mathbf{u})$$
⁽¹⁷⁾

$$\boldsymbol{\varphi}_{2}(\mathbf{s}) = -\boldsymbol{\varphi}_{2}(\mathbf{s}) \tag{18}$$

Коэффициент корреляции фаз для фиксированной первой и сдвинутой второй карты имеет вид:

$$C(\mathbf{u}) = \alpha \sum_{\mathbf{s}} F_1(\mathbf{s}) F_2(\mathbf{s}) \exp\{i[\varphi_2(\mathbf{s}) - \varphi_1(\mathbf{s})]\}$$

= $\alpha \sum_{\mathbf{s}} F_1(\mathbf{s}) F_2(\mathbf{s}) \exp\{i[\varphi_2(\mathbf{s}) - \varphi_1(\mathbf{s})] \exp[-2\pi i(\mathbf{s}, \mathbf{u})]\}$ ⁽¹⁹⁾

Формула (19) имеет структуру преобразования Фурье, вычисленного с коэффициентами

$$H(\mathbf{s})\exp[i\psi(\mathbf{s})] = F_1(\mathbf{s})F_2(\mathbf{s})\exp\{i[\varphi_2'(\mathbf{s}) - \varphi_1(\mathbf{s})]\}$$
(20)

Алгоритм быстрого преобразовангия Фурье позволяет эффективно вычислить множество коэффициентов корреляции для сдвигов начала координат, ограниченных узлами некоторой сетки в элементарной ячейке.

Когда сдвиг начала координат дополняется сменой энантиоморфа (мы считаем в этом случае, что переход к энантиоморфу есть первое преобразование, а сдвиг начала координат - второе), они приводят к множеству коэффициентов корреляции, рассчитанных как преобразование Фурье с коэффициентами

$$H(\mathbf{s})\exp[i\psi(\mathbf{s})] = F_1(\mathbf{s})F_2(\mathbf{s})\exp\{i[-\varphi_2(\mathbf{s})-\varphi(\mathbf{s})]\}$$
(21)

Выбор оптимальной точки решетки (и энантиоморфа) достигается при этом прямым перебором в двух трехмерных множествах коэффициентов корреляции.

2.2. Нетривиальные пространственные группы

При работе с нетривиальными пространственными группами не все сдвиги начала координат сохраняют симметрию. аналогично, переход к энантиоморфу может быть запрещен симметриями пространственной группы.

В этом случае только те сдвиги, которые разрешены симметрией пространственной группы, могут рассматриваться при поиске наилучшего выравнивания. Как показано ниже, нет необходимости вычислять трехмерное преобразование Фурье в этом случае.

Если симметрия пространственной группы оставляет только конечное число допустимых вариантов выбора начала координат, то только они и должны быть проверены.

Другой случай наблюдается, когда множество допустимых сдвигов сводится к конечному числу вариантов в двух измерениях и произвольному сдвигу в третьем измерении. Предположим для простоты, что произвольный

сдвиг разрешен вдоль оси z. Пусть $\mathbf{U} = \left\{ \mathbf{u}^m \right\}_{m=1}^M$ есть множество разрешенных сдвигов в плоскости ху: $\mathbf{u}^m = \left(x^m, y^m, 0 \right)$. (Например, для группы Р2 **u** содержит четыре сдвига (0,0,0), (0,1/2,0), (1/2,0,0), (1/2,1/2,0); для группы Р6 оно содержит только элемент (0,0,0)). Любой допустимый сдвиг в этом случае может быть представлен как

$$(u_x, u_y, u_z) = (u_x^m, u_y^m, 0) + (0, 0, u_z)$$
 (22)
где $(u_x^m, u_y^m, 0) \in \mathbf{U}$ и $u_z \in (0, 1).$

Формула (19) может быть сведена в этом случае к

$$C^{m}(u_{z}) = \alpha \sum_{l} H^{m}(l) \exp[i\psi^{m}(l)] \exp(-2\pi i l u_{z})$$
(23)

c

$$H^{m}(l)\exp[i\psi^{m}(l)] = \sum_{h,k} F_{1}(hkl)F_{2}(hkl)$$

$$\times \exp\left\{i[\varphi_{2}(hkl) - \varphi_{1}(hkl)] - 2\pi i \left(hu_{x}^{m} - ku_{y}^{m}\right)\right\}$$
(24)

Итак, в этом случае возможно выполнить одномерное преобразование Фурье для каждого сдвига из множества U (и перехода к энантиоморфу, если он допустим) и искать затем наилучшее значение коэффициента корреляции

среди М (или 2М, если переход к энантиоморфу допустим) одномерных множеств.

Ааналогичные формулы могут быть написаны также, если произвольный сдвиг разрешен вдоль оси, отличной от z. Разница только в том, по каким индексам идет суммирование в формуле (24).

3. Использование асимметричной части множества структурных факторов

Обычно при практической работе используется только асимметричная часть набора структурных факторов. Мы сведем сейчас суммирование в (13), (14) и (24) к асимметричной части множества структурных факторов. Пусть S - полное множество структурных факторов, используемое при расчете карт

 $\rho_1(\mathbf{r})$ и $\rho_2(\mathbf{r});$ пусть $\Gamma = \{(G_v, t_v)\}_{v=1}^n$ - соответствующая пространственная группа, и S_{asym} - асимметричная часть множества S, т.е.

- все структурные факторы набора S могут быть сгенерированы из элементов набора S_{asym} при помощи преобразований симметрии группы Г и эрмитовой симметрии;
- - для каждого $\mathbf{s} \in S_{asym}$ это единственная точка орбиты $\left\{ \pm G^T \mathbf{s} \right\}_{\nu=1}^n$, которая принадлежит S_{asym} .

Если мы обозначим

$$m(\mathbf{s})$$
-количество различных точек орбиты $\left\{\pm G^T \mathbf{s}\right\}_{\nu=1}^n$ (25)

то выражение (14) может быть сведено к

$$\boldsymbol{\alpha} = \left\{ \sum_{\mathbf{s} \in S_{asym}} m(\mathbf{s}) F_1(\mathbf{s})^2 \sum_{\mathbf{s} \in S_{asym}} m(\mathbf{s}) F_2(\mathbf{s})^2 \right\}^{-1/2}$$
(26)

Чтобы свести суммирование в (13) и (24) к S_{аsym}, необходимо принимать во внимание, как пространственная группа ограничивает возможные сдвиги начала координат. Необходимо учитывать также наличие систематических погасаний. В разделе "Допустимые сдвиги начала координат" показано, что для непогасающих рефлексов

$$\exp\left[-2\pi i(\mathbf{s}, G_{\nu}\mathbf{u})\right] = \exp\left[-2\pi i(\mathbf{s}, \mathbf{u})\right]$$
(27)

для всех допустимых сдвигов **u** и всех $\nu = 1, ..., n$. Общая формула симметрии в обратном пространстве

$$F(G_{\nu}^{T}\mathbf{s})\exp[i\varphi(G_{\nu}^{T}\mathbf{s})] = F(\mathbf{s})\exp\{i\varphi(\mathbf{s}) - 2\pi(\mathbf{s},t_{\nu})\}$$
(28)

дает в этом случае

$$F_{1}(G_{\nu}^{T}\mathbf{s})F_{2}(G_{\nu}^{T}\mathbf{s})$$

$$\times \exp\left\{i\left[\varphi_{2}(G_{\nu}^{T}\mathbf{s})-\varphi_{1}(G_{\nu}^{T}\mathbf{s})\right]-2\pi i\left(G_{\nu}^{T}\mathbf{s},\mathbf{u}^{m}\right)\right\}$$

$$=F_{1}(\mathbf{s})F_{1}(\mathbf{s})\exp\left\{i\left[\varphi_{2}(\mathbf{s})-\varphi_{1}(\mathbf{s})\right]-2\pi i\left(\mathbf{s},G_{\nu}\mathbf{u}^{m}\right)\right\}$$

$$=F_{1}(\mathbf{s})F_{1}(\mathbf{s})\exp\left\{i\left[\varphi_{2}(\mathbf{s})-\varphi_{1}(\mathbf{s})\right]-2\pi i\left(\mathbf{s},\mathbf{u}^{m}\right)\right\}$$
(29)

поэтому все точки орбиты $\{G_{\nu}^{T}\mathbf{s}\}_{\nu=1}^{n}$ дают одинаковый вклад в рассматриваемые суммы. Это означает, что вместо (13) можно написать $C = \alpha \sum_{\mathbf{s} \in S_{asym}} m(\mathbf{s}) F_{1}(\mathbf{s}) F_{2}(\mathbf{s}) \cos[\varphi_{2}(\mathbf{s}) - \varphi_{1}(\mathbf{s})]$ (30)

где $m(\mathbf{s})$ - по-прежнему число различных точек орбиты $\{\pm G^T \mathbf{s}\}_{\nu=1}^n$ Сумма (24) рассчитывается для конкретного значения l, поэтому для l≠0 в сумме присутствует только половина орбиты $\{\pm G^T \mathbf{s}\}_{\nu=1}^n$, и мы можем записать сумму для $\mathbf{s} = (hkl) \in S_{asym}$ как

$$H^{m}(l)\exp[i\psi^{m}(l)] = \sum_{h,k} p(hkl)F_{1}(hkl)F_{2}(hkl)$$

$$s \in S_{asym}$$

$$(31)$$

$$\times \exp\left\{i[\varphi_{2}(hkl) - \varphi_{1}(hkl)] - 2\pi i \left(hu_{x}^{m} - ku_{y}^{m}\right)\right\}$$

где p(hkl) - это количество различных точек на орбите $\{\pm G^T \mathbf{s}\}_{\nu=1}^n$ с фиксированным значением l. Для l=0 это равно мультипликатору m(hk0), определенному в (25). В других случаях $p(\mathbf{s}) = m(\mathbf{s}) / 2$.(подчеркнем, что мы рассматриваем суммирование (24) для пространственных групп, допускающих произвольный сдвигом вдоль оси z).

4. "Перевернутые" карты

Ранее упоминалось, что некоторые методы нахождения фаз при очень низком разрешении имеют тенденцию иногда выдавать в качестве решения "перевернутый" синтез $--\rho(\mathbf{r})$ вместо настоящего. Другими словами, может быть найдена область растворителя вместо области молекулы. Мы оставляем дискуссию о причинах этого явления упоминаемым работам, а здесь только рассмотрим, как такая дополнительная "степень свободы" для разрешенных преобразований карт влияет на процедуру выравнивания карт.

Возможность смены знака синтеза (или, что то же самое, увеличение всех фаз на π) означает, что мы должны искать максимальное значение $\max |C(\mathbf{t})|$ вместо $\max C(\mathbf{t})$. Поскольку

$$\max_{\mathbf{t}} |C(\mathbf{t})| = \max \left\{ \max_{\mathbf{t}} C(\mathbf{t}) - \min_{\mathbf{t}} C(\mathbf{t}) \right\},\$$

то мы должны искать как максимальное, так и минимальное значение коэффициента корреляции и брать в качестве оптимального то, у которого абсолютная величина больше. Поэтому при рассмотрении карт с дополнительной возможностью смены знака синтеза нет необходимости вычислять новые значения коэффициентов корреляции, а изменяется только поиск оптимального значения.

5. Разрешенные сдвиги начала координат.

Пусть карта $\rho(\mathbf{r})$ обладает симметрией группы $\Gamma = \{(G_v, t_v)\}_{v=1}^n$, т.е. $\rho(G_v \mathbf{r} + t_v) = \rho(\mathbf{r})$ для каждого $\mathbf{r}, v = 1, ..., n$. Пусть $\Gamma^t = \{(E, \mathbf{t})\}_{\mu=1}^m$ будет ее подгруппа нетривиальных трансляций . Подгруппа Γ^t содержит только тождественное преобразование (m=1) для примитивной решетки. Для двух векторов \mathbf{u} и \mathbf{v} мы запишем $u = v \big|_{\text{mod } \Gamma^t}$ если существует $(E, \mathbf{t}_{\mu}) \in \Gamma^t$ такое, что $\mathbf{u} = \mathbf{v} + \mathbf{t}_{\mu} \big|_{\text{mod } l}$.

Мы говорим, что **u** есть разрешенный сдвиг начала координат, если карта $\rho'(\mathbf{r}) = \rho(\mathbf{r} + \mathbf{u})$ имеет ту же группу симметрий Γ , как и $\rho(\mathbf{r})$. Аналогично, мы говорим, что Γ допускает переход к энантиоморфу, если $\rho'(\mathbf{r}) = \rho(-\mathbf{u})$ имеет ту же груупу симметрии Γ . Множество разрешенных сдвигов начала координат определяется следующей Леммой. Лемма 1. Функция $\rho(\mathbf{r} + \mathbf{u})$ имеет ту же группу симметрии, что и $\rho(\mathbf{r})$, тогда и только тогда, когда

$$(G_{\nu} - E)\mathbf{u} = 0 \Big|_{\text{mod }\Gamma^{t}}$$
для всех $\nu = 1,...,n$ (32)

Доказательство.

Пусть Γ - пространственная группа $ho(\mathbf{r})$ и \mathbf{u} такой, что $\mathbf{u} = G_v \mathbf{u} - \mathbf{t}_\mu$, где $(E, \mathbf{t}_\mu) \in \Gamma^t$. Тогда

$$\rho'(G_{\nu}\mathbf{r} + \mathbf{t}_{\nu}) = \rho(G_{\nu}\mathbf{r} + \mathbf{t}_{\nu} + \mathbf{u}) =$$

$$\rho(G_{\nu}(\mathbf{r} + \mathbf{u}) + \mathbf{t}_{\nu} - \mathbf{t}_{\mu}) =$$

$$\rho(G_{\nu}(\mathbf{r} + \mathbf{u}) + \mathbf{t}_{\nu}) =$$

$$\rho(\mathbf{r} + \mathbf{u}) = \rho'(\mathbf{r})$$

Поэтому $ho'(\mathbf{r})$ имеет все симметрии из Γ .

Наоборот, пусть $ho'(\mathbf{r})$ обладает некоторой симметрией из Γ :

 $\rho'(G_{\nu}\mathbf{r} + \mathbf{t}_{\nu}) = \rho'(\mathbf{r})$ для каждого \mathbf{r} , т.е. $\rho'(G_{\nu}\mathbf{r} + \mathbf{t}_{\nu} - \mathbf{u}) = \rho'(\mathbf{r} - \mathbf{u})$.Полагая $\mathbf{r} - \mathbf{u} = \mathbf{v}$ мы имеем $\rho[G_{\nu}(\mathbf{v} + \mathbf{u}) + \mathbf{t}_{\nu} - \mathbf{u}] = \rho(\mathbf{v})$ или $\rho[G_{\nu}\mathbf{v}_{\nu} + \mathbf{t}_{\nu} + (G_{\nu} - E)\mathbf{u}] = \rho(\mathbf{v})$ для каждого \mathbf{v} . Это означает, что преобразование $[G_{\nu}\mathbf{v}_{\nu} + \mathbf{t}_{\nu} + (G_{\nu} - E)\mathbf{u}] \in \Gamma$, и поэтому его вектор трансляции может отличаться от \mathbf{t}_{ν} только на \mathbf{t}_{μ} (с некоторым $\mu = 1, ..., m$. Поэтому $(G_{\nu} - E)\mathbf{u} = \mathbf{t}_{\mu}$, где $(E, \mathbf{t}_{\mu}) \in \Gamma^{t}$.

Теперь мы готовы доказать (27). Для пространственной группы с примитивной решеткой (т.е. когда Γ содержит только тождественное преобразование (E,0)) это немедленно следует из $(G_v - E)\mathbf{u} = 0|_{\text{mod }l}$. Если решетка центрированная, т.е. Γ содержит преобразование $(E, \mathbf{t}_{\mu}), \mathbf{t}_{\mu} \neq 0|_{\text{mod }l}$, то $F(\mathbf{s}) \neq 0$ возможно только при $(\mathbf{s}, \mathbf{t}_{\mu}) = 0$. Поэтому $G_{\mu}\mathbf{u} = \mathbf{u} + \mathbf{t}_{\mu}$ снова означает (27) для непогасающих рефлексов.

Аналогичные рассмотрения показывают, что необходимые и достаточные условия для функции $ho(-{f r})$ иметь ту же пространственную группу Γ , что и $ho({f r})$ - это

$$2\mathbf{t}_{\nu} = 0 \Big|_{\text{mod } \Gamma^{t}} \text{ для всех } \nu = 1, \dots, n.$$
(34)

Для примитивной решетки это означает, что все компоненты сдвигов в уравнениях симметрии должны быть равны 0 или 1/2. Проверка списка пространственных групп показывает, что для групп с центрированной решеткой и без инверсии в начале координат единственными исключениями из этого правила являются группы $I4_132$ и $I\overline{4}32$.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

- 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.
- 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.
- 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 фазовой проблемы, различных вариантов решения 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 рибосомной частицы T50S из *Thermus thermophilus* при разрешении 40Å.

СПИСОК ЛИТЕРАТУРЫ

Васильев В.Д. Морфологическая модель малой рибосомной 30S субчастицы. (1974). Доклады АН СССР,219, 994-995.

Дюран Б. & Оделл П. (1977). Кластерный анализ. М. Статистика.

- Петрова Т.Е., Лунин В.Ю., Лунина Н.Л., Сковорода Т.П. Выбор априорного распределения вероятностей координат атомов в макромолекулярных структурах на базе принципа максимального правдоподобия. (1998). Биофизика, принято в печать.
- Спирин А.С. (1986). Молекулярная биология. Структура рибосомы и биосинтез белка. М. Высшая школа.
- Уржумцев А.Г. Использование локального усреднения для анализа изображения макромолекул на картах электронной плотности. (1985). Препринт ОНТИ НЦБИ, АН СССР, Пущино.
- Фан Ликсин Структурная модель 30S рибосомной субчастицы thermus thermophilus с пространственным разрешением 35 Å по данным малоуглового рассеяния нейтронов. (1998). Диссертация на соискание ученой степени к.ф.-м.н. ИТЭБ РАН. Пущино.
- Чиргадзе Ю.Н., Невская Н.А., Фоменкова Н.П., Никонов С.В., Сергеев Ю.В., Бражников Е.В., Гарбер М.Б., Лунин В.Ю., Уржумцев А.Г., Вернослова Е.А. Пространственная структура гамма-кристаллина IIIb из хрусталика глаза теленка при разрешении 2,5Å. (1986). Доклады АН СССР, т.290, в.2, 492-495.
- Чиргадзе Ю.Н., Бражников Е.В., Никонов С.В., Фоменкова Н.П., Гарбер М.Б., Уржумцев А.Г., Лунин В.Ю., Чиргадзе Ю.Н., Некрасов Ю.В. Кристаллическая структура рибосомного фактора G из бактерий Thermus Thermophilus при низком разрешении. (1991). Доклады АН СССР, т.320, в.2, 488-491.
- Шевырев А.А. & Симонов В.И. Учет статистических ошибок в структурных амплитудах и эффекта обрыва ряда Фурье в синтезах распределения электронной плотности. (1981). Кристаллография, 26, 36-41.

- Andersson, K.M. & Hovmöller, S. Phasing Proteins at Low Resolution. (1996). Acta Cryst. D52, 1174-1180.
- Baker, D., Krukowski, A.E. & Agard, D. A. Uniqueness and the Ab Initio Phase Problem in Macromolecular Crystallography. (1993). Acta Cryst. D49, 186-192.
- Badger, J. & Caspar, D.L.D. Water structure in cubic crystal. (1991). Proc.Natl.Acad.Sci.USA 88, 622-626.
- Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R.A., Sweet, R., Frank, J., Moore, P.B. & Steitz, T.A. A 9 Å Resolution X-Ray Crystallographic Map of the Large Ribosomal Subunit. (1998). Cell 93, 1105-1115.
- Bentley, G.A., Lewit-Bentley, A., Liljas, L., Skoglund, U., Roth, M. & Unge, T. Structure of RNA in Satellite Tobacco Necrosis Virus. A Low Resolution Neutron Diffraction Study using H2O/2H2O Solven Contrast Variation. (1987). J.Mol.Biol. 194, 129-141.
- Blow, D.M. & Crick, F.H.C. The Treatment of Errors in the Isomorphous Replacement Method. (1959). Acta Cryst. 12, 794-802.
- Blow, D.M. & Rossmann, M.G. The Single Isomorphous Replacement Method. (1961). Acta Cryst. 14, 1195-1202.
- Bragg, W.L. & Perutz, M.F. The External Form of the Haemoglobin Molecule. I. (1952). Acta Cryst. 5, 277-283.
- Brazhnikov, E., Chirgadze, Yu., Aevarsson, A., Svensson, A., & Lunin, V. The Similarity of Results Obtained by Solvent Flattening and Low-Resolution Phase Retrieval for the Improvement of Protein Electron Density Map. (1993).
 3rd European Workshop of Biological Macromolecules, Centro di Cultura Scientifica "A.Volta", Como (Italy), May 21-25, 1993.
- Bricogne, G. Geometric Sources of Redundancy in Intensity Data and Their Use for Phase Determination. (1974). Acta Cryst, A30, 395-405.
- Bricogne, G. Maximum Entropy and the Foundations of Direct Methods. (1984). Acta Cryst. A40, 410-445.
- Bricogne, G. Direct Phase Determination by Entropy Maximization and Likelihood Rankind: Status Report and Perspectives. (1993). Acta Cryst. D49, 37-60.

- Bricogne, G. & Gilmore, C.J. A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. I. Theory, Algorithms and Strategy. (1990). Acta Cryst., A46, 284-297.
- Carbo, R., Leyda, L. & Arnau, M. How Similar is a Molecule to Another? An Electron Density Measure of similarity between Two Molecular Structures. (1980). Int. J. Quant. Chem. XVII, 1185-1189.
- Carter C.W., Jr., Crumbley, K.V., Coleman, D.E., Hage, F., & Bricogne, G. Direct Phase Determination for the Molecular Envelope of Tryptophanyl-tRNA Synthetase from Bacillus stearothermophilus by X-ray Contrast Variation. (1990). Acta Cryst. A46, 57-68.
- Chapman, M.S., Tsao Jun & Rossmann, M.G. Ab initio Phase Determination for Spherical Viruses: Parameter Determination for Spherical-Shell Models. (1992). Acta Cryst. A48, 301-312.
- Chirgadze, Yu.N., Nevskaya, N.A., Fomenkova, N.P., Nikonov, S.V., Sergeev, Yu.N., Brazhnikov, E.V., Garber, M.B., Lunin, V.Yu., Urzhumtsev, A.G. & Vernoslova, E.A. The space structure of gamma-crystallin IIIb from calf eye lens at 2.5Å resolution. (1986). Dokl.Akad.Nauk SSSR, 290, 2, 492-495.
- Cochran, W. A Relation between the Signs of Structure Factors. (1952). Acta Cryst., 5, 65-67.
- Cox, D.R. & Hinkley, D.V. (1974). Theoretical Statistics. Imperial College, London, Chapman and Hall.
- David, P.R. & Subbiah, S. Low-Resolution Real-Space Envelopes: the Application of the Condensing-Protocol Approach to the ab initio Macromolecular Phase Problem of a Variety of Examples. (1994). Acta Cryst. D50, 132-138.
- Filman, D.J., Miller, S.T. & Hogle, J.M. A Genetic Algorithm for the Ab initio Phasing of Icosahedral Viruses. In "ACA Annual Meeting, June 25 - July 1, 1994", p.67. INFORUM, Atlanta, Georgia, Atlanta Convention Center, 1994.
- Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Rademacher, M., Grassucci, R., Lata R.K. & Agarwal, R.K. A model for protein synthesis based on cryo-electron microscopy of the *E.coli* ribosome. (1995). *Nature* 376, 441-444.

- Harris, G.W. Fast Ab-initio Calculation of Solvent Envelopes for Protein Structures. (1995). Acta Cryst. D51, 695-702.
- Harrison, S.C. & Jack, A. Structure of Tomato Bushy Stunt Virus. III. Threedimensional X-ray Diffraction Analysis at 16 E Resolution. (1975). J.Mol.Biol. 97, 173-191.
- Harrison, R.W. Histogram Specification as a Method of Density Modification. (1988). J.Appl.Cryst. 21, 949-952.
- Hasek, J. & Schenk, H. On the Comparison of Different Sets of Structure Factor Phases. (1992). Acta Cryst. A48, 639-695.
- Hendrickson, W.A., Klippenstein, G.L. & Ward, K.B. Tetriary Structure of Myohemerythin at Low Resolution. (1975). Proc. Natl. Acad. Sci. USA, 72, 2160-2164.
- Jack, A. Direct Determination of X-ray Phases for Tobacco Mosaic Virus Protein using Non-crystallographic Symmetry. (1973). Acta Cryst. A29, 545-554.
- Jones, E.Y., Walker, N.P. & Stuart, D.I. Methodology Employed for the Structure Determination of Tumour Necrosis Factor, a Case of High Non-Crystallographic Symmetry. (1991). Acta Cryst. A47, 753-770.
- Klug, A. Joint Probability Distributions of Structure Factors and the Phase Problem. (1958). Acya Cryst. 11, 515-543.
- Lamzin, V.S. & Wilson, K.S. Automated Refinement of Protein Models. (1993). Acta Cryst, D49, 129-147.
- Leslie, A.G.V. A reciprocal-space method for calculating a molecular envelope using the algorithm of B.C.Wang. (1988). Acta Cryst. A43, 134-136.
- Lunin, V.Yu. & Urzhumtsev, A.G. Improvement of Protein Phases by Coarse Model Modification. (1984). Acta Cryst. A40, 269-277.
- Lunin, V.Yu. Use of the Information on Electron Density Distribution in Macromolecules. (1988). Acta Cryst. A44, 144-150.
- Lunin, V.Yu., Urzhumtsev, A.G. & Skovoroda, T.P. Direct Low-Resolution Phasing from Electron-Density Histograms in Protein Crystallography. (1990). Acta Cryst. A46, 540-544.

- Lunin, V.Yu. & Skovoroda, T.P. Frequency-Restrained Structure-Factor Refinement. I. Histogram Simulation. (1991). Acta Cryst. A47, 45-52.
- Lunin, V.Yu. & Vernoslova, E.A. Frequency-Restrained Structure-Factor Refinement. II. Comparison of Methods. (1991). Acta Cryst. A47, 238-243.
- Lunin, V.Yu. Electron-Density Histograms and the Phase Problem. (1993). Acta Cryst. D49, 90-99.
- Lunin, V.Yu. & Woolfson, M.M. Mean Phase Error and the Map Correlation Coefficient. (1993). Acta Cryst. D49, 530-533.
- Lunin, V.Yu., Lunina, N.L., Petrova, T.E., Vernoslova, E.A., Urzhumtsev, A.G. & Podjarny, A.D. On the ab-initio solution of the phase problem for macromolecules at very low resolution. (1994). Joint CCP4 and ESF-EACBM Newsletter 30, 37-44.
- Lunin,V.Y., Lunina, N.L., Petrova, T.E., Vernoslova, E.A., Urzhumtsev, A.G. & Podjarny, A.D. On the Ab initio solution of the Phase Problem for Macromolecules at Very Low Resolution: The Few Atoms Model Method. (1995). Acta cryst. D51, 896-903.
- Lunin, V.Yu. & Skovoroda, T.P. R-Free Likelihood Based Estimates of Errors for Phases Calculated from Atomic Models. (1995). Acta Cryst. A51, 880-887.
- Lunin, V.Yu. & Lunina, N.L. The map Correlation Coefficient for Optimally Superposed Maps. (1996). Acta Cryst. A52, 365-368.
- Lunin, V.Y. The likelihood based choice of priors in statistical approaches to the phase problem. (1998). Direct Methods for Solving Macromolecular Structures, edited by S.Fortier, pp.451-454. Kluwer Academic Publishers.
- Lunin,V.Y., Lunina, N.L., Petrova, T.E., Urzhumtsev, A.G. & Podjarny, A.D. On the Ab initio solution of the Phase Problem for Macromolecules at Very Low Resolution. II. Generalized Likelihood Based Approach to the Cluster Discrimination. (1998^a). Acta cryst. D54, 726-734.
- Lunin, V.Yu., Lunina, N.L., Petrova, T.E., Urzhumtsev, A.G. & Podjarny, A.D.
 Very low resolution ab-initio phasing. Problems and advances. (1998^b).
 Bulletin of the Czech and Slovak Crystallographic Association, 5, special issue A, ECM-18 abstracts, 131-132.

- Lunin, V.Y., Lunina, N.L. & Urzhumtsev, A.G. Seminvariant Density Decomposition and Connectivity Analysis and Their Application to Very Low Resolution Macromolecular Phasing. (1998). Acta Cryst. A, *submitted*.
- Luzzati, V., Mariani, P. & Delacroix, H. X-ray crystallography at macromolecular resolution: a solution of the phase problem. (1988). Makromol. Chem. Macromol. Symp. 15, 1-17.
- Main, P. Phase Determination Using Non-Crystallographic Symmetry. (1976). Acta Cryst. 23, 50-54.
- Main, P. A Formula for Electron Density Histograms for Equal-Atom Structures. (1990). Acta Cryst. A46, 507-509.
- Malhotra, A., Penczek, P., Agrawal, R.K., Gabashvili, I.S., Grassucci, R.A., Jünemann, N., Burkhardt, N., Nierhaus, K.H. & Frank, J. *Escherichia coli* 70 S Ribosome at 15 Å Resolution by Cryo-electron Microscopy: Localization of fMet-tRNA_f^{Met} and Fitting of L1 Protein. (1998) *J.Mol.Biol.*, 280, 103-116.
- Mathiesen, R.H. & Mo, F. Application of Known Triplet Phases in the Crystallographic Study of Bovine Pancreatic Trypsin Inhibitor. II. Study at 2.0 E Resolution. (1998). Acta Cryst. D54, 237-242.
- Miller, S.T., Hogle, J.M. & Filman, D.J. A genetic Algorithm for the Ab-initio Phasing of Icosahedral Viruses. (1996). Acta Cryst. D52, 235-251.
- Miller, R. & Weeks, C.M. Shake-and-bake: applications and advances. (1998). Direct Methods for Solving Macromolecular Structures, edited by S.Fortier, pp.389-400. Kluwer Academic Publishers.
- Mishnev, A.F. & Woolfson, M.M. On the Application of Phase Relationships to Complex Structures. Relationships to Complex Structures. XXXIV. VFOM – a New Figure of Merit for Protein Phase Sets at Moderate Resolution. (1994). Acta Cryst. D50, 842-846.
- Mo, F., Mathiessen, R.H., Hauback, B.C. & Adman, E.T. Application of Known Xray Phases in the Crystallographic Study of a Small Protein. (1996). Acta Cyst., D52, 893-900.

- Mo, F., Mathiesen, R.H., Alzari, P.M., Lescar, J. & Rasmussen, B. Low-resolution triplet phases for two proteins estimated from three-beam diffraction experiments-acquisition and possible applications. (1998). Eighteenth European Crystallographic Meeting, Praha, Czech Republic, 15/20 August 1998, Book of Abstracts, 484.
- Moras, D., Lorber, B., Romby, P., Ebel, J.-P, Giegй, R., Lewit-Bentley, A. & Roth, M. Yeast tRNAAsp –Aspartyl-tRNA Synthetase: The Crystalline Complex. (1983). J.Biomol.Structure & Dynamics 1, 209-223.
- Namba, K. & Stubbs, G. Solving the Phase Problem in Fiber Diffraction. Application to Tobacco Mosaic Virus at 3.6 E Resolution. (1985). A41, 252-262.
- Navaza, J. AmoRe: an Automated Package for Molecular Replacement. (1994). Acta Cryst. A50, 157-163.
- Navaza, K. & Vernoslova, E. On the Fast Translation Functions for Molecular Replacement. (1995). Acta Cryst., A51, 445-449.
- Perutz, M.E. (1956) Isomorphous Replacement and Phase Determination in Noncentrosymmetric Space Groups. Acta Cryst. 9, 867-873.
- Petrova, T.E., Lunin, V.Y. & Podjarny, A.D. Likelihood-based search of the centre of a macromolecular object. (1998^a). Bulletin of the Czech and Slovak Crystallographic Association, 5, special issue B, ECM-18 abstracts, 481-482.
- Petrova, T.E., Lunin, V.Y. & Podjarny, A.D. Likelihood-Based Search for the Macromolecular Position in the Crystalline Unit Cell. (1998^b). Acta Cryst. D, *submitted*.
- Podjarny, A.D. & Yonath, A. Use of Matrix Direct Methods for Low-Resolution Phase Extension for tRNA. (1977). Acta Cryst. A33, 655-661.
- Podjarny, A.D., Schevitz, R.W. & Sigler, P.B. Phasing Low-Resolution Macromolecular Structure Factors by Matrical Direct Methods. (1981). Acta Cryst. A37, 662-668.
- Podjarny, A.D., Rees, B., Thierry, J.-C., Cavarelli, J., Jesior, J.C., Roth, M., Lewitt-Bentley, A., Kahn, R., Lorber, B., Ebel, J.-P., Giegй, R. & Moras, D. Yeast

tRNAAsp –Aspartyl-tRNA Synthetase Complex: Low Resolution Crystal Structure. (1987). J.Biomol.Struct. & Dynamics 5, 187-198.

- Podjarny, A.D., Rees, B. & Urzhumtsev, A.G. Density Modification in X-Ray Crystallography. (1996). In Methods in Molecular Biology, 56, 205-226.
- Podjarny, A.D., Urzhumtsev, A.G. Low resolution phasing. (1997). In *Methods in Enzymology*, Academic Press, San Diego., C.W.Carter, Jr., R.M.Sweet, eds. 276A, 641-658
- Rayment, I., Johnson, J.E., Suck, D., Akimoto, T., Rossmann, M.G., Lonberg-Holm, K., Korant, B.D., Johnson, J.E. An 11 E –Resolution Electron Density Map of Southern Bean Mosaic Virus. (1978). Acta Cryst. B34, 567-578.
- Rayment, I., Baker, T.S., Caspar, D.L.D. & Murakami, W.T. Polyoma virus capsid structure at 22.5 E resolution. (1982). Nature 295, 110-119.
- Rayment, I. Molecular Replacement Method at Low Resolution: Optimum Strategy and Intrisic Limitation as Determined by Calculations on Icosahedral Virus Models. (1983). Acta Cryst. A39, 102-116.
- Read, R.J. Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors. (1986). Acta Cryst. A42, 140-149.
- Read, R.J. & Moult, J. Fitting Electron Density by Systematic Search. (1992). Acta Cryst. A48, 104-113.
- Rossmann, M.G. & Blow, D.M. Determination of Phases by the Conditions of Non-Crystallographic Symmetry. (1963). Acta Cryst. 16, 39-45.
- Rossmann, M.G. (1972). *The Molecular Replacement Method.*, Gordon & Breach; New York, London, Paris.
- Rossmann, M.G. The Molecular Replacement Method. (1990). Acta Cryst., A46, 73-82.
- Roth, M. Best Density Maps in Low-Resolution Crystallography with Contrast Variation. (1987). Acta Cryst. A43, 780-787.
- Shapiro, A., Botha, J.D., Pastore, A. & Lesk, A.M. A Method for Multiple Superposition of Structures. (1992). Acta Cryst. A48, 11-14.

- Sheldrick, G.M. SHELX applications to macromolecules. (1998). Direct Methods for Solving Macromolecular Structures, edited by S.Fortier, pp.401-411. Kluwer Academic Publishers.
- Simonov, V.I. Phase refinement by the method of modification and Fourier transformation of an approximate electron density distribution. (1976). In Crystallographic Computing Techniques. Editted by Ahmed, F.R., Huml, K. & Sedláček, 138-143. Copenhagen: Munksgaard.
- Stark, H., Mueller, F., Orlova, E.V., Schatz, M., Dube, P., Erdemir, T., Zemlin, F., Brimacombe, R. & van Heel, M. The 70S *Escherichia coli* ribosome at 23 Å resolution: fitting the ribosomal RNA. (1995). *Structure* 3, 815-821.
- Stenkamp, R.E. Low Resolution Reflections: Phases, Finding Solvent Molecules and Refinement Rubredoxin as a Test Case. (1994). In "ACA Annual Meeting, June 25 - July 1, 1994", p.40. INFORUM, Atlanta, Georgia, Atlanta Convention Center.
- Subbiah, S. Low-Resolution Real-Space Envelopes: An Approach to the Ab Initio Macromolecular Phase Problem. (1991). Science 252, 128-133.
- Subbiah, S. Low-Resolution Real-Space Envelopes: Improvements to the Condensing Protocol Approach and a New Method to Fix the Sign of Such Envelopes. (1993). Acta Cryst. D49 108-119.
- 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., Solution Scattering Analysis of the 70S *Escherichia coli* Ribosome by Contrast Variation. I. Invariants and Validation of Electron Microscopy Models, 1997, *J.Mol.Biol.*, **271**, 588-601.
- Svergun, D.I., Burhardt, N., Pedersen, J.S., Koch, M.H.J., Volkov, V.V., Kozin, M.B., Meerwink, W., Stuhrmann, H.B., Diedrich, G. & Nierhaus, K.H., Solution Scattering Stryctural Analysis of the 70S *Escherichia coli* Ribosome by Contrast Variation. II. A Model of the Ribosome and its RNA at 3.5 nm Resolution, 1997, *J.Mol.Biol.*, 271, 602-618.

- Tsoucaris, G. A New Method for Phase Determination. The "Maximum Determinant Rule". (1970). Acta Cryst. A26, 492-499.
- Urzhumtsev, A.G., Lunin, V.Yu. & Luzianina, T.B. Bounding a Molecule in a Noisy Synthesis. (1989). Acta Cryst., A45, 34-39.
- Urzhumtsev, A.G. Low-Resolution Phases: Influence on SIR Syntheses and Retrieval with Double-Step Filtration. (1991). Acta Cryst., A47, 794-801.
- Urzhumtsev, A.G., Podjarny, A.D. & Navaza, J. On the solution of the molecular replacement problem at very low resolution. (1994). Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 30, 29-36.
- Urzhumtsev, A.G. & Podjarny, A.D. On the Solution of the Molecular-Replacement Problem at Very Low Resolution: Application to Large Complexes. (1995). Acta Cryst, D51, 888-895.
- Urzhumtsev, A.G. & Podjarny, A.D. On the problem of solvent modeling in macromolecular crystals using diffraction data: 1. The low-resolution range. (1995). Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, 32, 12-16.
- Urzhumtsev, A.G., Skovoroda, T.P. & Lunin, V.Y. A procedure compartible with X-PLOR for the calculation of electron-density maps weighted using an R-free-likelihood-based approach. (1996). J.Appl.Cryst. 29, 741-744.
- Urzhumtsev, A.G., Vernoslova, E.A. & Podjarny, A.D. Approaches to Very Low Resolution Phasing of the Ribosome 50S Particle from Thermus thermophilus by the Few-Atoms-Models and Molecular Replacement Methods. (1996). Acta Cryst. D52, 1092-1097.
- Urzhumtsev, A.G., Lunin, V.Y. & Lunina, N.L. Seminvariant density decomposition and connectivity analysis in the low resolution macromolecular phasing. (1998). Bulletin of the Czech and Slovak Crystallographic Association, 5, special issue B, ECM-18 abstracts, 482-483.
- Vasiliev, V.D. Electron microscopy study of 70S ribosomes of Escherichial coli. (1971). FEBS Letters, 14, 203-205.
- Volkmann, N., Hottentrgger, S., Hansen, H.A., Zayzsev-Bashan, A., Sharon, R., Yonath, A. & Wittmann, H.G. Characterization and Preliminary

Crystallographic Studies on Large Ribosomal Subunits from *Thermus thermophilus*. (1990). J.Mol.Biol. 216, 239-241.

- Volkman, N., Schlünzen, F., Urzhumtsev, A.G., Vernoslova, E.A., Podjarny, A.D., Roth, M., Pebay-Peyroula, E., Berkovitch-Yellin, Z., Zaytzev-Bashan, A. & Yonath, A. On ab-initio phasing of ribosomal particles ar very low resolution. (1995). JOINT CCP4 and ESF-EACBM Newsletter on Protein Crystallography, 32, 23-32.
- Wang, B.C. Resolution of phase ambiguity (1985). Methods Enzymol., 115, 90-112.
- Ward, K.B., Hendrickson, W.A. & Klippenstein, G.L. Quaternary and tertiary structure of haemerythrin. (1975). Nature 257,818-821.
- Weckert, E. & Hьmmer,K. Multiple-Beam X-ray Diffraction for Physical Determination of Reflection Phases and its Applications. (1997). Acta Cryst. A53, 108-143.
- Weckert, E., Hulzer, K., Schroer, K.,Zellner, J. & Hьmmer,K. Phasing of small and macromolecular structures by three-beam interference. (1997). Seventeenth European Crystallographic Meeting, Lisboa, Portugal, 24/28 August 1997, Book of Abstracts, 7.
- Weckert, E., Hulzer, K., Mbller, R., Schroer, K. & Zellner, J. Using maximumentropy methods to exploit measured triplet phases for the structure determination of small proteins. (1998). Eighteenth European Crystallographic Meeting, Praha, Czech Republic, 15/20 August 1998, Book of Abstracts, 483-484.
- White, P.S. & Woolfson,, M.M. The Application of Phase Relationships to Complex Structures. VII. Magic Integers. (1975). Acta Cryst., A29, 133-137.
- Woolfson, M.M. Structure Determination by the Method of Permutation Syntheses.(1954). Acta Cryst. 7, 65-67.
- Woolfson, M.M. & Yao Jia-xing On the Application of Phase Relationships to Complex Structures. XXX. Ab initio Solution of a Small Protein by SAYTAN. (1990). Acta Cryst. A46, 409-413.

- Xiang,S., Carter, C.W.Jr, Bricogne, G. & Gilmore, C.J. Entropy Maximization Contrained by Solvent Flatness: a New Method for Macromolecular Phase Extension and Map Improvement. (1993). Acta Cryst. D49, 193-212.
- Zhang, K.Y.J. & Main, P. Histogram Mathing as a New Density Modification Technique for Phase Refinement and Extension of Protein Molecules. (1990). Acta Cryst. A46, 41-46.