

АКАДЕМИЯ НАУК СССР
НАУЧНО - ИССЛЕДОВАТЕЛЬСКИЙ ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР

На правах рукописи
УДК 548.737

Уржумцев Александр Георгиевич

ИСПОЛЬЗОВАНИЕ СМЕШАННОЙ АТОМНОЙ МОДЕЛИ
ДЛЯ УЛУЧШЕНИЯ НАБОРОВ ФАЗ СТРУКТУРНЫХ ФАКТОРОВ
ПРИ РЕНТГЕНОСТРУКТУРНОМ ИССЛЕДОВАНИИ МАКРОМОЛЕКУЛ

01.04.18 - Кристаллография и кристаллофизика

Диссертация
на соискание ученой степени
кандидата физико-математических наук

Научные руководители -
доктор биологических наук
Ю.Н.Чиргадзе
кандидат физико-математических наук
В.Ю.Лунин

Пущино - 1985 г.

О Г Л А В Л Е Н И Е

ВВЕДЕНИЕ	5
ГЛАВА I. МЕТОДЫ УЛУЧШЕНИЯ НАБОРОВ ФАЗ СТРУКТУРНЫХ ФАКТОРОВ ПРИ ИССЛЕДОВАНИИ МАКРОМОЛЕКУЛ (Литературный обзор)	II
I.1. Проблема улучшения наборов фаз в белковой кристаллографии	II
I.2. Уравнения Сэйра, их аналоги и следствия ...	15
I.3. Метод модификации функции распределения электронной плотности	23
I.4. Неотрицательность функции распределения электронной плотности. Принцип максимума определителя	28
I.5. Использование геометрической информации ...	33
I.6. Использование моделей молекулы	38
I.7. Комбинирование дополнительной информации различных типов	50
I.8. Сравнение методов улучшения наборов фаз структурных факторов для макромолекул	56
ГЛАВА 2. УЛУЧШЕНИЕ НАБОРОВ ФАЗ СТРУКТУРНЫХ ФАКТОРОВ ПУТЕМ ПОСТРОЕНИЯ И МОДИФИКАЦИИ ГРУБОЙ ЛИБО СМЕШАННОЙ АТОМНОЙ МОДЕЛИ	67
2.1. Общее описание метода	67
2.2. Построение грубой модели	74
2.3. Модификация грубой модели	85
2.4. Частичная реконструкция грубой модели	92
2.5. Построение заключительного синтеза	94
2.6. Анализ качества полученных данных	96

2.7. Использование дополнительной информации ...	I00
2.8. Использование смешанной атомной модели	I03
2.9. Программное обеспечение	I10
ГЛАВА 3. ПРОВЕРКА МЕТОДА НА МОДЕЛЬНЫХ ДАННЫХ	I13
3.1. Общая организация численных экспериментов .	I13
3.2. Численный эксперимент по расширению набора фаз	I14
3.3. Численный эксперимент по уточнению значений фаз	I26
ГЛАВА 4. УЛУЧШЕНИЕ НАБОРА ФАЗ ДЛЯ γ -КРИСТАЛЛИНА Ш	
С ИСПОЛЬЗОВАНИЕМ ГРУБОЙ АТОМНОЙ МОДЕЛИ	I30
4.1. Общая характеристика γ -кристаллина Ш ...	I30
4.2. Построение грубой модели	I31
4.3. Модификация грубой модели	I34
4.4. Анализ качества наборов фаз	I35
4.5. Построение и анализ заключительного синтеза	I40
ГЛАВА 5. УТОЧНЕНИЕ НАБОРА ФАЗ ДЛЯ γ -КРИСТАЛЛИНА Ш	
С ИСПОЛЬЗОВАНИЕМ СМЕШАННОЙ АТОМНОЙ МОДЕЛИ ...	I46
5.1. Описание частичной модели	I46
5.2. Уточнение частичной модели	I47
5.3. Использование смешанной атомной модели	I49
5.4. Результат первого этапа улучшения набора фаз с помощью смешанной атомной модели	I55
5.5. Использование смешанной атомной модели для улучшения набора фаз (второй этап)	I56
5.6. Эффективность методики для улучшения фаз структурных факторов γ -кристаллина Ш ...	I62

ГЛАВА 6. УЛУЧШЕНИЕ НАБОРА ФАЗ СТРУКТУРНЫХ ФАКТОРОВ

ДЛЯ АСПАРТАТ-АМИНОТРАНСФЕРАЗЫ	I67
6.1. Общие сведения об аспартат-аминотрансферазе	I67
6.2. Построение грубой модели	I68
6.3. Модификация грубой модели	I70
6.4. Использование разностных синтезов	I71
6.5. Построение и анализ заключительного синтеза	I75
ВЫВОДЫ	I77
ЛИТЕРАТУРА	

В В Е Д Е Н И Е

Рентгеноструктурный анализ является одним из важнейших методов исследования пространственной структуры макромолекул. С повышением в последнее время интереса к структуре и функциям белков, нуклеиновых кислот и вирусов повышается и роль рентгеноструктурного анализа в этих исследованиях.

Важнейшим результатом применения метода рентгеноструктурного анализа является трехмерная атомная модель молекулы, позволяющая проводить дальнейшие исследования структуры.

Стандартный путь от рентгеновского эксперимента до получения окончательной атомной модели при исследовании макромолекул состоит из ряда этапов (см., например, [I]):

- 1) кристаллизация нативного белка;
- 2) получение тяжелоатомных производных;
- 3) проведение рентгеновского эксперимента, сбор данных;
- 4) первичная обработка данных, определение модулей структурных факторов;
- 5) определение положений тяжелых атомов;
- 6) расчет фаз структурных факторов, построение карт распределения электронной плотности;
- 7) интерпретация карт распределения электронной плотности, построение атомной модели;
- 8) уточнение атомной модели.

Среди этих этапов в настоящее время наименее алгоритмизированным является этап построения атомной модели. Основная причина сложностей заключается в недостаточно высоком качестве карт для больших молекул. Это обусловлено ошибками в используемых для построения карт значениях модулей и фаз

структурных факторов и существенной неполнотой набора данных.

Модули структурных факторов определяются непосредственно по экспериментальным данным. Основными методами получения информации о фазах при исследовании макромолекул являются методы изоморфного замещения [2], аномального рассеяния [3], молекулярного замещения [4]. Эти методы опираются на дополнительные гипотезы о наличии изоморфизма, идентичности пространственной структуры двух молекул и т.п., справедливые лишь приближенно. Поэтому найденные с помощью таких методов значения фаз могут существенно отличаться от истинных величин, а фазы отражений высокого разрешения практически всегда остаются неопределенными. Влияние ошибок в значениях фаз на качество карт распределения электронной плотности существенно сильнее влияния погрешностей экспериментального определения значений модулей структурных факторов. Поэтому в данной работе внимание будет сконцентрировано на уменьшении ошибок в определении фаз, значения же модулей будут считаться известными практически точно.

Таким образом, один из возможных подходов к повышению качества карт распределения электронной плотности может быть сформулирован так: считая, что модули заданного набора структурных факторов известны точно и что для некоторых структурных факторов из этого набора известны приближенные значения фаз, попытаться наиболее точно определить значения фаз полного набора структурных факторов, привлекая для этого дополнительную информацию.

Возможны и другие подходы к решению задачи улучшения качества карт. Например, можно ставить задачу поиска функ-

ции $Q(\vec{\zeta})$, "не противоречащей" имеющейся информации о модулях и фазах структурных факторов и отвечающей некоторым дополнительным требованиям (см., например, [5]). Такой подход отличается от первого, в частности, тем, что независимыми переменными здесь являются значения функции $Q(\vec{\zeta})$ во всех точках асимметричной части элементарной ячейки (в практических применениях – значения $Q(\vec{\zeta})$ во всех узлах некоторой сетки), то есть, фактически, разрешается использование высокочастотных гармоник, для которых эксперимент не дает никаких данных. Этот подход рассматриваться в работе не будет.

Выбор для исследования все более сложных объектов повышает требования к качеству карт распределения электронной плотности потому, что соответственно усложняется и построение атомных моделей. Большинство исследований пространственной структуры макромолекул, выполненных с помощью метода рентгеноструктурного анализа в последние годы, содержат в качестве одного из этапов работы улучшение исходных карт. Практически все методы, используемые для этого в настоящее время, действуют в рамках первого подхода – улучшения набора фаз. Обзору этих методов посвящена глава I данной работы. Проведенный анализ методов свидетельствует об ограниченной области их применения и необходимости развития новых методик, более широко использующих дополнительную информацию различных типов.

Основная часть работы посвящена описанию практической реализации методики улучшения набора фаз для макромолекул путем построения и модификации грубой либо смешанной атомной модели. В методическом плане центральной здесь является методика использования грубой модели, описание которой посвя-

щена практически вся глава 2. Отдельный параграф посвящен методике использования смешанной атомной модели.

Метод улучшения набора фаз путем построения и модификации грубой атомной модели опирается на идею, предложенную в 1977 году Агарвалом и Айзексом [6]. Эта идея заключается в улучшении набора фаз путем организации следующей процедуры:

- 1) по имеющемуся набору модулей и фаз структурных факторов рассчитывается функция распределения электронной плотности $\rho(\vec{r})$;
- 2) по функции $\rho(\vec{r})$ автоматически строится атомная модель, основное требование к которой – достаточно хорошее качество воспроизведения функции распределения электронной плотности $\rho(\vec{r})$; по атомного соответствия модели исследуемой структуре при этом не требуется;
- 3) проводится модификация построенной модели путем минимизации критерия типа R -фактора;
- 4) по модифицированной модели рассчитывается набор структурных факторов, фазы которых вместе с экспериментально определенными модулями могут быть использованы для построения новой, улучшенной функции распределения электронной плотности; при недостаточно высоком качестве полученной функции процедура может быть повторена.

Заметим, что аналогичная идея применительно к этапу работы при более низком разрешении была независимо высказана и развита Боровиковым, Вайнштейном, Гельфандом и Калининым [7,8].

Используя такую идею в качестве отправной, мы разработали новую методику, которая существенно отличается от рекомендаций Агарвала и Айзекса как по вопросам построения модели, так и по вопросам модификации модели и последующего ис-

пользования фазовой информации. Излагаемый ниже подход позволил получить в тестовой ситуации вдвое меньшую ошибку предсказания фаз по сравнению с ошибкой, полученной Агарвалом и Айзексом в экспериментах, близких к нашим.

Дальнейшим развитием идеи является излагаемая ниже методика использования смешанной атомной модели. Эта методика позволяет улучшить набор фаз в ситуации, когда построенная исследователем атомная модель соответствует лишь части структуры и недостаточно полна, чтобы улучшать набор фаз и модель за счет ее уточнения. Мы предложили улучшать набор фаз в такой ситуации путем построения и модификации смешанной атомной модели. Здесь смешанная модель состоит из атомов частичной модели, дополненных фиктивными атомами грубой модели в тех областях элементарной ячейки, где интерпретация синтеза не была проведена. Расширение проинтерпретированной части синтеза по мере улучшения карт позволяет плавно переходить к обычному кристаллографическому уточнению атомной модели.

При разработке методики были проведены численные эксперименты на модельных данных по структуре белка актинидина, координаты атомов которого были взяты из Банка белковых данных [9]. При расширении набора фаз с 3.0 \AA (464I отражение) до 2.0 \AA удалось определить значения фаз для 10368 отражений разрешением $3.0 \text{ \AA} > d \geq 2.0 \text{ \AA}$ со средней погрешностью 35° . В другом эксперименте при уточнении зашумленного набора фаз (средняя ошибка определения фаз 37°) в зоне $d \geq 3.0 \text{ \AA}$ удалось снизить эту ошибку до 21° . Подробное описание численных экспериментов дано в главе 3.

В главе 4 описан пример использования разработанной ме-

тодики для расширения набора фаз структурных факторов белка γ -кристаллина Шь из глазной линзы теленка. Структура белка исследуется в Институте белка АН СССР под руководством д.б.н. Ю.Н.Чиргадзе. Исходный набор фаз структурных факторов разрешением до 3.0 \AA [10], полученный методами изоморфного замещения и аномального рассеяния, был уточнен и расширен до разрешения 2.7 \AA путем построения и модификации грубой атомной модели. Полученные данные были использованы для расчета синтеза электронной плотности, по которому, в отличие от исходного, оказалось возможным построение части атомной модели (определено около 69% общего числа неводородных атомов).

Дальнейшее улучшение набора фаз структурных факторов для γ -кристаллина Шь описано в главе 5. Результатом этой работы, выполненной с использованием смешанной атомной модели, явился синтез электронной плотности, по которому имевшаяся частичная модель была скорректирована и дополнена. Построенная таким образом модель со стандартной стереохимией содержала приблизительно 90% общего числа неводородных атомов (кристаллографический R-фактор для модели составлял 0.46), что позволило перейти к обычному кристаллографическому уточнению этой модели.

В главе 6 описано использование метода для расширения набора фаз структурных факторов аспартат-аминотрансферазы, структура которой исследуется в Институте кристаллографии АН СССР группой к.ф.-м.н. В.В.Борисова [11]. В результате выполненной работы удалось реально повысить разрешение карт распределения электронной плотности с 3.5 \AA до 2.8 \AA и тем самым облегчить анализ этих карт.

Работы по улучшению наборов фаз для γ -кристаллина Шь и

аспартат-аминотрансферазы подтвердили возможности метода, продемонстрированные в численных экспериментах.

ГЛАВА I

МЕТОДЫ УЛУЧШЕНИЯ НАБОРОВ ФАЗ СТРУКТУРНЫХ ФАКТОРОВ ПРИ ИССЛЕДОВАНИИ МАКРОМОЛЕКУЛ (Литературный обзор)

I.I. Проблема улучшения наборов фаз в белковой кристаллографии.

Результатом применения метода рентгеноструктурного анализа к исследованию того или иного вещества является пространственная атомная модель, описывающая структуру исследуемой молекулы в терминах координат \vec{z} ; центров атомов и параметров рассеяния этих атомов. Обычным путем построения такой модели является анализ функции $\rho(\vec{z})$ распределения электронной плотности в кристалле исследуемого вещества, которая может быть построена на основе экспериментальных данных:

$$\rho(\vec{z}) = \frac{1}{V} \sum_{\vec{s} \in R'} F(\vec{s}) e^{i\psi(\vec{s})} e^{-2\pi i (\vec{s}, \vec{z})} \quad (I)$$

Здесь V – объем элементарной ячейки,

R' – целочисленная решетка в обратном пространстве,

$F(\vec{s})$ – модуль структурного фактора $F(\vec{s}) = F(\vec{s}) e^{i\psi(\vec{s})}$,

$\psi(\vec{s})$ – фаза структурного фактора.

Рентгеновский эксперимент позволяет определить для некоторого набора отражений (обозначим этот набор S_1) модули структурных факторов $\{F(\vec{s})\}_{\vec{s} \in S_1}$, которые мы будем полагать известными точно. Дополнительные эксперименты с тяжелоатом-

ными производными (либо какие-нибудь альтернативные методы) позволяют получить информацию о фазах $\Psi(\xi)$ для набора отражений S_0 , который, вообще говоря, является частью набора S_1 . Типичной при исследовании белков является ситуация, когда метод изоморфного замещения дает возможность определить фазы структурных факторов разрешением не выше $3.5 - 3.0 \text{ \AA}$, в то время, как дифракционное поле кристаллов нативного белка простирается до $2.5 - 2.0 \text{ \AA}$.

Проблема, которой посвящена данная работа, заключается в том, чтобы, используя дополнительную информацию об исследуемом объекте, по значениям модулей $\{F(\xi)\}_{\xi \in S_1}$ и известной информации о фазах отражений из множества S_0

- I) уточнить информацию о фазах для отражений из S_0 - задача уточнения фаз,
- 2) определить неизвестные фазы для отражений из S_1 , не входящих во множество S_0 - задача расширения набора фаз.

Одновременное решение обеих задач I) и 2) далее будет называться улучшением набора фаз. Задача расширения набора фаз будет, в основном, рассматриваться в ситуации, типичной для исследования макромолекул - в ситуации повышения разрешения, т.е. будет подразумеваться, что S_1 - набор отражений более высокого разрешения, чем S_0 . Для исследования низкомолекулярных соединений характерна другая ситуация, когда исходный набор содержит фазы "сильных" отражений, а расширение набора происходит путем определения фаз более "слабых" отражений внутри той же зоны разрешения. Последняя задача в данной работе отдельно рассматриваться не будет, хотя четкую границу между этими двумя ситуациями провести затруднительно. Например, при использовании метода изоморфного замещения

фазы для некоторых "слабых" отражений не удается определить, и их определение также включается в процесс расширения набора фаз.

Информация о фазах может быть представлена в двух различных формах. В первом случае фаза $\varphi(\vec{s})$ задается ее некоторым (приближенным) значением $\tilde{\varphi}(\vec{s})$, и можно говорить о погрешности задания фазы $\delta_{\vec{s}} = |\tilde{\varphi}(\vec{s}) - \varphi(\vec{s})|$ либо о средней по набору S ошибке задания фаз

$$\bar{\delta} = \sum_{\vec{s} \in S} |\tilde{\varphi}(\vec{s}) - \varphi(\vec{s})| / M(S) \quad (2)$$

(Здесь и далее через $M(S)$ будем обозначать количество элементов множества S). Уточнение значений фаз в этом случае означает уменьшение погрешности.

Во втором случае информация о фазе $\varphi(\vec{s})$ задается в виде распределения вероятностей $P_{\vec{s}}(\varphi)$ для ее значений. Конкретное значение фазы можно определять, например, как наиболее вероятное значение

$$\tilde{\varphi}(\vec{s}) = \varphi_{MP}(\vec{s}) = \arg \max P_{\vec{s}}(\varphi) \quad (3)$$

либо как среднее значение

$$\tilde{\varphi}(\vec{s}) = \varphi_{AV}(\vec{s}) = \langle \varphi \rangle_{P_{\vec{s}}} = \int_0^{2\pi} \varphi P_{\vec{s}}(\varphi) d\varphi \quad (4)$$

Здесь надежность оценки $\tilde{\varphi}(\vec{s})$ определяется среднеквадратичным отклонением

$$\sigma^2 = \langle (\varphi - \tilde{\varphi}(\vec{s}))^2 \rangle_{P_{\vec{s}}} = \int_{\tilde{\varphi}(\vec{s}) - \pi}^{\tilde{\varphi}(\vec{s}) + \pi} (\varphi - \tilde{\varphi}(\vec{s}))^2 P_{\vec{s}}(\varphi) d\varphi \quad (5)$$

Уточнение значений фазы в этом случае означает представление информации в виде распределения $P_{\vec{s}}(\varphi)$ с меньшим значением σ^2 .

Основой для улучшения набора фаз является информация об

объекте, дополнительная к использованной ранее при определении исходных значений фаз. На практике используется сравнительно небольшое число типов такой информации:

- 1) атомарное строение молекулы;
- 2) дискретность (в грубом приближении) значений функции распределения электронной плотности;
- 3) неотрицательность функции распределения электронной плотности;
- 4) локальная (некристаллографическая) симметрия;
- 5) данные об области, занимаемой молекулой;
- 6) наличие модели объекта либо его части.

В практической работе обычно привлекается одновременно несколько типов такой информации.

Отметим, что, как правило, дополнительная информация является "описательной", и для ее использования требуется математическая переформулировка в виде тех либо иных ограничений. Можно выделить два основных широких класса способов использования дополнительной информации:

- 1) выписывание на основе дополнительной информации систем уравнений или неравенств, связывающих структурные факторы, и решение (обычно, приближенное) их тем либо иным способом;
- 2) использование фазовой информации, полученной при моделировании рассеяния рентгеновских лучей на различных моделях исследуемого объекта.

Ниже приведен обзор существующих методов улучшения наборов фаз для макромолекул и примеры практического использования этих методов.

I.2. Уравнения Сэйра, их аналоги и следствия.

I.2.1. Вывод уравнений Сэйра, их аналогов и следствий.

Для получения уравнений, связывающих структурные факторы, Сэйром [12] в 1952 г. был предложен следующий подход, основывающийся на предположении, что функция распределения электронной плотности молекулы формируется суммой вкладов одинаковых атомов, изображения которых отделены друг от друга. Последнее предположение может быть переформулировано как Гипотеза I. Функция $\rho(\vec{r})$ распределения электронной плотности внутри элементарной ячейки кристалла удовлетворяет условию

$$\rho(\vec{r}) = \sum_{j=1}^N \rho_0(\vec{r} - \vec{r}_j), \quad (6)$$

причем

$$\rho_0(\vec{r} - \vec{r}_j) \cdot \rho_0(\vec{r} - \vec{r}_k) \equiv 0 \quad \text{при } j \neq k. \quad (7)$$

Здесь \vec{r}_j — координаты центров атомов, $1 \leq j \leq N$,

$\rho_0(\vec{r}) = \rho_0(|\vec{r}|)$ — сферически симметричная функция, описывающая распределение электронной плотности для отдельного атома.

Для функций типа (6), образованных суммой вкладов отдельных атомов, структурные факторы представимы в виде

$$F(s)e^{is(\vec{s})} = \sum_{j=1}^N f_0(|\vec{s}|) e^{2\pi i (\vec{s}, \vec{r}_j)} = f_0(|\vec{s}|) \sum_{j=1}^N e^{2\pi i (\vec{s}, \vec{r}_j)}, \quad (8)$$

где

$$f_0(s) = \frac{2}{s} \int_0^\infty r \rho_0(r) \sin 2\pi s r dr \quad (9)$$

Возведя функцию $\rho(\vec{r})$ в квадрат и обозначив $\hat{\rho}(\vec{r}) = \rho^2(\vec{r})$, в силу гипотезы I имеем

$$\hat{\rho}(\vec{r}) = \sum_{j=1}^N \rho_0^2(\vec{r} - \vec{r}_j) \quad (10)$$

Структурные факторы, отвечающие $\hat{F}(\vec{s})$,

$$\hat{F}(\vec{s})e^{i\varphi(\vec{s})} = f_2(|\vec{s}|) \sum_{j=1}^N e^{2\pi i(\vec{s}, \vec{\zeta}_j)} = \frac{f_2(|\vec{s}|)}{f_0(|\vec{s}|)} f_0(|\vec{s}|) \sum_{j=1}^N e^{2\pi i(\vec{s}, \vec{\zeta}_j)}, \quad (II)$$

где

$$f_2(s) = \frac{2}{s} \int_0^\infty z \zeta_0^2(z) \sin 2\pi s z dz, \quad (I2)$$

отличаются от структурных факторов $F(\vec{s})e^{i\varphi(\vec{s})}$ лишь множителем $\alpha(\vec{s}) = f_2(|\vec{s}|)/f_0(|\vec{s}|)$, не зависящим от координат атомов,

$$\hat{F}(\vec{s}) = \alpha(\vec{s}) F(\vec{s}) \quad (I3)$$

Используя свойство преобразования Фурье переводить произведение двух функций в свертку их преобразований Фурье, последнее равенство можно переписать в виде

$$\alpha(\vec{s}) \vec{F}(\vec{s}) = \sum_{\vec{u} \in \mathbb{R}^3} \vec{F}(\vec{u}) \vec{F}(\vec{s} - \vec{u}) \quad \text{при всех } \vec{s} \in \mathbb{R}^3 \quad (I4)$$

Систему уравнений (I4) принято называть системой уравнений Сэйра.

В ряде случаев удобно перейти к нормализованным структурным факторам

$$E(\vec{s})e^{i\varphi(\vec{s})} = N^{-1/2} \sum_{j=1}^N e^{2\pi i(\vec{s}, \vec{\zeta}_j)} = \frac{1}{N^{1/2} f_0(|\vec{s}|)} F(\vec{s})e^{i\varphi(\vec{s})}, \quad (I5)$$

отвечающим "обостренной" функции распределения электронной плотности (сумме "точечных" атомов)

$$e(\vec{\zeta}) = \sum_{j=1}^N \delta(\vec{\zeta} - \vec{\zeta}_j), \quad (I6)$$

где $\delta(\vec{\zeta})$ — трехмерная δ -функция Дирака. Для нормализованных структурных факторов также могут быть получены уравнения Сэйра, аналогичные (I4).

Считая, что модули $F_0(\vec{s})$ структурных факторов известны

из эксперимента, можно систему (I4) рассматривать как систему уравнений для определения фаз. Каждое из комплексных уравнений (I4) распадается на два действительных – для модульной и для фазовой части:

$$\alpha(\vec{s}) F_0(\vec{s}) = \left| \sum_{\vec{u}} F_0(\vec{u}) F_0(\vec{s}-\vec{u}) e^{i[\varphi(\vec{u}) + \varphi(\vec{s}-\vec{u})]} \right| \quad (I7A)$$

$$\varphi(\vec{s}) = \arg \sum_{\vec{u}} F_0(\vec{u}) F_0(\vec{s}-\vec{u}) e^{i[\varphi(\vec{u}) + \varphi(\vec{s}-\vec{u})]} \quad (I7B)$$

Отбросив уравнения для модулей (I7A), мы получим следствие – фазовую часть (I7B) уравнений, которая переписывается в виде

$$\operatorname{tg} \varphi(\vec{s}) = \frac{\sum_{\vec{u}} F_0(\vec{u}) F_0(\vec{s}-\vec{u}) \sin [\varphi(\vec{u}) + \varphi(\vec{s}-\vec{u})]}{\sum_{\vec{u}} F_0(\vec{u}) F_0(\vec{s}-\vec{u}) \cos [\varphi(\vec{u}) + \varphi(\vec{s}-\vec{u})]}, \quad (I8)$$

известном в литературе под названием тангенс-формулы. Аналогичные соотношения могут быть получены из рассмотрения нормализованных структурных факторов

$$\operatorname{tg} \varphi(\vec{s}) = \frac{\sum_{\vec{u}} E(\vec{u}) E(\vec{s}-\vec{u}) \sin [\varphi(\vec{u}) + \varphi(\vec{s}-\vec{u})]}{\sum_{\vec{u}} E(\vec{u}) E(\vec{s}-\vec{u}) \cos [\varphi(\vec{u}) + \varphi(\vec{s}-\vec{u})]} \quad (I9)$$

Отметим, что возможен другой, вероятностный подход к получению тангенс-формулы и ее модификаций, развитый в работах [I3–I7] и ряде других.

С вычислительной точки зрения улучшение набора фаз при помощи тангенс-формулы представляет собой решение системы (I7B) методом простой итерации (см., например, [I8]). Это означает, что строится следующий итерационный процесс:

- I) по имеющемуся набору значений фаз $\varphi_i(\vec{s})$ и экспериментально определенным значениям модулей $F_0(\vec{s})$ рассчитываются модифицированные значения фаз

$$\hat{\varphi}(\vec{s}) = \alpha \tau g \sum_{\vec{u}} F_0(\vec{u}) F_0(\vec{s}-\vec{u}) e^{i[\varphi(\vec{u}) + \varphi(\vec{s}-\vec{u})]} ; \quad (20)$$

- 2) в качестве следующего приближения значений фаз берется набор величин $\varphi^{j+1}(\vec{s}) = \hat{\varphi}(\vec{s})$;
- 3) итерации прекращаются, либо осуществляется возврат к пункту I).

Надежность выполнения уравнений (I7А, I7Б) определяется степенью достоверности гипотезы I в конкретной ситуации.

Для биологических макромолекул условие одинаковости атомов можно считать приближенно выполненным, поскольку эти молекулы состоят в основном из близких по рассеивающим способностям атомов углерода, азота, кислорода. Кроме того, существуют модификации [19-23] уравнений (I7), позволяющие учесть наличие разных типов атомов в структуре. Условие же разделенности атомов начинает удовлетворительно выполнятьсь лишь при работе с набором данных достаточно высокого разрешения, и это ограничивает возможности применения уравнений Сэйра для исследования больших молекул. Различные авторы пытались увеличить точность выполнения уравнений (I7) путем введения модульных [24] и фазовых [25, 26] поправок. Известны также работы, где авторы пытались уточнить тангенс-формулу с помощью дополнительных слагаемых, зависящих от трех либо четырех отражений [27].

Однако, даже при точном выполнении соотношений (I7), описанный выше метод улучшения набора фаз может не приводить к успеху. Во-первых, метод простой итерации может не сходиться. Во-вторых, система уравнений (I7Б) есть следствие полной системы уравнений Сэйра (I7), и возможно существование решений системы (I7Б), не являющихся решениями пол-

ной системы (I7). Введение в тангенс-формулу различных весовых сомножителей [28-30] не гарантирует сходимость к правильному набору значений.

1.2.2. Использование тангенс-формулы для улучшения наборов фаз структурных факторов макромолекул.

Улучшение наборов фаз структурных факторов с помощью тангенс-формулы, успешно применявшееся для исследования низкомолекулярных соединений (см., например, [31]), исходно было проблематичным для исследования больших молекул в силу приведенных выше причин, а также в силу вычислительных сложностей. Впервые возможность использования тангенс-формулы (I8) для улучшения наборов фаз структурных факторов белков была высказана Коултером [32] в 1965 г. В связи с разработкой Кули и Тьюки [33] алгоритма быстрого преобразования Фурье пригодной для практических вычислений стала итерационная процедура улучшения наборов фаз, вытекающая непосредственно из соотношения (I3):

- I) расчет $Q^j(\vec{s}) = \frac{1}{V} \sum_{\vec{s}} F_0(\vec{s}) e^{i\Phi^j(\vec{s})} e^{-2\pi i(\vec{s}, \vec{\tau})}$;
- 2) расчет $\hat{Q}(\vec{s}) = [Q^j(\vec{s})]^2$;
- 3) расчет структурных факторов $\hat{F}(\vec{s}) e^{i\hat{\Phi}(\vec{s})}$ функции $\hat{Q}(\vec{s})$;
- 4) определение очередного приближения к значениям фаз $\Phi^{j+1}(\vec{s}) = \hat{\Phi}(\vec{s})$.

В 1969 г. Вайнцирль, Айзенберг и Дикерсон [34] провели численные эксперименты на данных по модельному белку и по цитохрому С. Авторами была показана принципиальная возможность улучшения фаз структурных факторов белка при среднем (4.0-3.7 Å) разрешении. Ошибка значений фаз исходного набора заметно снижалась, хотя наблюдалась сходимость не к

точному, а к некоторому самосогласованному решению. Аналогичный факт был обнаружен также Рике и Липскомбом [35], которые исследовали возможность расширения наборов фаз белка при различном разрешении от 6 Å до 2 Å на примере данных по карбоксипептидазе А. Подробный анализ полученных функций распределения электронной плотности привел Рике и Липскомба к более сдержанной оценке возможностей использования тангенс-формулы в белковой кристаллографии. Тестирование метода проводилось, кроме того, на данных по миоглобинам [36-39].

Среди практических работ [40-45] с использованием тангенс-формулы для улучшения наборов фаз белков, как правило, небольших и при сравнительно высоком разрешении, в качестве наиболее успешной следует, по-видимому, отметить расширение Хендриксоном и Карлом [43] набора фаз структурных факторов кальций-связывающего белка с 2.0 до 1.85 Å. Интересна также работа Хендрикса [44], где продемонстрировано появление на синтезе ложных пиков.

I.2.3. Решение полной системы уравнений Сэйра.

Как уже отмечалось выше, система уравнений (I7Б) и тангенс-формула (I8) являются следствиями полной системы уравнений Сэйра (I7). Первая попытка использования информации, содержащейся в модульной части (I7А) уравнений Сэйра, была предпринята Хоппе [46], который предложил метод улучшения набора фаз путем минимизации критерия

$$R = \sum_{\zeta} (F_c(\zeta) - F_o(\zeta))^2, \quad (21)$$

где $F_c(\zeta)$ - модули структурных факторов, рассчитываемые с помощью уравнений Сэйра. Однако из-за вычислительных сложностей данный метод не был реализован.

Сэйром [47] в 1972 г. был предложен способ решения полной системы уравнений (I7) путем минимизации по переменным Ψ функционала

$$S(\Psi) = \frac{1}{2} \sum_{\vec{s}} \left| C_{\vec{s}} \vec{F}(\vec{s}) + \sum_{\vec{u}} \vec{F}(\vec{u}) \vec{F}(\vec{s}-\vec{u}) \right|^2 \quad (22)$$

где $C_{\vec{s}}$ – известные константы. Примерно в то же время Краббендалом и Крооном [23, 48] был предложен аналогичный по идеи метод, но из-за неудачно выбранного способа минимизации он был пригоден для исследования только небольших структур.

Метод Сэйра решения полной системы уравнений (I7) был проверен на тестовой структуре из 100 атомов [47] и на данных по рубредоксину [24], где была проведена серия численных экспериментов по расширению различных наборов фаз до разрешения 1.5 Å. При использовании в качестве исходного 2.0- и 2.5-ангстремных наборов фаз были получены удовлетворительные результаты, в отличие от расширения 3.0-ангстремного набора. Замечен факт наличия самосогласованного решения с отклонением значений фаз от точных величин в среднем на 32°. Характерно, что работа даже со столь небольшим белком потребовала огромных машинных ресурсов. Проверка работоспособности метода была проведена также Гассманом [49] на данных по миоглобину.

В последующих работах Сэйр и Тоупин [50, 51] радикально улучшили алгоритм минимизации, используя специфику функционала (22). В сущности, ими для данного частного случая был реализован алгоритм быстрого дифференцирования, рассмотренный позднее в общем виде в работе Кима и сотр. [52], а применительно к задачам рентгеноструктурного анализа – в рабо-

такх Лунина и Уржумцева [53-55].

Усовершенствованный метод Сэйра и Тоупина был использован Катфилдом и сотр. [56] для расширения набора фаз структурных факторов инсулина с 1.9 до 1.5 Å. Средняя погрешность определения фаз структурных факторов разрешением до 1.9 Å уменьшилась с 60.0° до 52.8°, а фазы, ранее неизвестные, были определены со средней ошибкой 60.5° [57].

I.2.4. Обобщение уравнений Сэйра.

Заметим, что свойство "атомности" (6), (7) функции распределения электронной плотности $\rho(\vec{z})$ сохраняется не только при преобразовании

$$\rho \rightarrow \rho^2, \quad (23)$$

но и при произвольном преобразовании

$$\rho \rightarrow \lambda(\rho) \quad (24)$$

где $\lambda(t)$ – произвольная функция. (Более того, как отметил Ротбауэр [22], свойство сохраняется и при преобразовании

$$\rho \rightarrow D[\rho] \quad (25)$$

где D – некоторый оператор дифференцирования.) Отсюда, аналогично уравнению (I3), из гипотезы I следует для произвольной функции $\lambda(t)$ уравнение

$$\hat{F}(\vec{s}) = \frac{f_o(|\vec{s}|)}{f_\lambda(|\vec{s}|)} \mathcal{F}_{\vec{s}} \left\{ \lambda \left[\frac{1}{\sqrt{V}} \sum_{\vec{u}} \hat{F}(\vec{u}) e^{-2\pi i (\vec{u}, \vec{s})} \right] \right\}, \quad (26)$$

где $\mathcal{F}_{\vec{s}}(a(\vec{s}))$ означает преобразование Фурье функции $a(\vec{s})$, вычисленное в точке \vec{s} обратного пространства, а

$$f_\lambda(s) = \frac{2}{s} \int_0^\infty z \lambda(\rho_o(z)) \sin 2\pi s z \, dz \quad (27)$$

Фазовая часть системы имеет вид

$$\varphi(\vec{s}) = \alpha \tau g \tilde{F}_{\vec{s}} \left\{ \lambda \left[\frac{1}{V} \sum_{\vec{u}} \tilde{F}(\vec{u}) e^{-2\pi i (\vec{u}, \vec{s})} \right] \right\} \quad (28)$$

Решение ее методом простой итерации с разнообразными функциями $\lambda(t)$ широко применялось на практике под названием метода модификации функции распределения электронной плотности. Более подробно этот метод рассмотрен в п. I.3.

I.3. Метод модификации функции распределения электронной плотности.

I.3.1. Функции с дискретным набором значений.

Как было показано выше в п. I.2., справедливость гипотезы I о разделенности изображений атомов структуры приводит к соотношению (26) между структурными факторами. Методом улучшения набора фаз путем модификации функции распределения электронной плотности (методом "модификации электронной плотности") с функцией $\lambda(t)$ называют итерационную процедуру решения фазовой части (28) системы уравнений (26):

- 1) рассчитывается $\hat{\rho}^j(\vec{r}) = \frac{1}{V} \sum_{\vec{s}} F_0(\vec{s}) e^{i\varphi^j(\vec{s})} e^{-2\pi i (\vec{s}, \vec{r})}$;
- 2) вычисляется модифицированная функция $\hat{\rho}(\vec{r}) = \lambda(\hat{\rho}^j(\vec{r}))$;
- 3) рассчитываются структурные факторы $\hat{F}(\vec{s}) e^{i\hat{\varphi}(\vec{s})}$ функции $\hat{\rho}(\vec{r})$;
- 4) определяется следующее приближение значений фаз $\varphi^{j+1}(\vec{s}) = \hat{\varphi}(\vec{s})$.

Покажем существование другого подхода [55] к выводу уравнений (28).

На некоторых этапах исследования структур можно приблизенно считать, что функция $\hat{\rho}(\vec{r})$ принимает только два значения (при надлежащей нормировке $\hat{\rho}(\vec{r})$ – значения 0

или I). Данное свойство

$$\wp(\vec{\tau}) = \begin{cases} 0 \\ 1 \end{cases} \quad (29)$$

эквивалентно выполнению условия

$$\wp(\vec{\tau}) = \wp^2(\vec{\tau}), \quad (30)$$

то есть справедливости следующей системы уравнений для структурных факторов

$$\vec{F}(\vec{s}) = \mathfrak{F}_{\vec{s}}(\wp^2) = \frac{1}{\sqrt{\sum_{\vec{\tau}}}} \sum_{\vec{\tau}} \wp^2(\vec{\tau}) e^{2\pi i (\vec{s}, \vec{\tau})} \quad (31)$$

Фазовая часть этой системы комплексных уравнений

$$\varphi(\vec{s}) = \arg \sum_{\vec{u}} F(\vec{u}) F(\vec{s} - \vec{u}) e^{i[\varphi(\vec{u}) + \varphi(\vec{s} - \vec{u})]} \quad (32)$$

полностью совпадает с фазовой частью (I7Б) системы уравнений Сэйра.

Аналогично, свойство функции $\wp(\vec{\tau})$ иметь своими значениями конечный набор величин $\{t_1, \dots, t_k\}$ эквивалентно выполнению условия

$$\wp(\vec{\tau}) = \lambda(\wp(\vec{\tau})), \quad (33)$$

где $\lambda(t)$ такова, что t_1, \dots, t_k – все корни уравнения

$$t = \lambda(t) \quad (34)$$

Соответствующие уравнения для структурных факторов имеют вид

$$\vec{F}(\vec{s}) = \mathfrak{F}_{\vec{s}}(\lambda(\wp(\vec{\tau}))), \quad (35)$$

и их фазовая часть полностью совпадает с фазовой частью (28) системы уравнений (26).

Таким образом, метод модификации электронной плотности

может трактоваться и как следствие из гипотезы о дискретности набора значений функции $Q(\vec{\zeta})$, а применимость метода не обязана ограничиваться рамками достоверности гипотезы о разделенности изображений атомов.

С другой стороны, из показанного выше следует, что системе (28) удовлетворяют функции из различных классов, и полученное решение может не удовлетворять предписанному заранее свойству.

I.3.2. Примеры применения метода.

Среди первых работ с пересчетом фаз по модифицированной функции распределения электронной плотности необходимо отметить работы Кураши [58] и Хоппе [59]. В частности, эти работы интересны тем, что в них предложены две идеи выбора функций модификации, используемые до настоящего времени:

- 1) идея неотрицательности функции распределения электронной плотности [58] и
- 2) идея схожести функции распределения электронной плотности при фиксированном разрешении и ее квадрата (либо других степеней) при более низком разрешении [59].

Впервые расширение набора фаз структурных факторов белка данным методом было выполнено в 1964 г. Хоппе и Гассманом [60]. Для структурных факторов миоглобина фазы I.5-ангстремного набора были пересчитаны по квадрату функции распределения электронной плотности, рассчитанной по данным разрешением до 2.0 \AA . Итерационная процедура здесь еще не была реализована.

Среди вычислительных вопросов, возникающих в связи с реализацией метода, отдельное внимание было уделено выбору шага сетки для вычисления функции $Q(\vec{\zeta})$ и модифицированной

функции $\lambda(Q(\xi))$ [61-63]. В большом количестве работ [64-72] обсуждался вопрос выбора функции модификации $\lambda(t)$ с целью включения наибольшей дополнительной информации и улучшения сходимости метода. Примеры таких функций приведены на рис. I.

Метод улучшения набора фаз путем модификации функции распределения электронной плотности был успешно опробован на ряде тестов и применялся в исследованиях структуры небольших молекул [70, 73-78]. Однако проверка метода на данных по белковым структурам [66, 71, 75, 79, 80] не давала однозначного ответа на вопрос об эффективности данного подхода при исследовании макромолекул. Лишь в 1983 г. Канилло и сотр. [81] провели серию успешных численных экспериментов на данных по миоглобину (расширение набора фаз с 4.0 до 1.8 Å). В качестве функции модификации $\lambda(t)$ ими была использована простейшая функция

$$\lambda(t) = \begin{cases} 1, & t > \bar{t} \\ 0, & t \leq \bar{t} \end{cases} \quad (36)$$

Тем не менее, правильный выбор методики позволил Канилло и сотр. даже при простейшей $\lambda(t)$ определить неизвестные фазы с точностью 39°.

Из исследовательских работ с применением метода модификации электронной плотности известна работа Мавридис и сотр. [82]. К сожалению, опубликованные результаты по уточнению 2.8-ангстремного набора фаз структурных факторов 2-кето-3-декси-6-фосфоглюконат альдолазы не дают представления об эффективности использования метода. Имеются примеры удачного применения метода для расширения наборов фаз структурных факторов α -химотрипсина [83], транспортных РНК [84, 85] и центральной части нуклеосомы [86].

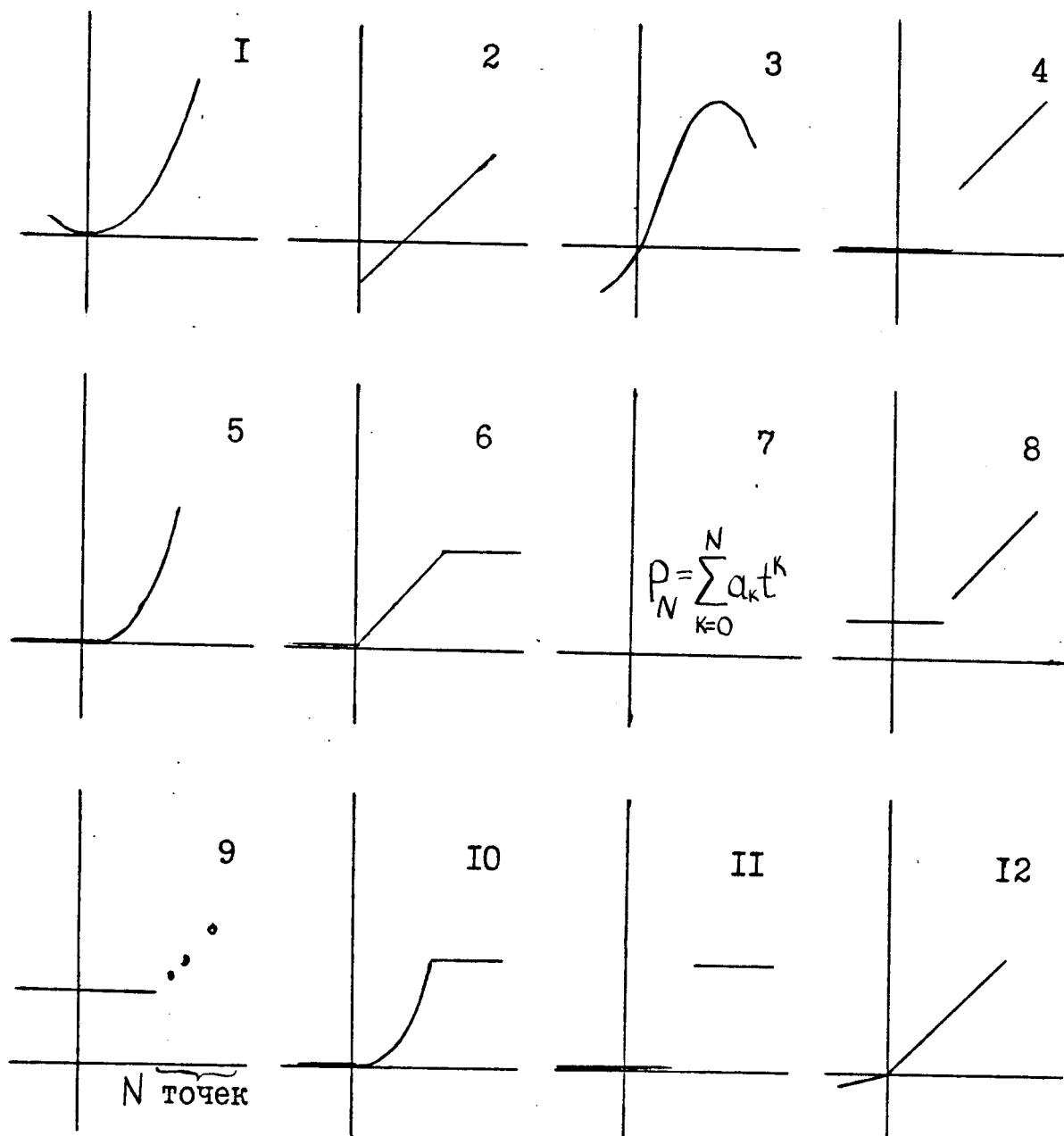


Рис. I. Примеры функций модификации $\lambda(t)$.

- | | |
|---------------------------------------|--------------------------------------|
| I. Сэйр, 1952, [12] | 7. Гассман, 1975, [68] |
| 2. Кураши, 1953, [58] | 8. Симонов, 1976, [70] |
| 3. Хоппе, Гассман, 1968, [60] | 9. Симонов, 1976, [70] |
| 4. Картха, 1969, [65] | 10. Никсон, Норт, 1976, [71] |
| 5. Баррет, Цвик, 1971, [66] | II. Вайнштейн и сотр.,
1977, [72] |
| 6. Гассман, Цехмайстер,
1972, [67] | 12. Шевиш и сотр., 1981, [84] |

I.4. Неотрицательность функции распределения электронной плотности. Принцип максимума определителя.

I.4.I. Непосредственное использование свойства неотрицательности.

Свойство неотрицательности функции распределения электронной плотности резко сужает произвол выбора фаз структурных факторов, и использование этого свойства представляется эффективным в задачах определения и улучшения наборов фаз. Заметим, что при практической работе, когда набор структурных факторов является конечным и имеет ограниченное разрешение, данное свойство функции распределения электронной плотности теряет свою силу (результаты выполненных нами расчетов воспроизведены на рис. 2). Однако, в более или менее грубом приближении можно считать это свойство выполненным и пытаться использовать его для улучшения наборов фаз.

Одним, наиболее распространенным способом учета свойства неотрицательности является выбор соответствующих функций модификации в методе модификации электронной плотности (п. I.3), где идея "исправления" функции $Q(\vec{\zeta})$ на участках ее отрицательности [58] реализовывалась тем или иным образом [65-72, 84].

Другой способ непосредственного использования неотрицательности $Q(\vec{\zeta})$ был предложен в 1976 г. Дэвисом и Роллетом [87]. Исходной информацией в данном методе являются значения модулей $F_o(\vec{s})$ и фаз $\Psi_o(\vec{s})$ некоторого набора отражений $\vec{s} \in P$ и значения модулей $F_o(\vec{s})$ дополнительного набора отражений $\vec{s} \in M$. Метод заключается в поиске на множестве функций, являющихся квадратами тригонометрических полиномов заданной степени, функции, "ближайшей" в пространстве коэффициентов

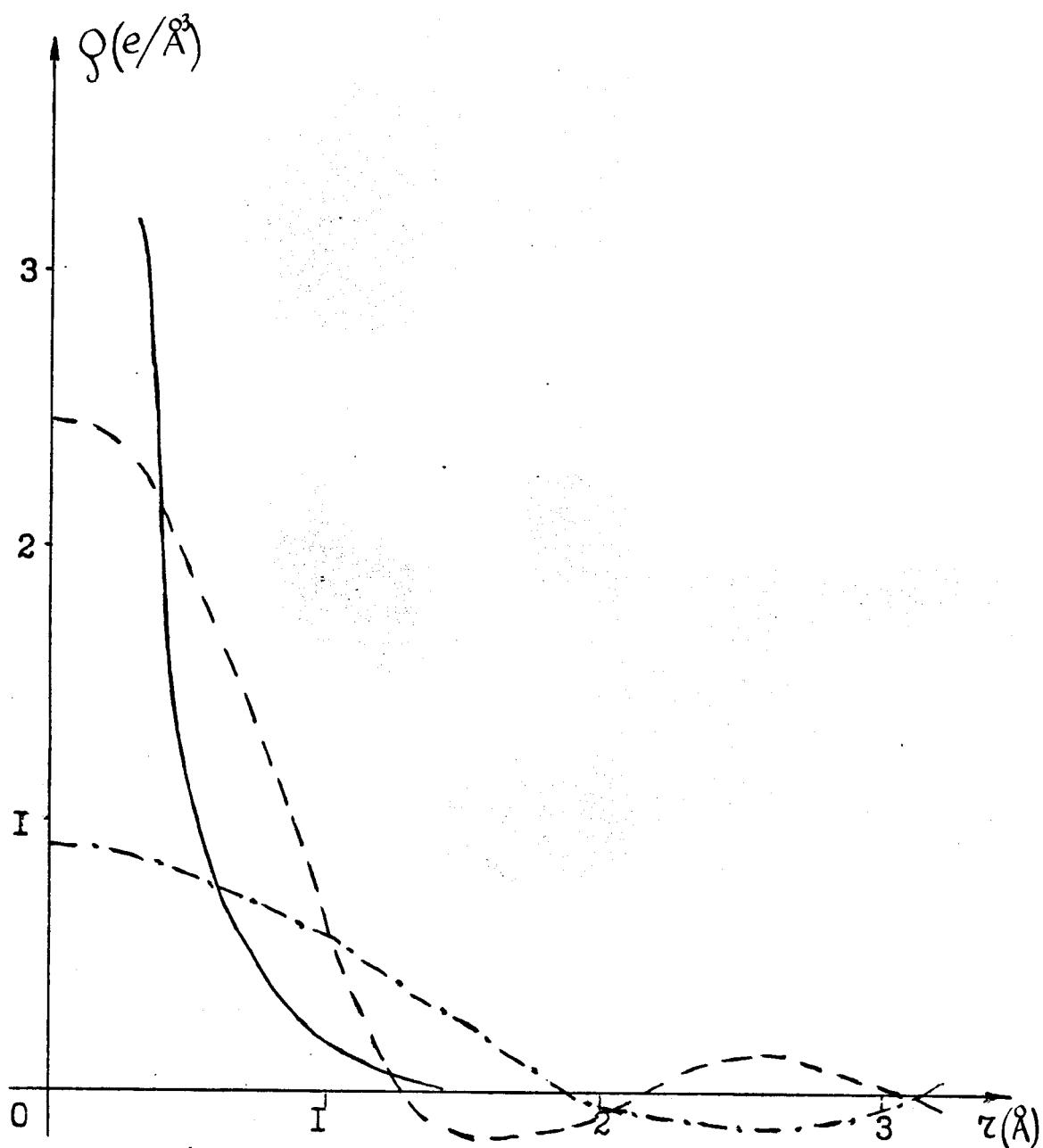


Рис. 2. Изображение покоящегося атома азота при различном разрешении.

- — — идеальное изображение
- - - разрешение 2.0 \AA
- · - · - разрешение 3.0 \AA

Фурье к исходной: $Q(\{\zeta\}) \Rightarrow \min$,

$$Q(\{\zeta\}) = \sum_{\vec{s} \in P} w_{II}(\vec{s}) [F_o(\vec{s}) - F_c(\vec{s}) \cos(\varphi_o(\vec{s}) - \varphi_c(\vec{s}))]^2 + \\ + \sum_{\vec{s} \in P} w_I(\vec{s}) [F_c(\vec{s}) \sin(\varphi_o(\vec{s}) - \varphi_c(\vec{s}))]^2 + \sum_{\vec{s} \in M} w_M(\vec{s}) [M F_o(\vec{s}) - F_c(\vec{s})]^2 \quad (37)$$

Здесь $\zeta(\vec{s})$ – коэффициенты Фурье функции $\hat{Q}(\vec{\zeta})$,

$F_c(\vec{s}) e^{i\varphi_c(\vec{s})} = \sum_{\vec{u}} c(\vec{u}) \zeta(\vec{s} - \vec{u})$ – коэффициенты Фурье искомой функции $Q_c(\vec{\zeta}) = \hat{Q}^2(\vec{\zeta})$,

w_I, w_{II}, w_M, M – весовые функции.

Фазы $\varphi_c(\vec{s})$ коэффициентов Фурье функции $Q_c(\vec{\zeta})$, найденной в результате минимизации функционала (37), используются в качестве улучшенного набора значений фаз структурных факторов.

Следует отметить, что близкий по идеи метод повышения разрешения для задач астрофизики был предложен и проверен на тестах Бираудом в 1969 г. [88].

Дэвис и Роллет опробовали предложенный ими метод на малой структуре, состоявшей из 16 случайным образом расположенных атомов. Проверка метода на данных по большим молекулам проведена не была.

I.4.2. Принцип максимума определителя.

В 1950 г. Карл и Хауптман [89], используя результаты советских математиков Ахиезера и Крейна [90, 91], показали, что свойство неотрицательности $Q(\vec{\zeta}) \geq 0$ эквивалентно свойству неотрицательности определителей матриц специального вида (матриц Карла-Хауптмана):

$$\det K_F[\vec{s}_1, \dots, \vec{s}_m] \geq 0, \quad (38)$$

где

$$K_F[\vec{s}_1, \dots, \vec{s}_m] = \begin{pmatrix} \vec{F}(\vec{s}_1 - \vec{s}_1) & \vec{F}(\vec{s}_1 - \vec{s}_2) & \dots & \vec{F}(\vec{s}_1 - \vec{s}_m) \\ \dots & \dots & \dots & \dots \\ \vec{F}(\vec{s}_m - \vec{s}_1) & \vec{F}(\vec{s}_m - \vec{s}_2) & \dots & \vec{F}(\vec{s}_m - \vec{s}_m) \end{pmatrix}, \quad (39)$$

$\vec{s}_1, \dots, \vec{s}_m$ – произвольные векторы целочисленной решетки \mathbb{R}^l обратного пространства,

M – произвольное натуральное число.

Поскольку "обостренное" распределение $e(\vec{\zeta})$ электронной плотности (16) также неотрицательно, то аналогичное (38) свойство имеет место и для определителей матриц

$$K_E[\vec{s}_1, \dots, \vec{s}_m] = \begin{pmatrix} \vec{E}(\vec{s}_1 - \vec{s}_1) & \vec{E}(\vec{s}_1 - \vec{s}_2) & \dots & \vec{E}(\vec{s}_1 - \vec{s}_m) \\ \dots & \dots & \dots & \dots \\ \vec{E}(\vec{s}_m - \vec{s}_1) & \vec{E}(\vec{s}_m - \vec{s}_2) & \dots & \vec{E}(\vec{s}_m - \vec{s}_m) \end{pmatrix} \quad (40)$$

Нахождение значений фаз структурных факторов больших молекул из системы неравенств (38) являлось затруднительным. Однако в 1969–1970 годах де Ранго и Цукарисом [92–95], исходя из вероятностных соображений, был сформулирован важный принцип, получивший название принципа максимума определителя:

пусть известны значения модулей и фаз всех $\vec{E}(\vec{s}_j - \vec{s}_k)$, $1 \leq j, k \leq m-1$ и значения модулей $E(\vec{s}_j - \vec{s}_m)$, $1 \leq j \leq m-1$; тогда наиболее вероятными значениями фаз $\Psi(\vec{s}_j - \vec{s}_m)$, $1 \leq j \leq m-1$ будут те, которым соответствует максимальное значение определителя $\det K_E[\vec{s}_1, \dots, \vec{s}_m]$.

Этот принцип был положен де Ранго и Цукарисом в основу следующей итерационной процедуры улучшения набора фаз:

- 1) строятся один либо несколько определителей матриц K_E размерности M , где фазы отражений, входящих в главный минор (миноры) порядка $M-1$, известны;
- 2) фазы, соответствующие отражениям последней, M -ой строки и M -го столбца находятся путем максимализации соответствующего определителя (определителей);

3) отражения, фазы которых найдены таким образом, используются для определения фаз следующих отражений.

В дальнейшем метод разрабатывался и усовершенствовался рядом авторов (см., например, [96-101]). Подчеркнем вычислительную сложность метода: для достаточной точности определения фаз при работе с макромолекулами необходимы матрицы порядка нескольких сотен.

Первые работы по улучшению наборов фаз для макромолекул с использованием метода максимума определителя были выполнены авторами метода [102,103]. На данных по миоглобину (4.0\AA) и инсулину (1.9\AA) исследовались практические возможности уточнения и расширения наборов фаз. Тестовые работы были выполнены также Подъярни и сотр. [104] на данных по триклинической модификации лизоцима (расширение набора фаз с 3.3 до 2.5\AA). В дальнейшем метод использовался для улучшения наборов фаз структурных факторов белков [105] и нуклеиновых кислот [99,106-108]. В работах Подъярни и сотр. по исследованию структуры транспортных РНК [99,106-108] определение фаз проводилось не только для отражений более высокого разрешения, но и для ряда сильных рефлексов низкого разрешения. Включение последних в расчет синтезов дало заметный положительный эффект. В этих работах подчеркнуты высокие требования к качеству исходного набора фаз для успешного использования метода и предложены некоторые критерии для оценки такого качества.

В заключение обратим внимание на появление в последнее время ряда работ (например, [109,110]), в которых указывается на связь данного метода с другим методом улучшения карт распределения электронной плотности – методом максимума эн-

тропии [5].

I.5. Использование геометрической информации.

I.5.1. Принцип использования геометрической информации.

Основная идея использования геометрической информации о структуре состоит в определении некоторых соотношений, которым должна удовлетворять функция распределения электронной плотности в заданных точках пространства, и в последующем рассмотрении этих соотношений как системы уравнений относительно фаз структурных факторов [55].

Наиболее простым примером является ситуация, когда известно распределение $\hat{\rho}(\vec{z})$ электронной плотности в некоторой области Ω элементарной ячейки. Математически это требование может быть записано в виде

$$\rho(\vec{z}) = \lambda_1[\rho(\vec{z})] = \begin{cases} \hat{\rho}(\vec{z}), & \vec{z} \in \Omega \\ \rho(\vec{z}) & \text{для остальных точек, } \vec{z} \in \overline{\Omega} \end{cases} \quad (41)$$

Конкретно, если установлено, что молекула локализована в области Ω , то можно приближенно считать, что электронная плотность вне молекулы постоянна

$$\rho(\vec{z}) = \lambda_1^o[\rho(\vec{z})] = \begin{cases} \rho_0, & \vec{z} \in \Omega \\ \rho(\vec{z}), & \vec{z} \in \overline{\Omega} \end{cases} \quad (42)$$

Другим примером соотношений являются ограничения, возникающие из-за наличия в кристалле явления локальной симметрии. Это означает, что для точек из некоторой ограниченной области Ω имеет место равенство

$$\rho(G\vec{z} + \vec{t}) = \rho(\vec{z}), \quad \vec{z} \in \Omega, \quad (43)$$

где G – матрица вращения, \vec{t} – вектор трансляции. Подчеркнем, что, в отличие от кристаллографической симметрии, где (43) справедливо для всех точек пространства, здесь соотношение выполняется лишь локально для $\vec{\zeta} \in \Omega$. Возможно существование в кристалле одновременно нескольких локальных симметрий, но для простоты ограничимся рассмотрением лишь одного такого преобразования.

Свойство (43) эквивалентно следующему равенству

$$Q(\vec{\zeta}) = \lambda_2 [Q(\vec{\zeta})] = \begin{cases} Q(G\vec{\zeta} + \vec{t}), & \vec{\zeta} \in \Omega \\ Q(\vec{\zeta}) & \text{для остальных точек} \end{cases} \quad (44)$$

или, в более симметричном виде,

$$Q(\vec{\zeta}) = \lambda_2^0 [Q(\vec{\zeta})] = \begin{cases} \frac{1}{2} [Q(\vec{\zeta}) + Q(G\vec{\zeta} + \vec{t})], & \vec{\zeta} \in \Omega \\ \frac{1}{2} [Q(G^{-1}(\vec{\zeta} - \vec{t})) + Q(\vec{\zeta})], & \vec{\zeta} \in G\Omega + \vec{t} \\ Q(\vec{\zeta}) & \text{для остальных точек} \end{cases} \quad (45)$$

Более того, можно рассматривать [55] преобразования более общего вида, например,

$$Q(\vec{\zeta}) = \lambda_2^1 [Q(\vec{\zeta})] = \begin{cases} \tau(Q(\vec{\zeta}), Q(G\vec{\zeta} + \vec{t})), & \vec{\zeta} \in \Omega \\ \tau(Q(G^{-1}(\vec{\zeta} - \vec{t})), Q(\vec{\zeta})), & \vec{\zeta} \in G\Omega + \vec{t} \\ Q(\vec{\zeta}) & \text{для остальных точек} \end{cases} \quad (46)$$

где $\tau(x, y)$ – произвольная функция, такая, что

$$\tau(x, x) = x \quad (47)$$

(в преобразовании (45) $\tau(x, y) = \frac{1}{2}(x + y)$).

В обоих разобранных выше примерах учет дополнительной информации сводился к составлению системы уравнений вида

$$Q(\vec{\zeta}) = \lambda [\{Q(\vec{\zeta})\}] \quad (48)$$

или соответствующих уравнений для структурных факторов

$$F(\vec{s})e^{i\varphi(\vec{s})} = \mathcal{F}_{\vec{s}} \left\{ \lambda \left[\left\{ \frac{1}{V} \sum_{\vec{u}} F(\vec{u}) e^{i\varphi(\vec{u})} e^{-2\pi i(\vec{u}, \vec{z})} \right\} \right] \right\}, \quad \vec{s} \in \mathbb{R}^1 \quad (49)$$

где $\mathcal{F}_{\vec{s}} \{ \alpha(\vec{z}) \}$ - преобразование Фурье функции $\alpha(\vec{z})$, вычисленное в точке \vec{s} ,

\mathbb{R}^1 - целочисленная решетка в обратном пространстве.

Полагая модули структурных факторов известными из эксперимента, можно действовать аналогично описанным выше в п.п. I.2-I.4 методам, решая методом итераций фазовую часть системы (49)

$$\varphi(\vec{s}) = \arg \left\{ \mathcal{F}_{\vec{s}} \left\{ \lambda \left[\left\{ \frac{1}{V} \sum_{\vec{u}} F_0(\vec{u}) e^{i\varphi(\vec{u})} e^{-2\pi i(\vec{u}, \vec{z})} \right\} \right] \right\} \right\} \quad (50)$$

Наиболее выгодной с вычислительной точки зрения будет являться следующая итерационная процедура:

1) расчет $Q^j(\vec{z}) = \frac{1}{V} \sum_{\vec{s}} F_0(\vec{s}) e^{i\varphi^j(\vec{s})} e^{-2\pi i(\vec{s}, \vec{z})}$;

2) преобразование $\hat{Q}(\vec{z}) = \lambda \left[\{ Q^j(\vec{z}) \} \right]$;

3) расчет структурных факторов $\hat{F}(\vec{s}) e^{i\hat{\varphi}(\vec{s})}$ функции $\hat{Q}(\vec{z})$;

4) определение новых значений фаз $\varphi^{j+1}(\vec{s}) = \hat{\varphi}(\vec{s})$.

Априорно при этом сохраняются общие недостатки, отмеченные выше при рассмотрении некоторых методов улучшения наборов фаз:

1) не используется половина уравнений ("модульная" часть), т.е. теряется часть информации;

2) возникает возможность существования самосогласованных решений, не удовлетворяющих исходным требованиям;

3) возможно отсутствие сходимости метода простой итерации.

Однако, при данном способе учета геометрической информации

такие недостатки, по-видимому, не являются решающими, и описанная выше итерационная процедура успешно использовалась в большинстве практических работ (см., например, обзор [III]).

Вопросы использования конкретно локальной (некристаллографической) симметрии подробно разобраны Вагиным [II2].

В заключение необходимо отметить, что, хотя при выводе различных соотношений геометрическая информация записывается в виде точных равенств, такая запись является лишь некоторым приближением. В частности, уравнения локальной симметрии неточны, во-первых, по физической природе явления, во-вторых, в силу погрешности определения параметров преобразования симметрии и, в-третьих, в силу ограниченности разрешения.

I.5.2. Примеры использования геометрической информации.

Впервые возможность использования геометрической информации для решения фазовой проблемы была продемонстрирована Россманом и Блоу [II3-II5] в начале 60-х годов. Мэйном и Россманом [II6,II7] и Кроутером [II8,II9] был предложен способ решения системы соответствующих уравнений для структурных факторов в обратном пространстве. Несколько позднее был предложен фактически эквивалентный, но существенно более простой алгоритм решения системы с помощью перехода в прямое пространство [I20-I22], описанный в предыдущем пункте. В настоящее время наиболее известными реализациями этого алгоритма являются программы Бриконя [I23,I24] и Джонсона [I25]. В СССР ряд работ был выполнен с использованием комплекса программ Вагина [II2].

Первые практические работы по использованию геометрических свойств структуры [I26-I29] являются фактически работами по улучшению карт распределения электронной плотности,

но не по улучшению наборов фаз, так как исследователи ограничивались в них лишь усреднением функции распределения электронной плотности без последующего пересчета фаз. Пересчет фаз впервые был проведен в 1968 г. Мьюрхедом и сотр. [130] при исследовании структуры деоксигемоглобина человека на данных разрешением до 5.5 Å.

Активное использование данного метода началось в середине 70-х годов. В большом количестве работ по исследованию пространственной структуры белков [123, 124, 131-154] продемонстрирован широкий диапазон применимости метода. Некоторые методические особенности содержат работы Флеттерика и Штейца [138-140] и Агарда и Штроуда [151]. Среди работ советских исследователей известны работы по улучшению наборов фаз структурных факторов каталазы *Penicillium Vitale* (разрешение 3.0 Å), аспартат-аминотрансферазы (3.5 Å) [112] и рибонуклеазы B_i (3.2 Å) [155]. Отметим, что при использовании этого метода существенную роль играет правильное проведение границ молекулы [132, 156-158], что порой является затруднительным.

Богатые возможности метода продемонстрированы Шевицем, Подъярни и сотр. [84, 85] при расширении набора фаз для транспортных РНК. Им удалось повысить разрешение карт распределения электронной плотности с 14.5 до 4.0 Å и включить в расчет новых карт дополнительно несколько сильных отражений низкого разрешения, для которых фазы также были определены этим методом.

Многочисленными являются работы, содержащие этап улучшения набора фаз методом усреднения по локальной симметрии для вирусов [125, 157-176], где такая симметрия встречается

наиболее часто. Интересна одна из первых работ по исследованию структуры вирусов [159], где проводилось усреднение как в прямом, так и в обратном пространстве. Следует также выделить работы Рэймента и сотр. [157,158,172], в которых анализ каждого из этапов метода и влияния различных факторов на результат улучшения выполнен наиболее тщательно.

I.6. Использование моделей молекулы.

I.6.1. Общие замечания.

Построение модели молекулы заключается в выборе гипотезы о том, как может быть устроена такая модель, и в построении в рамках определенного выше класса моделей некоторой конкретной, наилучшим образом соответствующей доступной информации о молекуле. Основная идея использования моделей для решения фазовой проблемы состоит в расчете структурных факторов по той или иной модели и приписывании их фаз экспериментально определенным значениям модулей. Позволяя фазам принимать не произвольные значения, а только те, которые соответствуют моделям определенного типа, мы тем самым существенно сужаем класс возможных решений за счет информации, которая учитывалась при выборе этого типа моделей.

В качестве примера можно рассмотреть моделирование рассеяния рентгеновских лучей на атомных моделях. Пусть известно, что структура описывается атомной моделью, состоящей из N рассеивающих центров ("атомов"), причем каждый из них характеризуется четырьмя параметрами (координатами центра и изотропным температурным фактором). Тогда все допустимые наборы фаз являются элементами подмножества, в котором допустима $4N$ -мерная параметризация. В то же время исходное

описание структуры в терминах структурных факторов определяется значениями, вообще говоря, бесконечного набора комплексных величин. Предъявление к атомной модели дополнительных требований, например, наличия локальной симметрии, определяет дальнейшее сужение множества допустимых значений фаз.

Наиболее простой путь использования моделей состоит, таким образом, в организации итерационной процедуры:

- 1) расчет функции распределения электронной плотности по набору экспериментально определенных значений модулей $\{F_0(\vec{s})\}$ и имеющемуся набору значений фаз $\{\varphi^j(\vec{s})\}$;
- 2) построение по картам электронной плотности некоторой (не обязательно атомной) модели;
- 3) расчет структурных факторов $F_C(\vec{s})e^{i\varphi_C(\vec{s})}$ по построенной модели;
- 4) использование $\varphi_C(\vec{s})$ в качестве новых значений фаз:
$$\varphi^{j+1}(\vec{s}) = \varphi_C(\vec{s}).$$

Более общий подход получается из тех соображений, что введение параметризации позволяет уточнять модель, построенную на этапе 2), и тем самым уточнять значения набора фаз $\{\varphi_C(\vec{s})\}$.

Выбор модели в каждом конкретном случае должен учитывать, с одной стороны, ограниченное количество экспериментальных данных, а с другой стороны – возможность более точного моделирования при увеличении числа параметров.

Подчеркнем, что в качестве модели может использоваться модель произвольного типа – атомная, "крупноблочная", модель электронной плотности и т.п. Богатый выбор моделей обуславливает широкие возможности метода.

I.6.2. Метод молекуллярного замещения.

Один из широко распространенных методов определения

значений фаз – метод молекулярного замещения (см., например, [III]) – может быть рассмотрен в рамках описанного подхода. В качестве модели белка здесь выбирается модель (атомная, электронной плотности) известного белка, близкого по структуре к исследуемому. Задача определения положения выбранной модели в ячейке исследуемого объекта решается с помощью функций вращения, трансляции и упаковки (см., например, [4, II2]). Таким образом, можно рассматривать метод молекулярного замещения как метод уточнения значений фаз структурных факторов путем уточнения параметров расположения рассеивающей модели в ячейке исследуемой структуры.

I.6.3. Использование "крупноблочных" моделей.

Исследование структур макромолекул при низком разрешении не позволяет вести построение и анализ атомных моделей. В этой ситуации рассеяние должно моделироваться на объектах, элементы которых соответствуют целой группе атомов. Такие модели, называемые "крупноблочными", удобны для работы потому, что они обычно задаются сравнительно небольшим числом параметров.

Впервые модель такого типа, представляющая собой шаровой слой, была использована Джонсоном и сотр. [I63] в 1976г. для определения набора фаз структурных факторов разрешением до 22.5 \AA вириуса южной бобовой мозаики. В серии работ Рэймента и сотр. [I57, I58, I72] фазы структурных факторов капсиды вириуса полиоми рассчитывались по модели, состоящей из совокупности пустотелых цилиндров. Параметры модели (внешний и внутренний радиусы и высота цилиндров) выбирались из условия наилучшего соответствия рассчитанных по модели модулей структурных факторов экспериментально определенным вели-

чинам. Полученный таким образом набор фаз использовался далее в качестве исходного для последующего уточнения методом усреднения функции распределения электронной плотности по локальной симметрии.

Более интересными по отношению к вопросу улучшения наборов фаз являются работы [7, 8, I77, I78]. Для белков типа миоглобина, состоящих в основном из α -спиралей и гема, в этих работах предлагается моделировать рассеяние на цилиндрах, закрученных с торцов. Параметры цилиндров (размеры, ориентация, масса) уточняются, исходя из требования минимальности функционала

$$\Phi = (1-t)\Phi_2 + t\Phi_D , \quad (51)$$

включающего в себя как соответствие рентгеновскому эксперименту

$$\Phi_2 = \sum_{\vec{s}} \left(F_C(\vec{s}) - F_O(\vec{s}) \right)^2 , \quad (52)$$

так и контроль за взаимным расположением элементов модели

$$\Phi_D = \frac{q}{V} \int Q_1(\vec{\varepsilon}) \left(\sum_{j=2}^{N_s} Q_j(\vec{\varepsilon}) \right) dV \quad (53)$$

Здесь $Q_j(\vec{\varepsilon})$ – функция распределения электронной плотности молекулы, полученной из исходной j -м преобразованием симметрии,

N_s – количество преобразований симметрии,

V – объем элементарной ячейки,

q, t – некоторые весовые множители.

Минимизация функционала Φ проводится методами нелокального поиска [I79].

Проверка метода на структуре леггемоглобина люпина жел-

того [7] и расшифровка структуры ромбической модификации леггемоглобина [177] продемонстрировали эффективность предложенной методики для расшифровки при низком разрешении структур белков типа миоглобина.

Недавно Рабинович и Шаккед [180] показали возможность применения аналогичного подхода к решению фазовой проблемы при низком разрешении для произвольных макромолекул.

1.6.4. Использование частичных атомных моделей.

При исследовании макромолекул в большинстве случаев по имеющимся данным удается построить атомную модель лишь некоторой части структуры. Если такая модель является достаточно полной (определено около 90% общего числа неводородных атомов), то фазы структурных факторов, рассчитанных по ней, по качеству обычно превосходят фазы, определяемые, например, с помощью метода изоморфного замещения, на что было обращено внимание в работе Ватенко и сотр. [181]. (Влияние полноты атомной модели и наличия ошибок в значениях координат центров атомов на точность воспроизведения фаз продемонстрировано в таблице I. Расчеты выполнялись в НИВЦ АН СССР Е.А. Верносоловой на атомных моделях актинидина. В координаты центров атомов вносились случайные ошибки, распределенные сферически симметрично с нормальным законом распределения и заданными дисперсиями, дающими различные величины среднего смещения атомов от точных позиций. В наименее полной модели оставлены лишь атомы главной цепи. Остальные модели состояли из атомов главной цепи и соответственно 50%, 75% и 100% атомов боковых групп.) Таким образом, набор фаз структурных факторов, рассчитанных по частичной модели, может использоваться либо для расчета новых карт распределения электронной плотности с

Таблица I. Влияние полноты атомной модели и ошибок определения координат атомов на точность воспроизведения фаз структурных факторов (на примере данных по актинидину).

		Интервалы разрешения по $ \tilde{S} ^2$										
		.000-	.025-	.050-	.075-	.100-	.125-	.150-	.175-	.200-	.225-	.250-
Модель ошибки	пол- $\langle \Delta \vec{r} \rangle$.000-	.025-	.050-	.075-	.100-	.125-	.150-	.175-	.200-	.225-	.000-
0.0 Å	100%	0	0	0	0	0	0	0	0	0	0	0
	90%	24	23	19	20	20	19	20	18	18	18	19
	75%	33	35	28	28	28	29	30	28	26	27	29
0.4 Å	50%	40	44	44	43	42	40	43	40	37	38	41
	100%	10	16	21	26	32	34	39	41	39	42	34
	90%	25	28	27	33	36	38	43	44	43	46	39
0.8 Å	75%	34	37	34	39	41	44	47	48	45	49	44
	50%	38	50	49	49	52	51	55	53	52	55	52
	100%	19	32	39	50	58	61	67	70	69	72	60
Количество отражений	90%	28	38	43	54	61	62	68	70	70	73	62
	75%	35	47	48	57	62	65	70	72	70	75	65
	50%	39	56	58	63	67	67	73	73	74	77	69
	320	681	957	1078	1278	1475	1502	1691	1783	1948	12713	

целью корректировки модели, либо в качестве исходного для уточнения тем или иным способом [71, 182-185]. Более гибким является подход, включающий предварительное уточнение частичной модели и расчет фаз по уже уточненной частичной модели [186-195]. Влияние полноты исходной модели на результат уточнения наглядно продемонстрировано Ремингтоном и сотр. [195].

Отдельного упоминания заслуживают работы Райса [196, 197] по исследованию фосфоглицерат киназы при среднем разрешении (2.5 Å). Сложность этого исследования заключалась в отсутствии информации о первичной последовательности и существенной неполноте модели (было идентифицировано лишь около 2/3 общего числа неводородных атомов). Тем не менее, путем тщательного уточнения этой модели по программе Хендриксона и Коннера [198] и комбинирования фазовой информации с исходной, определенной по методу изоморфного замещения, удалось провести улучшение набора фаз и построение полной атомной модели по улучшенным синтезам. По-видимому, здесь оказалось существенным использование стереохимических ограничений при уточнении частичной модели.

I.6.5. Использование грубых атомных моделей.

В ряде исследований построение даже частичной атомной модели макромолекулы оказывается затруднительным из-за низкого качества карт распределения электронной плотности при среднем разрешении. В то же время значительное количество модулей структурных факторов более высокого разрешения остаются неиспользованными при расчете этих карт, поскольку оказалось невозможным определить для них значения соответствующих фаз. Агарвалом и Айзексом [6] в 1977 г. был предложен

оригинальный метод расширения набора фаз в данной ситуации. Ключевая идея метода состоит в построении по имеющемуся исходному синтезу электронной плотности атомной модели, фиктивные "атомы" которой не обязаны соответствовать каким-либо конкретным атомам структуры, а лишь используются для моделирования рассеяния рентгеновских лучей. Основное требование, предъявляемое к такой модели, заключается в достаточно хорошем воспроизведении исходного синтеза, чего более или менее удачно можно добиться, располагая фиктивные атомы в точках локальных максимумов функции распределения электронной плотности. Для построения такой модели могут быть предложены точные алгоритмы, и, следовательно, возможно ее автоматическое быстрое построение.

Построенную модель предлагалось уточнять по полному набору экспериментально определенных модулей структурных факторов путем минимизации [199]

$$R_2 = \sum_{\vec{s}} w_{\vec{s}} (F_c(\vec{s}) - F_o(\vec{s}))^2 \quad (54)$$

При этом фазы структурных факторов, рассчитанных по модифицированной модели, могут использоваться для построения новых, улучшенных карт распределения электронной плотности уже более высокого разрешения.

Описанный выше метод был опробован на данных по инсулину для расширения набора фаз с 3.0 до 1.9 и до 1.5 Å. В качестве точных значений были взяты значения фаз структурных факторов, рассчитанных по ранее уточненной атомной модели инсулина. Результатом проведенного расширения исходного набора фаз, определенных методом изоморфного замещения, явился набор фаз со средним отклонением от точных значений 65° по

отражениям разрешением $1.9 \text{ \AA} \leq d < 3.0 \text{ \AA}$ и 70° - по полному набору данных.

Эванс и Худсон [200] использовали этот метод для уточнения значений фаз структурных факторов фосфофруктокиназы (разрешение 2.4 \AA). К сожалению, опубликованная информация не дает сведений ни о конкретных деталях работы, ни о конечном эффекте уточнения.

Дальнейшего развития со стороны зарубежных исследователей, в том числе и его авторов, метод не получил. Лишь совсем недавно появилась работа Бентли и Льюит-Бентли [86], в которой α -спиральные участки моделировались цепочкой специальным образом расположенных "атомов", которые затем уточнялись по данным низкого разрешения (25 \AA , 16 \AA).

В начале 80-х годов Луниным и Уржумцевым [201] были выполнены реализация описанной выше идеи Агарвала и Айзекса и ее дальнейшее развитие. Здесь от модели не требовалось не только поатомного соответствия структуре, но и выполнения каких-либо стереохимических ограничений. Такие модели, используемые лишь для моделирования рассеяния рентгеновских лучей, получили название грубых моделей. Проведенные численные эксперименты на данных по структуре актинидина продемонстрировали неэффективность ряда практических рекомендаций Агарвала и Айзекса. Предложенная Лунинным и Уржумцевым новая методика позволила определить по значениям фаз 464I структурного фактора низкого и среднего разрешения (до 3.0 \AA) и по значениям модулей структурных факторов более высокого разрешения (до 2.0 \AA) значения неизвестных фаз с точностью 35° [201]. В другом эксперименте - по уточнению зашумленного набора фаз - удалось снизить ошибку с 37° до 21° для фаз

структурных факторов разрешением до 3.0 Å [202].

Разработанное программное обеспечение и методика успешно использовались для расширения наборов фаз структурных факторов γ -кристаллина Шб с 3.0 Å до 2.7 Å [203-206] и аспартат-аминотрансферазы с 3.5 Å до 2.8 Å.

I.6.6. Использование смешанных моделей.

В ситуации, когда использование грубой модели уже не дает существенного улучшения карт распределения электронной плотности (или метод почему-либо не может использоваться), а частичная модель еще недостаточно полная, дальнейшего прогресса в уточнении фаз и, следовательно, в уточнении и дополнении частичной модели можно достичь за счет использования смешанной модели, состоящей из атомов частичной модели и какой-либо дополнительной грубой модели. Общий алгоритм использования смешанной модели близок к описанным выше алгоритмам для частичной и грубой моделей и состоит из этапов:

- 1) построение (корректировка, дополнение) частичной модели;
- 2) построение смешанной модели путем дополнения частичной модели некоторой грубой на тех участках, где частичная модель отсутствует;
- 3) модификация смешанной модели либо отдельных ее компонентов (данный этап может и отсутствовать);
- 4) использование фазовой информации, полученной по (модифицированной) смешанной модели, для определения уточненных значений фаз.

Заметим, что с улучшением набора фаз появляется возможность дополнить частичную модель. Поэтому описанный выше итерационный процесс позволяет, вообще говоря, постепенно перехо-

дить от улучшения фаз к обычному кристаллографическому уточнению атомной модели.

Один из вариантов данного подхода был реализован в 1982 г. Батом и Блоу [207]. В качестве грубой модели ими использовалась модель электронной плотности, которая строилась по исходной функции распределения электронной плотности путем зануления ее вне области молекулы. Область молекулы $\Omega_{\text{МОЛ}}$ определялась как совокупность связных областей, обладающих следующими свойствами:

- 1) значения функции распределения электронной плотности в области выше некоторого заданного уровня;
- 2) размеры области не меньше некоторой заданной величины;
- 3) область должна быть связана с областью $\Omega_{\text{РМ}}$, занимаемой частичной моделью; сама область $\Omega_{\text{РМ}}$ в область молекулы $\Omega_{\text{МОЛ}}$ не включается [208].

Отметим, что Бат и Блоу допускали отсутствие частичной модели. В этом случае при построении грубой модели пункт 3) не рассматривается, а метод фактически вырождается в метод модификации функции распределения электронной плотности со специальным образом строящейся функцией модификации.

В результате, метод Бата и Блоу сводится к построению по исходному синтезу электронной плотности $Q^j(\vec{\zeta})$ новой функции

$$\hat{Q}(\vec{\zeta}) = \begin{cases} Q_{\text{РМ}}(\vec{\zeta}), & \vec{\zeta} \in \Omega_{\text{РМ}} \\ Q^j(\vec{\zeta}), & \vec{\zeta} \in \Omega_{\text{МОЛ}} \\ 0 & \text{для остальных точек} \end{cases}, \quad (55)$$

где $Q_{\text{РМ}}(\vec{\zeta})$ – функция распределения электронной плотности, рассчитанная по частичной модели. Структурные факторы, рас-

считанные по $\hat{Q}(\vec{\zeta})$, путем комбинирования полученной информации с исходной используются для определения нового набора фаз $\{\varphi^{j+1}(\vec{s})\}$, который, в свою очередь, используется вместе с экспериментально определенными значениями модулей $F_0(\vec{s})$ для построения нового синтеза $Q^{j+1}(\vec{\zeta})$.

Метод Бата и Блоу был успешно использован при исследовании тирозил-тРНК-синтетазы [207, 209, 210].

Важным шагом в последующем развитии метода явилось предложение его создателей после построения смешанной модели проводить уточнение ее атомной части, зафиксировав грубую модель электронной плотности [211]. Введение этого этапа сделало метод Бата и Блоу более гибким и универсальным.

Луниным и сотр. [212] был разработан метод улучшения набора фаз путем построения и модификации смешанной атомной модели. В этом методе частичная модель дополняется грубой атомной моделью, построенной исходя из требования удовлетворительного воспроизведения непроинтерпретированной части исходного синтеза электронной плотности. Существенным достоинством такой модели является возможность модификации как частичной, так и грубой модели, что дает дополнительные возможности по сравнению с методом Бата и Блоу. С помощью метода смешанной атомной модели было выполнено уточнение значений фаз структурных факторов χ -кристаллина Ш при разрешении 2.7 Å [212, 213].

I.6.7. Заключительные замечания.

Практически независимое и одновременное развитие различных методик уточнения значений и расширения наборов фаз для макромолекул с использованием грубых и частичных моделей [6, 7, 196, 201, 207, 211, 212], по-видимому, связано с общей

ситуацией, создавшейся в кристаллографии макромолекул, когда по данным, например, метода изоморфного замещения оказывается возможным построить в лучшем случае атомную модель лишь части молекулы. Дальнейшее развитие этих и аналогичных методов следует, по-видимому, связывать с привлечением дополнительной информации, например, стереохимических сведений об организации биологических макромолекул. Усложнение критериев качества модели и увеличение размеров исследуемых структур требуют, в свою очередь, развития применяемых вычислительных методов. В частности, при программной реализации методов, использующих модификации моделей, необходимо активно использовать алгоритм быстрого преобразования Фурье [33, 214] и алгоритм быстрого дифференцирования [52-55].

I.7. Комбинирование дополнительной информации различных типов.

I.7.I. Возможные способы использования дополнительной информации различных типов.

При выполнении практических работ естественным является желание исследователя использовать всю имеющуюся дополнительную информацию об изучаемой структуре. В зависимости от конкретной ситуации такое использование может реализовываться различными способами. Условно их можно разделить на следующие группы:

- 1) одновременное использование информации различных типов в одном и том же методе;
- 2) последовательное применение различных методов улучшения набора фаз;

3) комбинирование фазовой информации, полученной различными методами.

Наиболее типичным примером первой группы способов является одновременный учет границ молекулы и явления локальной симметрии в методе усреднения функции распределения электронной плотности [122]. Несложно также учесть информацию о расположении молекулы и в методе модификации функции распределения электронной плотности [84]. В принципе понятно, как можно учесть известные границы молекулы или наличие локальной симметрии в методах, использующих те или иные модели (см., например, [203]). Подробнее на этой группе способов мы останавливаться не будем, так же, как и на второй группе потому, что последовательное применение различных методов улучшения набора фаз является в идеином плане наиболее понятным, хотя здесь может возникнуть ряд методических вопросов.

Отдельного внимания заслуживают вопросы комбинирования фазовой информации, тем более, что имеются примеры как удачного [44], так и неудачного использования [215] такого комбинирования.

I.7.2. Комбинирование фазовой информации.

Вопрос комбинирования фазовой информации возникает, как правило, в ситуации, когда в распоряжении исследователя имеется несколько (например, две) функций распределения вероятностей значений фаз $P_j(\varphi; \vec{s})$, полученных при различной дополнительной информации и исходных гипотезах. Возникает задача построения комбинированного распределения

$$P(\varphi; \vec{s}) = f(P_1(\varphi; \vec{s}), \dots, P_M(\varphi; \vec{s})), \quad (56)$$

которое будет иметь среднеквадратичное отклонение меньшее, чем среднеквадратичные отклонения каждого из распределений $P_j(\varphi; \vec{s})$. Такое уменьшение может достигаться, например, за счет того, что для мультимодальных распределений сохраняются лишь общие пики.

Объединение фазовой информации обычно производят путем перемножения соответствующих распределений

$$P(\varphi; \vec{s}) = \prod_{j=1}^M P_j(\varphi; \vec{s}) \quad (57)$$

В большинстве работ такой подход либо вообще не обосновывается, либо приводится ссылка на "независимость распределений", хотя последнее не всегда соответствует действительности.

Практическое использование комбинирования информации сводится обычно к его применению совместно с каким-либо способом улучшения набора фаз, когда вновь полученная фазовая информация комбинируется с исходной, содержащей, например, данные о дополнительных экспериментах с тяжелоатомными производными. Конкретные реализации [122, 139, 171, 196, 205] связаны с выбором вида функций $P_j(\varphi; \vec{s})$, оценкой параметров этих распределений и подходом к построению комбинированного распределения.

I.7.3. Определение распределений вероятностей фаз.

Первые работы, в которых приводились различные формы представления функции распределения вероятностей фаз, найденных методом изоморфного замещения [216, 217] и рассчитанных по атомным моделям [218-220], появились на рубеже 50-х и 60-х годов. Более общая форма записи

$$P(\varphi; \vec{s}) = \exp \{A\vec{s} \cos \varphi + B\vec{s} \sin \varphi + C\vec{s} \cos 2\varphi + D\vec{s} \sin 2\varphi\}, \quad (58)$$

удобная для хранения и комбинирования информации, была предложена в 1970 г. Хендриксоном и Латтманом [221].

При представлении распределения вероятностей в форме (58) возникает вопрос об оценке параметров распределения $\{A_{\vec{s}}, B_{\vec{s}}, C_{\vec{s}}, D_{\vec{s}}\}_{\vec{s} \in S}$. Хендриксон и Латтман [221] дали способ таких оценок для фаз, определенных методами изоморфного замещения и аномального рассеяния. Для фаз, определенных прямыми методами, способы оценки параметров распределения указаны в [16, 222, 223].

Для распределений вероятностей фаз, рассчитанных по моделям, обычно используется упрощенный вариант формы (58) с $C_{\vec{s}} = D_{\vec{s}} = 0$:

$$P(\varphi; \vec{s}) = \alpha_{\vec{s}} \exp\{A_{\vec{s}} \cos \varphi + B_{\vec{s}} \sin \varphi\} = \alpha_{\vec{s}} \exp\{t_{\vec{s}} \cos(\varphi - \varphi_c(\vec{s}))\} \quad (59)$$

Возможный подход к оценке параметров $\{t_{\vec{s}}\}$ дан в работе Сима [219]:

$$t_{\vec{s}} = 2F_c(\vec{s})F_o(\vec{s}) / \sum_{j=1}^N f_j^2(|\vec{s}|) \quad (60)$$

(где $f_j(|\vec{s}|)$ - фактор рассеяния j -го атома, а сумма взята по всем атомам элементарной ячейки), а также в работах индийских исследователей [224].

Бриконем [123] был предложен эмпирический подход к оценке параметров распределения вероятностей фаз (59), рассчитанных по функции распределения электронной плотности.

$t_{\vec{s}}$ для такого распределения представляется ступенчатой функцией

$$\hat{t}(\vec{s}) = \left\langle |I_o(\vec{s}) - \theta I_c(\vec{s})| \right\rangle_{S_{j-1} \leq |\vec{s}| < S_j}^{-1} \quad \text{для } S_{j-1} \leq |\vec{s}| < S_j \quad (61)$$

Здесь $I_0(\vec{s})$ и $I_c(\vec{s})$ - экспериментально определенная и расчитанная по модели интенсивности отражения \vec{s} , s_j - точки разбиения интервала разрешения, $0 < \theta < I$ - некоторый коэффициент. В работе [171] предложена некоторая модификация оценки (61).

Оценка параметров $t\vec{s}$ (61) по Бриконю, несмотря на свой эмпирический характер, получила достаточно широкое распространение и использовалась также при оценке параметров распределения вероятностей фаз, полученных по частичной атомной модели [196].

Альтернативный способ оценки параметров $t\vec{s}$, основанный на принципе максимума правдоподобия, для распределений вероятностей фаз структурных факторов, рассчитанных по атомной модели, был предложен Лунинным [225].

Е.А.Вернословой в НИВЦ АН СССР было проведено сравнение оценок параметров $t\vec{s}$ описанными выше способами. По атомной модели актинида был рассчитан набор фаз Φ_0 . Затем в модель были внесены случайные независимые ошибки (среднее отклонение центров атомов от точных положений составило 0.4 \AA в одном случае и 0.8 \AA - в другом), по этим моделям были рассчитаны наборы фаз - соответственно Φ_1 и Φ_2 . Вторая, наиболее зашумленная модель M_2 уточнялась в обратном пространстве последовательно по данным разрешением до 3.0 \AA (соответствующий набор фаз - Φ_3) и до 2.0 \AA (набор фаз Φ_4). Анализ результатов сравнения оценок среднеквадратичных отклонений фаз, основанных на оценке $t\vec{s}$ тем либо иным способом (рис. 3), приводит к выводу, что в настоящее время отсутствуют способы надежной оценки параметров $t\vec{s}$ распределений (59), определяемых по уточняющейся атомной модели.

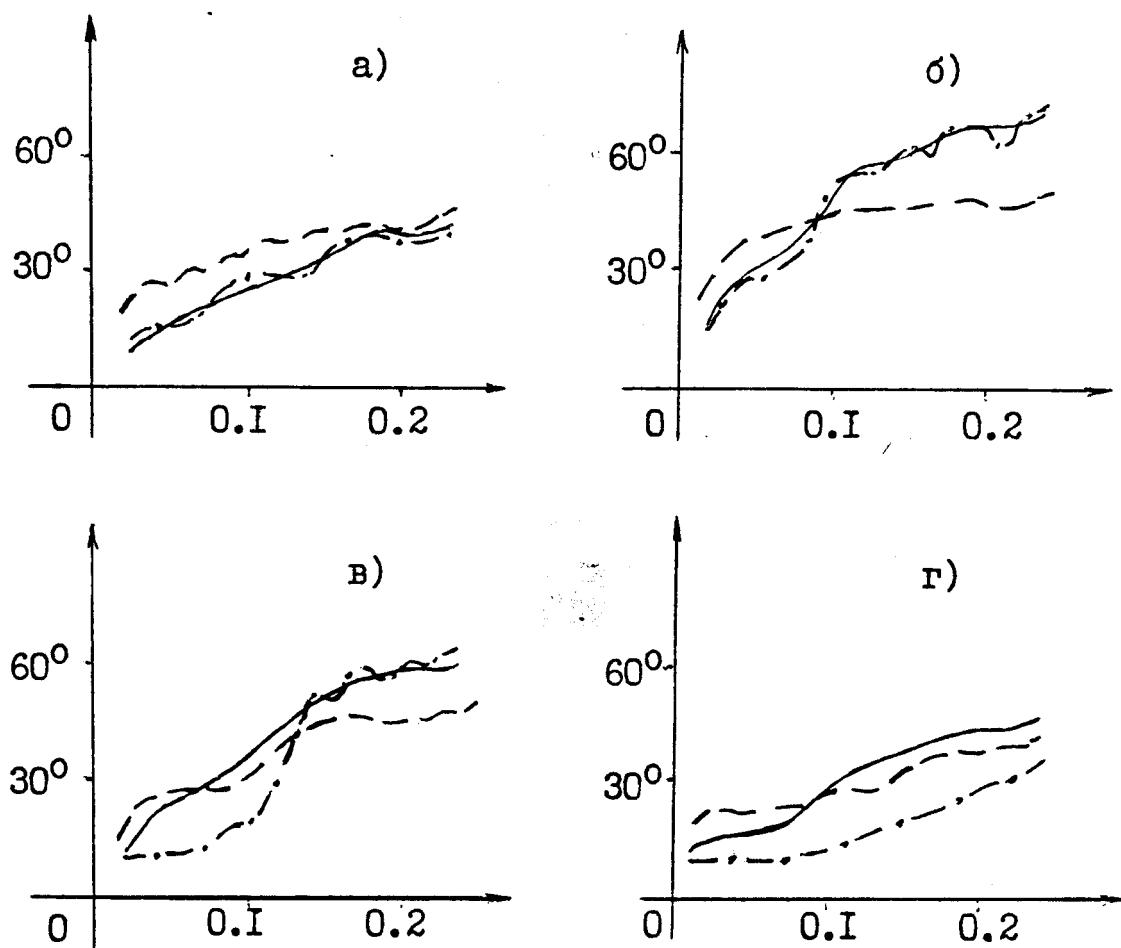


Рис. 3. Сравнение оценок погрешностей фаз структурных факторов, рассчитанных по различным моделям белка (на примере данных по актинидину).

- модель M_1 , среднее смещение атомов 0.4 \AA ;
- модель M_2 , среднее смещение атомов 0.8 \AA ;
- модель M_3 (уточнение M_2 по данным $d \geq 3.0 \text{ \AA}$);
- модель M_4 (уточнение M_3 по данным $d \geq 2.0 \text{ \AA}$).

— — — истинная погрешность
— - - - оценка по Бриконю
— - - . - оценка из принципа максимума правдоподобия

I.8. Сравнение методов улучшения наборов фаз структурных факторов для макромолекул.

Сравнение результатов применения различных методов улучшения наборов фаз представляет собой значительный интерес особенно в силу большого разнообразия этих методов. Такое сравнение важно для определения достоинств и недостатков методов и границ их применимости.

Непосредственный анализ результатов тестирования различных методов вызывает определенные трудности потому, что тестирование выполнялось на различных объектах, исходные и контрольные наборы данных определялись по-разному, выбирались различные средства контроля и т.п. В частности, приходится отказаться от использования работ, выполненных на малых структурах (в том числе весьма интересных работ [26, 76, 183]), поскольку выше уже отмечалась специфика методов улучшения наборов фаз для макромолекул. Также затруднителен анализ работ, где единственным критерием является визуальное сравнение синтезов (например, [71, 167]), в силу субъективности данного критерия, хотя при практических исследованиях именно такое сравнение является решающим критерием успеха или неудачи улучшения набора фаз.

Данные по ряду численных экспериментов включены в таблицу 2. Из работ с большим количеством экспериментов, например, [157], в таблице приведены данные лишь по наиболее существенным.

В таблице 2 использованы следующие обозначения:

\mathcal{Y}_{Mk} (\mathcal{Y}_{Sk} , \mathcal{Y}_{Dk}) — набор фаз, определенных по методу изоморфного замещения (соответственно по одному или двум изоморфным производным)

Ψ_M - набор фаз, рассчитанных по атомной модели

Ψ_{CM} - набор фаз, рассчитанных по загруженной модели

Ψ^N - зашумленный набор фаз

$(\Psi^{LS}, \Psi^{TAN}, \Psi^{AV}, \Psi^{COMB})$ Ψ^{REF} - уточненный набор фаз (в том числе соответственно путем решения системы уравнений Сэйра, по тангенс-формуле, по методу усреднения по локальной симметрии, из комбинирования полученного распределения с исходным)

F_0, F_M - наборы модулей структурных факторов, экспериментально определенных и рассчитанных по модели

$\langle \Psi - \Psi_0 \rangle = \sum_{S \in S} \Delta \Psi(S) / M(S)$, где $\Delta \Psi(S) = \min \{ |\Psi(S) - \Psi_0(S)|, 360 - |\Psi(S) - \Psi_0(S)| \}$,
а $M(S)$ - количество элементов множества S

$\langle \Psi - \Psi_0 \rangle_0, \langle \Psi - \Psi_0 \rangle_+, \langle \Psi - \Psi_0 \rangle_-, \langle \Psi - \Psi_0 \rangle_K$ - средние отклонения фаз соответственно по исходному набору отражений, по добавленным отражениям, по полному набору отражений, по K наиболее "сильным" отражениям

$\langle \Psi - \Psi_0 \rangle^{CS}, \langle \Psi - \Psi_0 \rangle^{NSCS}$ - то же самое соответственно для центро-симметричных и нецентросимметричных отражений

$M/A = \Omega_{MAX} / \Omega_{AVER}$, где Ω_{MAX} и Ω_{AVER} - максимальное и среднее значения функции распределения электронной плотности.

Таблица 2. Сравнение методов улучшения наборов фаз для макромолекул.

#	метод	авторы работы, год, ссылка	структура	разрешение, Å (количества отражений)	исходные данные	средства контроля	результат	примечание
1	Макси- мум опред.	де Ранго и сотр., 1975, [102]	инсулин	3.8 → 1.9 (1800 доп. отражений)	φ_M	$\langle \varphi - \varphi_M \rangle_+$	15°	порядок матриц 400
2	Макси- мум опред.	Подъярни и сотр., 1976, [104]	триклини- ческая моди- фикация лизоцима	3.3 → 2.5 (1400 → 2000)	F_M, φ_M F_O, φ_M	$\langle \varphi - \varphi_M \rangle_+$ $\langle \varphi - \varphi_M \rangle_+$	65° 75°	порядок матриц 336
3	Макси- мум опред.	Подъярни и сотр., 1981, [108]	дрожжевая тРНК	$[32-14] \rightarrow$ $\rightarrow [100-14]$ (107 → 135)	φ_{MIR}	$\langle \varphi - \varphi_M \rangle_+=$	73°	порядок матриц 4

Таблица 2. (Продолжение).

4	танг.-формула	Венцирль и сотр., 1969. [34]	модель	$4.0 \rightarrow 4.0$	$\left\{ \begin{array}{l} \langle \varphi_M^N - \varphi_M^C \rangle_{\text{НЧС}} = 56^\circ \\ \langle \varphi_M^N - \varphi_M^C \rangle_{\text{ЧС}} = 4^\circ \end{array} \right.$	$\langle \varphi - \varphi_M^N \rangle_{\text{НЧС}} = 28^\circ$	после 3-го цикла ошибки начали возрастать
5	танг.-формула	Рике и сотр., 1969, [35]	карбокси-пептидаза	$4.0 \rightarrow 4.0$ (700 \rightarrow 700)	$\left\{ \begin{array}{l} \varphi_M^{N2} = 45^\circ \\ \langle \varphi_M^{N2} - \varphi_M \rangle = 45^\circ \end{array} \right.$	$\langle \varphi - \varphi_M^{N2} \rangle = 170^\circ$	
				$6.0 \rightarrow 4.5$ (306 \rightarrow 894)	φ_{MIR}	$\langle \varphi - \varphi_{MIR} \rangle_+$	63°
				$4.7 \rightarrow 3.5$ (621 \rightarrow 2085)	φ_{MIR}^T	$\langle \varphi - \varphi_{MIR} \rangle_+$	77°
				$3.5 \rightarrow 2.8$ (2085 \rightarrow 2485)	φ_{MIR}	$\langle \varphi - \varphi_{MIR} \rangle_+$	70°
				$2.8 \rightarrow 2.0$	φ_{MIR}	$\langle \varphi - \varphi_{MIR} \rangle$	74°

Таблица 2. (Продолжение).

				φ_{MIR}	$\langle \varphi - \varphi_{MIR} \rangle$	63^0
6	танг.- форму- ла	Коултер, Дьюар, 1971, [37]	многолобин	$2.0 \rightarrow 1.4$ (1546-3191)	φ_{MIR}	
7	танг.- форму- ла	Хендрик- сон, 1975, [44]	кальций- связыв. белок	$2.0 \rightarrow 1.85$ (5055-8783)	$\begin{cases} \varphi_{MIR} \\ \langle \varphi_{MIR} - \varphi_M \rangle = 59^0 \end{cases}$ $\begin{cases} \langle \varphi_{MIR} - \varphi_M \rangle_{600} = 50^0 \\ \langle \varphi_{MIR} - \varphi_M \rangle_{600} \end{cases}$	$\langle \varphi_{TMR} - \varphi_M \rangle$ $\langle \varphi_{TMR} - \varphi_M \rangle_{600}$ $\langle \varphi_{comb} - \varphi_M \rangle_{600}$ $\langle \varphi_{comb} - \varphi_M \rangle_{600}$
8	ур-ния Сэйра, МНК	Сэйр, 1974, [24]	рубре- доксин	$2.5 \rightarrow 1.5$ (1608-5033) $2.0 \rightarrow 1.5$ (2813-5033)	φ_{MIR} сравнение карт	$\langle \varphi - \varphi_M \rangle =$ 47° сравнение карт

Таблица 2. (Продолжение).

9	ур-ния Сэйра, МНК	Катышид и сотр., 1975, [56]	Инсулин	$I.9 \rightarrow I.5$ (6300 \rightarrow $\rightarrow 10000$)	$\left\{ \langle \varphi_{MIR} - \varphi_M \rangle_0 = 60^\circ \right.$	$\langle \varphi - \varphi_M \rangle_0$ $\langle \varphi - \varphi_M \rangle_4$	52° 55°	см. данные: Агарвал, Ай- зекс, 1977, [6]
10	ур-ния Сэйра, МНК	Гассман, 1976, [49]	многлобин	$2.0 \rightarrow I.4$ (10000 \rightarrow $\rightarrow 17000$)	φ_M	$\langle \varphi - \varphi_M \rangle$	45°	
II	модиф. электр. плотн.	Коллинз и сотр., 1976, [80]	рубре- доксин	$2.0 \rightarrow I.5$ (28I2 \rightarrow 5033) $2.5 \rightarrow I.5$ (16I4 \rightarrow 5033)	$\left\{ \langle \varphi_{MIR} - \varphi_{MIR}^{LS} \rangle = 44^\circ \right.$ φ_{MIR}	$\langle \varphi - \varphi_{MIR}^{LS} \rangle_0$ $\langle \varphi - \varphi_{MIR}^{LS} \rangle_4$ $\langle \varphi - \varphi_{MIR}^{LS} \rangle$	37° 34°	близко к результатам предыдущего эксперимента
						$\langle \varphi - \varphi_{MIR}^{LS} \rangle$	$\langle \varphi - \varphi_{MIR}^{LS} \rangle$	улучшения нет улучшение есть

Таблица 2. (Продолжение).

I2	модиф. электр. плотн.	Каннилю и сотр., 1983, [81]	многолобин	$4.0 \rightarrow 4.0$ (II84-II84)	$\left\{ \begin{array}{l} \langle \varphi_M^N - \varphi_M \rangle = 30^\circ \\ \langle \varphi_M^N - \varphi_M \rangle_t = 25^\circ \end{array} \right.$	$\langle \varphi_M^N - \varphi_M \rangle$ $\langle \varphi_M^N - \varphi_M \rangle_t$	25°
				$4.0 \rightarrow 1.8$ (II84-II658)	φ_M^N	$\langle \varphi_M^N - \varphi_M \rangle$ $\langle \varphi_M^N - \varphi_M \rangle_t$	39°
				$4.0 \rightarrow 1.8$ (II84-II658)	φ_M^N	$\langle \varphi_M^N - \varphi_M \rangle$ $\langle \varphi_M^N - \varphi_M \rangle_t$	22°
						$\langle \varphi_M^N - \varphi_M \rangle$ $\langle \varphi_M^N - \varphi_M \rangle_t$	50°
I3	усред- нение по лок. систм.	Арго и сотр., 1975, [132]	глицер- альдегид- 3-фосфат дегидро- геназа	$6.0 \rightarrow 4.9$ (4284-6916) $\rightarrow 6.3$	φ_{SIR}	$\langle \varphi_S - \varphi_{DIR} \rangle$ $\langle \varphi_S - \varphi_{DIR} \rangle_t$	77° 8 циклов
					φ_{CM}	$\langle \varphi_S - \varphi_{DIR} \rangle$ $\langle \varphi_S - \varphi_{DIR} \rangle_t$	77° 10 циклов
							67°

Таблица 2. (Продолжение).

14	усреднение по зоне	Шевиц и сотр., 1981, [84]	Дрожжевая ТРНК	[14.0-4.5] → [14.0-4.5] (2633 → 2633)	$\left\{ \begin{array}{l} \varphi_{MIR} \\ \langle \varphi_M - \varphi_M \rangle = 68^\circ \end{array} \right.$	$\langle \varphi - \varphi_M \rangle_0$	49°	7 циклов
				[14.0-4.5] → [100.0-4.5] (2633 → 2662)	φ_{MIR}	$\langle \varphi - \varphi_M \rangle_0$	43°	II циклов
				[14.0-4.5] → [100.0-4.0] (2633 → 3545)	φ_{MIR}	$\langle \varphi - \varphi_M \rangle_0$	22°	7 циклов

сумм. и модиф.
электр. плотн.

Таблица 2. (Продолжение).

I5	усреднение по лок. симм.	Рэймент и сотр., 1983, [157]	модель капсиды вируса полиомы	φ_m $30.0 \rightarrow 30.0$ (90% → 100%)	$\langle \varphi - \varphi_m \rangle_t$	15^0	4 цикла
				φ_{cm1} $\langle \varphi_m - \varphi_m \rangle = 72^0$	$\langle \varphi - \varphi_m \rangle_t$	27^0	15 циклов
				φ_{cm2} $\langle \varphi_m - \varphi_m \rangle = 93^0$	$\langle \varphi - \varphi_m \rangle_t$	51^0	15 циклов
				φ_{cm1}	$\langle \varphi - \varphi_m \rangle_t$	28^0	21 цикл
				φ_{cm1} $\langle \varphi_m - \varphi_m \rangle = 90^0$	$\langle \varphi - \varphi_m \rangle_t$	48^0	15 циклов
				φ_{cm1} $\langle \varphi_m - \varphi_m \rangle = 32^0$	$\langle \varphi - \varphi_m \rangle_t$	17^0	15 циклов
				φ_{cm1} $\langle \varphi_m - \varphi_m \rangle = 35^0$	$\langle \varphi - \varphi_m \rangle_t$	35^0	15 циклов

Таблица 2. (Продолжение).

I6	груб. атомн. модель	Агаравал, Айзекс, 1977, [6]	инсулит	3.0 → 2.0 (→ 6300)	φ_{MIR}	M/A = 6	$\langle \varphi - \varphi_M \rangle =$ $\varphi_{\text{MIR}}^{\text{REF}}$	II 8 циклов уточнения	65°
				2.0 → I.7	M/A			III 8 циклов уточнения	30
I7	груб. атомн. модель	Лунин, Уржумцев, 1983, [202]	актинидия	I.7 → I.5 (→ I3400)	$\varphi_{\text{MIR}}^{\text{REF}}$	M/A	$\langle \varphi - \varphi_M \rangle$	III для синтеза 1.5 Å по модели	29
									70°
I8	груб. атомн. модель	Лунин, Уржумцев, 1983, [202]	актинидия	3.0 → 2.0 (464I → → I5009)	φ_M		$\langle \varphi - \varphi_M \rangle =$ $\langle \varphi - \varphi_M \rangle_+$	III 2I°	31°
				3.0 → 3.0 (464I → 464I)	φ_M^N		$\langle \varphi - \varphi_M \rangle =$ $\langle \varphi - \varphi_M \rangle = 370$		35°

Анализ таблицы 2 приводит к следующим выводам.

- I) Методы максимума определителя, улучшения по тангенс-формуле, решения системы уравнений Сэйра, основывающиеся на гипотезах разделенности изображений атомов и неотрицательности функции распределения электронной плотности, эффективны лишь при достаточно высоком разрешении. При снижении разрешения возможен отказ метода. Ошибки определения фаз заметно нарастают со снижением "силы" отражений.
- 2) Метод модификации функции распределения электронной плотности может эффективно применяться даже с достаточно простой функцией модификации.
- 3) Метод усреднения функции распределения электронной плотности по локальной симметрии позволяет эффективно работать на данных низкого разрешения. Большое количество успешных практических работ свидетельствуют о пригодности метода и при среднем разрешении. Существенным ограничением является необходимость достаточно аккуратного проведения границ молекулы. Необходимо также наличие явления локальной симметрии в кристаллах исследуемой структуры.
- 4) Методы, использующие моделирование рассеяния рентгеновских лучей на тех либо иных объектах, могут быть успешно использованы при работе с данными среднего и низкого разрешения. Эти методы не требуют дополнительных ограничений на структуру. Может быть использована дополнительная информация различных типов, что не всегда легко реализуется в других известных методах.
- 5) Низкое качество исходного набора фаз может существенно осложнить работу метода. При достаточно больших ошибках в

исходных значениях фаз улучшения этих значений может не произойти.

- 6) Существенной является применяемая методика (ср. данные таблицы 2, эксперименты II и I2, I6 и I7), а также выбор параметров в рамках используемой методики.
- 7) Развитые к настоящему моменту методы не являются универсальными, и требуется разработка новых методов, позволяющих использовать одновременно всю имеющуюся дополнительную информацию.
- 8) Наибольшее развитие в последние годы получают методы, использующие модели того либо иного типа для расчета по ним новых значений фаз.

ГЛАВА 2

УЛУЧШЕНИЕ НАБОРОВ ФАЗ СТРУКТУРНЫХ ФАКТОРОВ ПУТЕМ ПОСТРОЕНИЯ И МОДИФИКАЦИИ ГРУБОЙ ЛИБО СМЕШАННОЙ АТОМНОЙ МОДЕЛИ

2.1. Общее описание метода.

2.1.1. Постановка задачи.

При рентгеноструктурном исследовании биологических макромолекул этапу построения атомной модели структуры предшествует этап определения набора фаз структурных факторов и построения соответствующих карт распределения электронной плотности. При этом довольно часто, несмотря на то, что полный набор модулей структурных факторов для нативного белка имеет относительно высокое разрешение ($2.5\text{--}2.0 \text{ \AA}$), на эта-