

РОССИЙСКАЯ АКАДЕМИЯ НАУК

ОРДENA ТРУДОВОГО КРАСНОГО ЗНАМЕНИ
ИНСТИТУТ КРИСТАЛЛОГРАФИИ им. А.В.ШУБНИКОВА

На правах рукописи

ЛУНИН
ВЛАДИМИР ЮРЬЕВИЧ

УДК 548.737

ИСПОЛЬЗОВАНИЕ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК СИНТЕЗОВ ФУРЬЕ
ЭЛЕКТРОННОЙ ПЛОТНОСТИ ДЛЯ РЕШЕНИЯ ФАЗОВОЙ ПРОБЛЕМЫ В
КРИСТАЛЛОГРАФИИ БЕЛКА

Специальность 01.04.18 – Кристаллография,
физика кристаллов

Автореферат
диссертации на соискание ученой степени
доктора физико-математических наук

Москва 1992

Работа выполнена в Научно-исследовательском
вычислительном центре АН СССР (Пущино)

Официальные оппоненты :

Доктор физико-математических наук, профессор Н.С.Андреева
Доктор физико-математических наук В.Р.Мелик-Адамян
Доктор физико-математических наук Н.Н.Козлов

Ведущая организация : Институт теоретической и экспериментальной
биофизики РАН

Зашита диссертации состоится "... *август* 1992 г.
в 10 часов 30 минут на заседании Специализированного
совета Д 002.58.01 при Институте кристаллографии РАН
по адресу : 117333 Москва, Ленинский проспект, 59.

С диссертацией можно ознакомиться в библиотеке
Института кристаллографии РАН

Автореферат разослан "..." 199 ... г.

Ученый секретарь Специализированного совета
кандидат физико-математических наук

В.М.Каневский

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования.

Метод рентгеноструктурного анализа (РСА) постоянно привлекает внимание широкого класса исследователей, работающих в областях науки, использующих информацию о структуре вещества на атомно-молекулярном уровне. Это связано с тем, что метод РСА является практически единственным методом, позволяющим определять структуру исследуемого объекта на атомном уровне, то есть определять координаты в трехмерном пространстве всех атомов, входящих в состав исследуемого вещества.

Принципиальной особенностью метода РСА является неполнота данных, получаемых непосредственно в рентгеновском эксперименте. Эксперимент позволяет измерить лишь величины модулей F_s комплексных коэффициентов (структурных факторов) в разложении функции распределения электронной плотности в ряд Фурье :

$$\rho(r) = \frac{1}{V_{cell}} \sum_s F_s e^{i\varphi_s} e^{-2\pi i(s, r)} \quad (I)$$

Определение фаз φ_s структурных факторов представляет собой центральную проблему ("фазовую проблему") рентгеноструктурного анализа, и успех всей работы по определению структуры определяется во многом тем, насколько точно удалось решить фазовую проблему.

В силу неполноты данных рентгеновского эксперимента для решения фазовой проблемы необходима какая-то дополнительная информация об исследуемом объекте. В настоящее время на практике применяются, в основном, два класса методов решения фазовой проблемы. Для определения структуры низкомолекулярных соединений (100-150 атомов) применяются "прямые методы определения фаз", дополнительной информацией для которых является "атомность" исследуемого объекта. Основой для решения фазовой проблемы при исследовании биологических макромолекул является метод изоморфного замещения, в котором в качестве добавочной информации об исследуемом объекте (нативном белке) выступают данные по дополнительным рентгеновским экспериментам с близкими соединениями (изоморфными производными), отличающимися от нативного белка локальными добавками ("тяжелыми" атомами), не исказящими нативную структуру. В этом методе тяжесть решения фазовой проблемы смещается в биохимическую область. Получение изоморфных тяжелоатомных производных является сложной задачей, которую далеко не всегда удается решить. Даже в тех случаях,

когда производные вещества удается получить, изоморфизм может иметь место лишь приближенно – присоединяемые тяжелые атомы могут вносить некоторые искажения в нативную структуру. Это приводит (в зависимости от степени нарушения изоморфизма) к более или менее грубым ошибкам при расчете фаз и, как следствие, к усложнению задачи интерпретации в структурных терминах получаемой функции распределения электронной плотности (вплоть до полной неинтерпретируемости).

Указанные сложности применения метода изоморфного замещения стимулировали в последние годы значительные усилия по поиску дополнительных источников информации о структуре макромолекул и методов использования такой дополнительной информации для решения двух задач:

1) повышения интерпретируемости синтезов (уточнение значений фаз, определенных с ошибкой и, возможно, доопределение значений некоторых фаз, не определенных ранее);

2) решения фазовой проблемы для макромолекул при отсутствии тяжелоатомных производных.

Диссертационная работа выполнена по плану научно-исследовательских работ Научно-исследовательского вычислительного центра АН СССР (комплексная научно-техническая программа 0.74.05, номер государственной регистрации 0187.0 067038).

Цель работы заключалась в разработке и проверке применимости к решению фазовой проблемы новых подходов, основанных на вовлечении в работу не использовавшихся ранее характеристик распределения электронной плотности в кристаллах белков – гистограмм синтезов Фурье электронной плотности конечного разрешения.

Научная новизна работы.

Автором диссертации впервые (в 1984 г.) предложено использовать в качестве дополнительного источника информации о белках распределения частот, с которыми встречаются различные значения на синтезах электронной плотности конечного разрешения. (Несколько позже близкие идеи были высказаны независимо рядом зарубежных исследователей([1,2,3])). Впервые предложены подходы к практическому использованию этой информации в процессе решения фазовой проблемы и найдены пути к получению требуемой информации для исследуемого белка. Практическая реализация предложенных методов потребовала решения широкого круга разнообразных проблем, как теоретического, так и прикладного характера.

Научно-практическая значимость работы.

Гистограммы синтезов электронной плотности конечного разрешения дают новую информацию о кристаллах белков, практически не использовавшуюся ранее. Это позволяет рассчитывать на возможность продвижения в работе по определению структуры биологических макромолекул в ряде ситуаций, когда "классических" методов становится недостаточно. Разработанные и реализованные алгоритмы практического использования этой информации позволили провести тестирование ее полезности и применить новые методы к исследованию реальных структур (γ -кристаллина III β и фактора элонгации G). Теоретический анализ методов использования гистограммной информации привел к более четкому пониманию "механизма работы" и границ применимости одного из старейших методов улучшения карт распределения электронной плотности — метода модификации электронной плотности.

Апробация работы.

Результаты исследований, изложенные в диссертации, докладывались и обсуждались на следующих конференциях:

Четвертый всесоюзный симпозиум "Структура биологических макромолекул". (Звенигород, 1984 г.).

Международная школа-симпозиум по структуре биологических макромолекул. (Пущино, 1986 г.).

Международная школа "Кристаллография в молекулярной биологии". (Эриче, Италия, 1988).

Первая международная конференция "Структура и функция рибонуклеаз". (Москва, 1988).

Двенадцатая Европейская кристаллографическая конференция. (Москва, 1989).

Пятнадцатый международный конгресс по кристаллографии. (Бордо, Франция, 1990).

Школа по кристаллографическим вычислениям. (Стратсбург, Франция, 1990).

Структура и объем диссертации.

Диссертация состоит из введения, шести глав, заключения, основных выводов, и списка цитируемой литературы. Диссертация содержит 191 страницу, включая 14 таблиц и 28 рисунков. Список цитируемой литературы содержит 126 наименований.

Публикации.

По теме диссертации опубликовано 28 печатных работ, список основных публикаций приведен в конце авторефера.

СОДЕРЖАНИЕ РАБОТЫ.

Во введении приводятся краткие сведения из рентгеновской кристаллографии, поясняющие постановку рассматриваемых далее проблем, излагается история вопроса и дается краткая аннотация результатов, полученных в работе.

Глава I. ГИСТОГРАММА СИНТЕЗА ЭЛЕКТРОННОЙ ПЛОТНОСТИ КОНЕЧНОГО РАЗРЕШЕНИЯ - НОВЫЙ ИСТОЧНИК ИНФОРМАЦИИ О КРИСТАЛЛАХ БЕЛКОВ.

В этой главе вводится в рассмотрение новый тип дополнительной информации о распределении электронной плотности в белках - гистограммы синтезов Фурье электронной плотности конечного разрешения - и обсуждаются основные свойства таких гистограмм.

Гистограмма, отвечающая функции распределения электронной плотности.

Основой предлагаемых в диссертации подходов является попытка установить какие значения может принимать искомая функция $\rho(r)$ и как часто принимает функция $\rho(r)$ каждое из возможных значений. Наиболее прямой практический подход к ответу на этот вопрос может быть следующим. Введем в элементарной ячейке V равномерную сетку, и пусть $\{\rho_j\}_{j=1}^N$ - совокупность значений функции $\rho(r)$, вычисленных в узлах этой сетки. Разобьем интервал $(\rho_{\min}, \rho_{\max})$, в котором лежат значения $\{\rho_j\}$, на заданное число K равных частей (бинов) и для каждого бина определим частоту попадания значений ρ_j в этот бин

$$\hat{v}_k = n_k / N , \quad k = 1, \dots, K .$$

Здесь n_k - число точек сетки со значениями ρ_j , лежащими в k -ом бине, то есть таких точек, что

$$\frac{\rho_{\max} - \rho_{\min}}{K} \leq \rho_j \leq \rho_{\min} + j \frac{\rho_{\max} - \rho_{\min}}{K} ,$$

N - общее число точек сетки. Совокупность (распределение) частот $\{\hat{v}_k\}_{k=1}^K$ мы будем называть гистограммой, отвечающей функции $\rho(r)$.

Иногда более удобна для работы нормированная гистограмма

$$v_k = n_k / (\Delta_k N) , \quad k = 1, \dots, K , \quad (2)$$

где Δ_k обозначает длину k -того бина. В этом случае вероятность обнаружения точки (при случайному выборе узла сетки в области V) со значением ρ_j , лежащим в k -ом бине, равна $v_k \Delta_k$.

Полученные расчетом по формуле (2) значения частот зависят, строго говоря, не только от исследуемой функции $\rho(r)$, но и от

того, как введены бины. Чтобы избавиться от этой зависимости, мы можем ввести меру на области значений исследуемой функции более аккуратно. Определим для функции $\rho(r)$ кумулятивную функцию

$$N(t) = \frac{1}{|V|} \text{mes} \{ r : \rho(r) \leq t \}$$

и плотность кумулятивной функции

$$\nu(t) = \frac{d}{dt} N(t) = \frac{1}{|V|} \frac{d}{dt} \text{mes} \{ r : \rho(r) \leq t \}.$$

(Здесь и далее $\{r: A\}$ обозначает множество точек из области V , для которых выполнено условие A , $\text{mes } S$ обозначает объем области S , $|V| = \text{mes } V$ – объем области V). Функции $N(t)$ и $\nu(t)$ зависят только от исследуемой функции $\rho(r)$ и не связаны с выбором сетки в исследуемой области и разбиения вещественной оси на бины. Величина $\nu(t)\Delta t$ при малых Δt представляет собой вероятность встретить (при случайному выборе точки из области V) значение $\rho(r)$, лежащее в интервале $(\rho, \rho + \Delta t)$.

Нетрудно видеть, что нормированные частоты ν_k , вычисляемые по формуле (2), являются приближенными значениями функции $\nu(t)$ в точках t_k , отвечающих серединам бинов:

$$\nu(t_k) = \lim_{\Delta_k \rightarrow 0, N \rightarrow \infty} \nu_k.$$

Мы будем далее использовать термин "гистограмма" как для обозначения наборов частот $\{\nu_k\}_{k=1}^K$ и $\{\hat{\nu}_k\}_{k=1}^K$, так и для обозначения плотности $\nu(t)$ кумулятивной функции. Использование функции $\nu(t)$ более удобно при рассмотрении теоретических вопросов, в то время, как на практике удобнее оперировать с частотами.

Гистограммы синтезов Фурье конечного разрешения.

На рис.1 показан типичный вид гистограммы, отвечающей синтезу электронной плотности среднего разрешения (в данном случае 4 Å) для белка. Эта гистограмма имеет характерную асимметричную форму, которая присуща и гистограммам, отвечающим синтезам электронной плотности для других белков.

Наиболее важным свойством, определяющим дальнейшее использование гистограмм, является то, что гистограмма оказывается чувствительной к наличию ошибок в фазах структурных факторов и к отсутствию части структурных факторов при расчете синтеза. На рис.2 показано как влияет на форму гистограммы замена точных значений фаз структурных факторов на случайные значения и как влияет исключение из синтеза около 18% рефлексов вблизи оси 1 обратного пространства. Эта чувствительность гистограммы к

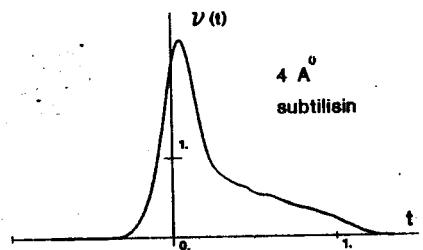


Рис.1. Гистограмма синтеза электронной плотности разрешения 4Å для белка субтилазина.

ошибкам позволяет надеяться на то, что гистограмма синтеза электронной плотности может служить индикатором правильности определения фаз структурных факторов.

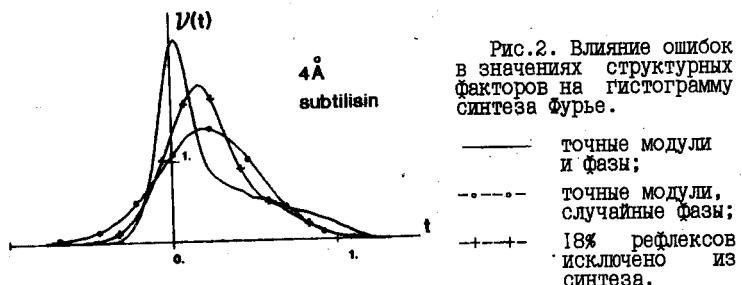


Рис.2. Влияние ошибок в значениях структурных факторов на гистограмму синтеза Фурье.

- точные модули и фазы;
- - - точные модули, случайные фазы;
- +---+ 18% рефлексов исключено из синтеза.

На практике мы всегда имеем дело не с "идеальным" распределением электронной плотности $\rho(r)$, а синтезом некоторого конечного разрешения d_{\min} (мы называем сумму (I) синтезом разрешения d_{\min} , если в него включены все слагаемые, отвечающие узлам в решетке обратного пространства $|s| \leq I/d_{\min}$). Рис.3 показывает как изменение разрешения синтеза Фурье влияет на форму соответствующей гистограммы.

Более подробный анализ обнаруживает, также, чувствительность гистограмм к среднему значению температурного фактора атомов.

Глава 2. ПРЕДСКАЗАНИЕ ГИСТОГРАММ ДЛЯ БЕЛКОВ С НЕИЗВЕСТНОЙ ПРОСТРАНСТВЕННОЙ СТРУКТУРОЙ.

В этой главе показывается, что гистограмма, отвечающая точному синтезу электронной плотности для кристалла исследуемого белка (то есть синтезу Фурье, рассчитанному с правильными значениями модулей и фаз структурных факторов) может быть

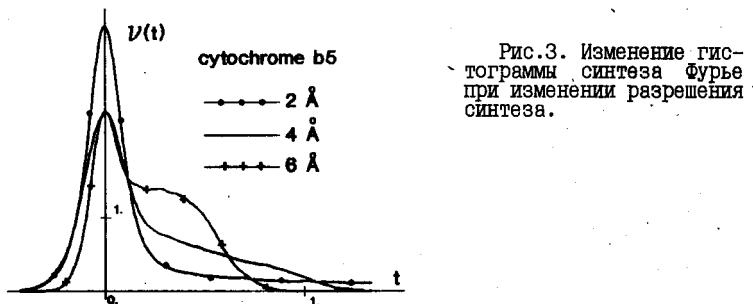


Рис.3. Изменение гистограммы синтеза Фурье при изменении разрешения синтеза.

получена до того, как решена фазовая проблема для данного белка. Тем самым, такая гистограмма может выступать в процессе решения фазовой проблемы источником дополнительной информации об исследуемом объекте, ограничивая класс возможных решений фазовой проблемы: только такие наборы фаз являются приемлемыми, которые приводят к синтезу, обладающему предписанной гистограммой.

Эмпирическая модель для описания гистограмм.

Анализ гистограмм, отвечающих синтезам одного и того же разрешения для разных белков показывает (рис.4), что эти

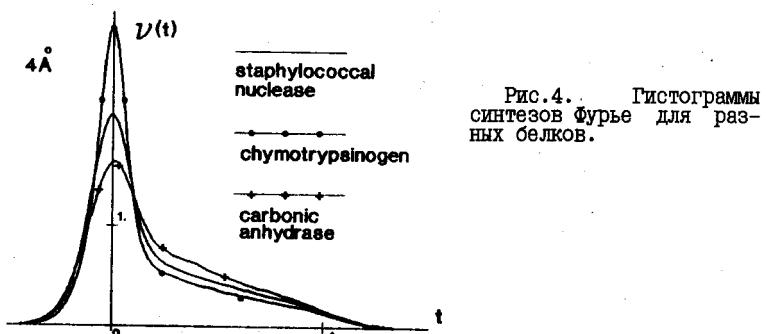


Рис.4. Гистограммы синтезов Фурье для разных белков.

гистограммы, хотя и имеют сходную форму, но не совпадают друг с другом, то есть гистограмма, отвечающая какому-то белку с известной структурой, не может быть непосредственно использована как эталон гистограммы для другого белка. Однако, ситуация несколько улучшается, если применить к гистограммам дополнительную перенормировку, переходя к "нормированным объемам"

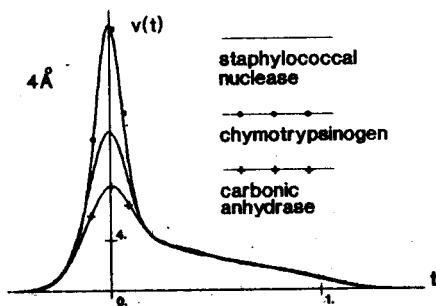


Рис.5. Гистограммы синтезов Фурье для тех же белков, что и на рис.4, после перенормировки.

$$v_k = v_k |V| / F_{ooo}, \quad k=1, \dots, K,$$

$$v(t) = v(t) |V| / F_{ooo}.$$

Здесь $|V|$ – объем, а F_{ooo} – суммарный электронный заряд элементарной ячейки. (При малых Δt , $v(t)\Delta t$ – это приходящийся на один электрон объем области в элементарной ячейке, в которой значения $r(r)$ лежат в интервале $(t, t+\Delta t)$). Как видно из рис.5, нормализованные гистограммы имеют похожие участки, соответствующие средним и большим значениям r , а также, наименьшим значениям r . (Подчеркнем, что поскольку мы исследуем синтезы конечного разрешения, на этих синтезах обязательно будут области с отрицательными значениями $r(r)$ и, более того, эти области будут сосредоточены в основном в области молекулы).

Рис.5 позволяет высказать гипотезу, что "нормированный объем" $v(r)$ для значений r , встречающихся только в области молекулы, одинаков для всех белков и может быть описан стандартным (одинаковым для всех белков) распределением $v^0(t)$. При этом, конечно, распределение $v^0(t)$ меняется при изменении разрешения синтеза. Мы не можем определить непосредственно из графиков, приведенных на Рис.5, величины $v^0(t)$ для близких к нулю значений t . Такие значения встречаются не только в области молекулы, но и в "волнах обрыва ряда Фурье", и в кривой $v(t)$ содержится информация о "смеси" таких значений. Однако более аккуратный анализ кривых $v(t)$ для разных белков позволяет определить значения стандартного распределения $v^0(t)$ для всех значений t .

Анализ функций $v(t)$, отвечающих кристаллам разных белков,

позволяет заметить, что высота центрального пика на соответствующих графиках оказывается прямо пропорциональной величине $|V|/F_{ooo}$, то есть чем "свободней" размещаются молекулы белка в кристаллической ячейке, тем больше высота этого пика. Это позволяет выдвинуть гипотезу, что объем области в межмолекулярном пространстве элементарной ячейки, в которой значения функции $\rho(r)$ попадают в некоторый интервал $(t, t+\Delta t)$, прямо пропорционален объему межмолекулярного пространства. Это дает эмпирическую модель для распределения нормированных объемов

$$v(t) = v^0(t) + \left(|V|/F_{ooo} - \int_{-\infty}^{\infty} v^0(x) dx \right) q^0(t), \quad (3)$$

или, что же, к модель для гистограммы

$$v(t) = \frac{F_{ooo}}{|V|} v^0(t) + \left(1 - \frac{F_{ooo}}{|V|} \int_{-\infty}^{\infty} v^0(x) dx \right) q^0(t). \quad (4)$$

В этих формулах $v^0(t)$ – одинаковая для всех белков функция, описывающая распределение значений $\rho(r)$ внутри области молекул, а $q^0(t)$ – одинаковая для всех белков функция, описывающая распределение значений $\rho(r)$ в межмолекулярной области. (Фактически, $q^0(t)$ описывает распределение значений для "волны обрыва ряда Фурье".)

Дискретным аналогом формул (3)–(4) являются равенства

$$v_k = v_k^0 + \left(|V|/F_{ooo} - \sum_{j=1}^K v_j^0 \Delta_j \right) q_k^0 \quad (5)$$

$$v_k = \frac{F_{ooo}}{|V|} v_k^0 + \left(1 - \frac{F_{ooo}}{|V|} \sum_{j=1}^K v_j^0 \Delta_j \right) q_k^0. \quad (6)$$

Здесь зависящие от непрерывно меняющегося параметра t стандартные функции $v^0(t)$ и $q^0(t)$ заменены на наборы значений этих функций $v_k^0=v^0(t_k)$ и $q_k^0=q^0(t_k)$ в отдельных точках t_k .

Расчет стандартных распределений $\{v_k^0\}_{k=1}^K$ и $\{q_k^0\}_{k=1}^K$.

Для определения содержащихся в формулах (5)–(6) стандартных распределений использовался набор белков ("базисный набор"), атомная структура которых известна (таблица I). Для отобранных белков по атомной модели были найдены структурные факторы, построены синтезы электронной плотности и рассчитаны точные гистограммы. Стандартные распределения $\{v_k^0\}_{k=1}^K$ и $\{q_k^0\}_{k=1}^K$ были определены из условия наилучшего совпадения "теоретических" гистограмм, определенных по формуле (6) и точных гистограмм базисных белков.

На рис.6 показаны графики стандартных распределений $v^0(t)$ и $q^0(t)$, отвечающих разрешению 4 Å.

Таблица I.

белок	точность предсказания гистограммы Q_h	
	$d = 4\text{ \AA}$	$d = 10\text{ \AA}$
белки базисного набора		
Carbonic anhydrase	0.022	0.098
Chymotrypsinogen	0.033	0.051
Cytochrome b5	0.053	0.106
HIPPIP	0.028	0.135
B-J Protein	0.023	0.074
Insulin	0.024	0.408
Lysozyme	0.025	0.064
Mioglobin	0.056	0.039
Neurotoxin	0.088	0.212
Ovomucoid	0.030	0.145
Phospholipase	0.019	0.091
Plastocyanin	0.032	0.090
Prealbumin	0.020	0.063
Proteinase A	0.023	0.130
Ribonuclease	0.038	0.162
Staphilococcal nuclease	0.044	0.066
белки, не входившие в базисный набор		
Ubiquitin	0.045	0.176
Gmabrin	0.044	0.076
Avian pancreatic polypeptide	0.051	0.128
Rubredoxine	0.076	0.170
Concanavalin	0.064	0.073

$$Q_h = \sum_{k=1}^K |v_k^o - v_k^c| \Delta_k, \quad v_k^o - \text{значения частот для точного синтеза}; \\ v_k^c - \text{значения частот, рассчитанные по формуле (6)}.$$

Предсказание гистограмм.

После того, как стандартные распределения $\{v_k^o\}_{k=1}^K$ и $\{q_k^o\}_{k=1}^K$ (отвечающие разрешению d_{min}) определены, формулы (4) или (6) позволяют предсказывать гистограмму (отвечающую разрешению d_{min}) для произвольного белка, если только параметры V и F_{ooo} для этого белка известны.

На рис.7 показаны точная и соответствующая найденным $\{v_k^o\}_{k=1}^K$ и $\{q_k^o\}_{k=1}^K$ "теоретическая" гистограммы для белка миоглобина (наихудшее совпадение среди базисных белков). На рис.8 показаны точная и предсказанная гистограммы для белка конканавалина, не входившего в базисный набор белков.

Предсказание гистограмм для синтезов низкого разрешения.

Изложенная методика предсказания гистограмм дает приемлемую для практической работы точность при работе с синтезами среднего

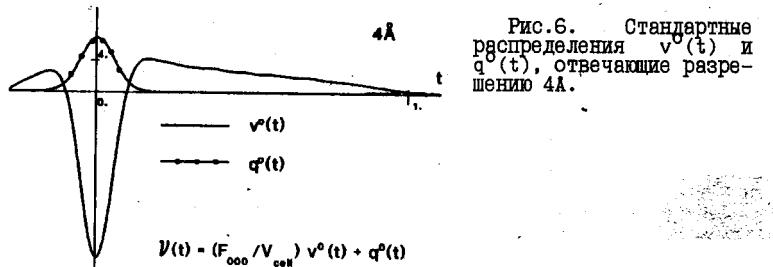


Рис.6. Стандартные распределения $v^0(t)$ и $q^0(t)$, отвечающие разрешению 4 Å.

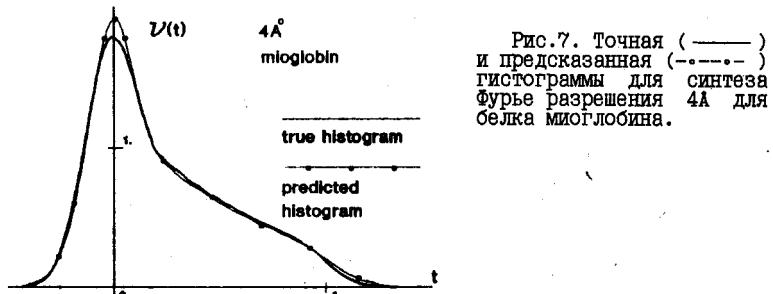


Рис.7. Точная (—) и предсказанная (- - -) гистограммы для синтеза Фурье разрешения 4 Å для белка миоглобина.

и высокого разрешения. Однако при переходе к синтезам низкого разрешения точность предсказания убывает. Это вынудило нас при работе с низким разрешением использовать для предсказания гистограмм иную методику, основанную на использовании атомной модели гомологичного белка или атомной модели, скомпонованной из кусков других белков, близкой по размерам и внешним очертаниям к модели исследуемого белка (такая информация может быть, например, известна из данных электронной микроскопии). В этом случае задача предсказания гистограммы может быть решена в два этапа:

- а) "размещение" атомной модели гомолога без самоналезаний в элементарной ячейке исследуемого белка;
- б) расчет гистограммы, отвечающей такой гипотетической модели (то есть расчет по атомной модели структурных факторов, затем расчет синтеза Фурье нужного разрешения и гистограммы для него).

Тесты показали, что рассчитываемые таким образом гистограммы слабо зависят от возможных изменений способа упаковки модели и могут быть использованы в практической работе.

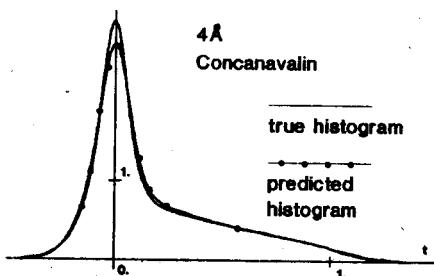


Рис.8. Точная (—) и предсказанная (- - -) гистограммы для синтеза Фурье разрешения 4А для белка конканавалина, не входившего в базисный набор белков.

Глава 3. ИСПОЛЬЗОВАНИЕ ГИСТОГРАММ ДЛЯ ВОССТАНОВЛЕНИЯ НЕДОСТАЮЩИХ СТРУКТУРНЫХ ФАКТОРОВ.

В этой главе показано, как зная гистограмму, которой должен обладать точный синтез электронной плотности, можно решать задачу определения значений фаз для части структурных факторов или, даже, восстанавливать значения некоторых структурных факторов полностью (и модули и фазы).

Постановка задачи восстановления недостающих структурных факторов.

Предположим, что перед нами стоит задача вычисления синтеза Фурье конечного разрешения

$$\rho(r) = \frac{1}{|V|} \sum_{|s| \leq s_{\max}} F(s) e^{i\phi(s)} e^{-2\pi i(s, r)}, \quad (7)$$

причем не все необходимые для расчета значения структурных факторов известны. Обозначим через S_d множество индексов s , отвечающих структурным факторам с известными модулем F^0 и фазой ϕ^0 , а S_u , отвечающих структурным факторам, для которых либо фаза, либо и модуль и фаза неизвестны. Для расчета синтеза (7) этим неизвестным структурным факторам нужно присвоить какие-то численные значения. Обычный путь — не включать эти рефлексы в синтез, то есть присвоить соответствующим структурным факторам нулевое значение — может приводить к заметным искажениям синтеза.

Мы можем попытаться сделать более обоснованный выбор значений для неизвестных структурных факторов, если у нас имеется какая-то дополнительная информация о том, какими свойствами должен обладать синтез $\rho(r)$, который мы хотим получить. В таком случае мы можем попытаться доопределить неизвестные величины так, чтобы получаемая функция в максимальной мере отвечала этим

дополнительным требованиям.

Допустим, что нам известна гистограмма $\{\nu_k^o\}_{k=1}^K$, которой должен обладать синтез Фурье $\rho(r)$ будучи рассчитан с правильными значениями всех структурных факторов (мы будем называть ее эталонной гистограммой). Тогда для каждого пробного набора неизвестных структурных факторов мы можем проверить, насколько хорошо он согласуется с этой гистограммой, проделав следующую цепочку вычислений:

а) введем в элементарной ячейке некоторую сетку и рассчитаем значения пробного синтеза в узлах этой сетки

$$\begin{aligned}\rho_j^c = \rho^c(r_j) &= \frac{1}{|\mathcal{V}|} \sum_{s \in S_d} F^c(s) e^{i\varphi^c(s)} e^{-2\pi i(s, r_j)} + \\ &+ \frac{1}{|\mathcal{V}|} \sum_{s \in S_u} F^c(s) e^{i\varphi^c(s)} e^{-2\pi i(s, r_j)};\end{aligned}\quad (8)$$

б) рассчитаем гистограмму $\{\nu_k^c\}_{k=1}^K$, отвечающую полученному синтезу;

в) сравним, насколько близки эталонная и рассчитанная гистограмма, например, воспользовавшись критерием близости гистограмм типа

$$Q(\rho^c) = \frac{1}{K} \sum_{k=1}^K \frac{(\nu_k^c - \nu_k^o)^2}{\nu_k^o}. \quad (9)$$

Естественно считать, что наилучшим образом согласуется с имеющейся гистограммой $\{\nu_k^o\}_{k=1}^K$ тот набор значений неизвестных структурных факторов, для которого величина (9) минимальна. То есть задача доопределения неизвестных значений структурных факторов может быть сформулирована как задача минимизации функции (9), в которой величины частот ν_k^c зависят от значений r_j^c пробного синтеза, а те в свою очередь определяются значениями F_s^c и φ_s^c искомых структурных факторов.

Разумеется, любые другие типы дополнительной информации об исследуемом объекте (наличие некристаллографической симметрии, информация об области, занятой растворителем и т.п.) могут тоже быть подключены в работу. Обычный подход в таких случаях – минимизация составного критерия, где каждое из слагаемых "отвечает" за выполнение одного из дополнительных условий.

Квазигистограммы.

Минимизация критерия (9) представляет собой сложную вычислительную задачу, осложняющуюся еще и тем, что к функции (9) неприменимы методы, использующие информацию о производных (методы

минимизации первого порядка). Дело в том, что, как правило, при "малых изменениях" варьируемых параметров F_s и φ_s значения ρ_j^c немного изменяются, но остаются лежать в тех же бинах, что и раньше. Поэтому значения частот ν_k^c не меняются при малых изменениях пробных значений структурных факторов, и все производные критерия (9) по варьируемым параметрам равны нулю. Поэтому, для практической работы использовался несколько иной критерий качества пробного набора структурных факторов, более удобный для проведения процесса минимизации.

Вычисляемые согласно формуле (2) частоты $\{\nu_k^c\}_{k=1}^K$ могут быть определены формулой

$$\nu_k^c = \frac{1}{N} \sum_{j=1}^N \lambda^*(t_k - \rho_j) \quad (10)$$

где

$$\lambda^*(t) = \begin{cases} 1/\Delta & \text{при } |t| \leq \Delta/2, \\ 0 & \text{при } |t| > \Delta/2, \end{cases}$$

Δ — длина бинов, t_k — середины бинов.

"Плохие" свойства критерия (10) связаны, формально, с тем, что в вычислении величин ν_k^c участвует кусочно — постоянная функция $\lambda^*(t)$.

Введем

ОПРЕДЕЛЕНИЕ. Пусть $\lambda(t)$ — произвольная функция, для которой

$$\int_{-\infty}^{\infty} \lambda(\tau) d\tau = 1.$$

Будем называть квазичастотами (связанными с функцией $\lambda(t)$) величины, вычисляемые по формуле

$$\tilde{\nu}_k = \frac{1}{N} \sum_{j=1}^N \lambda(t_k - \rho_j),$$

Совокупность квазичастот $\{\tilde{\nu}_k\}_{k=1}^K$ будем называть квазигистограммой.

Если функция $\lambda(t)$ непрерывно дифференцируема (или, по крайней мере, кусочно-непрерывно дифференцируема), то квазичастоты гладким образом зависят от значений модулей и фаз структурных факторов, использованных при вычислении функции $\rho(r)$. Это позволяет использовать более удобный критерий качества пробного набора структурных факторов следующего вида.

Предположим, что нам известна эталонная квазигистограмма $\{\tilde{\nu}_k^o\}_{k=1}^K$, отвечающая искомой функции $\rho(r)$. Для пробного набора неизвестных структурных факторов $\{F^o(s)\exp(i\varphi(s))\}_{s \in S_u}$ определим критерий качества

$$Q = \frac{1}{K} \sum_{k=1}^K \frac{(\tilde{\nu}_k^o - \tilde{\nu}_k^e)^2}{\tilde{\nu}_k^o}. \quad (II)$$

Теперь задача определения неизвестных значений структурных факторов может быть сформулирована как задача минимизации значения критерия (II). Для минимизации этого критерия была написана специальная программа, реализующая алгоритм наискорейшего спуска с использованием алгоритмов быстрого преобразования Фурье и быстрого дифференцирования. При этом для вычисления квазичастот использовались кусочно – линейные функции вида

$$\lambda_\alpha(t) = \begin{cases} -(1/\alpha^2) |t| + 1/\alpha & \text{при } |t| \leq \alpha, \\ 0 & \text{при } |t| > \alpha. \end{cases} \quad (I2)$$

Основная идея перехода к квазичастотам – мы перестаем относить вклад отдельной точки сетки к одному бину и начинаем "распределять" его между несколькими соседними бинами. При этом, величины вкладов перераспределяются при изменениях значения ρ_j и делают тем самым квазичастоты чувствительными к малым изменениям ρ . Показано, что эталонные квазичастоты могут быть вычислены по эталонным частотам при помощи формулы

$$\tilde{\nu}_k = \frac{1}{N} \sum_{j=1}^N \lambda(t_k - \rho_j) \approx \int_{-\infty}^{\infty} \lambda(t-\tau) \nu(\tau) d\tau. \quad (I3)$$

Из этой формулы видно, также, что переход к квазичастотам предстает из себя некоторое "сглаживание" исходного распределения частот, и изменение критерия Q по сравнению с критерием (10) состоит в том, что мы переходим от сравнения гистограмм пробного и точного синтезов к сравнению некоторых усредненных характеристик гистограмм.

Тестовое восстановление пропущенных структурных факторов для субтилизина.

Для проверки работоспособности предложенного подхода была проведена серия тестов. В качестве тестового объекта была взята модель белка субтилизина [4], которая была размещена в пространственной ячейке с размерами $73 \times 64 \times 48$ Å в пространственной группе $P2_1 2_1 2_1$. По координатам атомов были рассчитаны структурные факторы и рассчитан синтез разрешения 4 Å (Рис.9^a). По этому синтезу были рассчитаны квазичастоты (отрезок (-0.5, 1.5) был разбит на 30 бинов, использовалась функция $\lambda(t)$ вида (I2) с $\alpha=5$).

Далее была проимитирована ситуация отсутствия информации о части модулей структурных факторов. Около 18% структурных факторов

(352 из 2104) были объявлены неизвестными и была поставлена задача определения их путем минимизации функции (II). В качестве набора S_u отсутствующих рефлексов выступал набор рефлексов, не отснятый по техническим причинам в одном из реальных рентгеновских экспериментов. Множество S_u было, в основном, сосредоточено около оси 1 обратного пространства.

Дальнейшие тесты проводились в двух модификациях. В первой из них предполагалось, что нам известны значения структурных факторов $F^o(s)\exp[i\phi^o(s)]$ для набора S_d рефлексов и известна эталонная квазигистограмма $\{\tilde{v}_k^{o,k}\}_{k=1}^K$. Задача заключалась в восстановлении и модулей, и фаз неизвестных структурных факторов. Во второй модификации мы считали, что для $s \in S_u$ неизвестны только фазы, а модули известны и задача состояла в определении неизвестных фаз.

Первый тест был посвящен попытке восстановить как фазы, так и модули структурных факторов с $s \in S_u$. В качестве стартовых для неизвестных структурных факторов были взяты нулевые значения. Одно из сечений стартового синтеза Фурье (то есть синтеза, построенного по неполному набору рефлексов) показано на Рис.9^b. В результате 10 циклов минимизации значение критерия (II) уменьшилось с 0.3×10^{-2} до 0.5×10^{-5} . При этом неизвестные ранее значения фаз определились со средней ошибкой в 37° , а значение R-фактора для восстановленных значений модулей структурных факторов составило 0.46. На Рис.9^c показан фрагмент синтеза, построенного с восстановленными значениями неизвестных структурных факторов. Мы видим здесь существенный прогресс по сравнению со стартовой картиной.

Во втором teste предполагалось, что для множества S_u модули структурных факторов известны, и проблема состоит лишь в восстановлении значений фаз. В качестве стартовых значений фаз при минимизации критерия (II) были взяты значения полученные в первом teste. Пять циклов минимизации привели к значению минимизируемого критерия 0.6×10^{-6} и средней ошибке определения фаз 33° . На рис.9^d показан фрагмент синтеза, при расчете которого использовались доопределенные таким образом значения фаз структурных факторов (значения модулей были взяты точными).

Восстановление величин структурных факторов для "сухой" формы γ -кристаллина III β .

Разработанная методика восстановления пропущенных структурных факторов была использована при исследовании "сухой" формы γ -кристаллина III β . Структура белка γ -кристаллина III β из глазной линзы теленка изучается в лаборатории д.б.н. Ю.Н.Чиргадзе

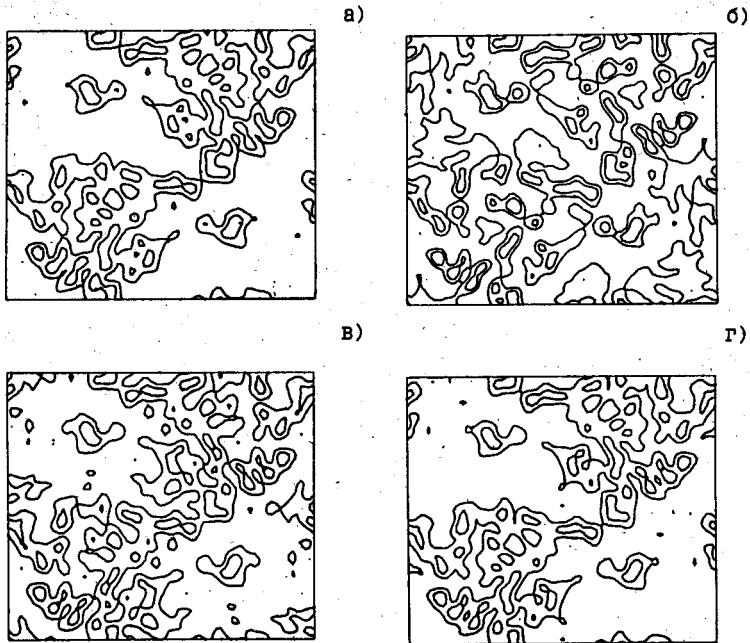


Рис.9. Сечение $z=12/48$ синтезов разрешения 4 Å для модели субтилизина : а) точный синтез; б) стартовый синтез (около 18% рефлексов исключено из синтеза; в) результат восстановления значений модулей и фаз удаленных рефлексов; г) результат восстановления фаз (при известных модулях) удаленных рефлексов.

в Институте белка АН СССР и в лаборатории проф. Т.Бланделла в Бирбек-колледже (Англия). Кристаллы белка принадлежат к пространственной группе $P2_12_12_1$ и имеют параметры элементарной ячейки $58.7 \times 69.5 \times 116.9$ Å. Структура γ -кристаллина была уточнена при разрешении 2.5 Å. Был также собран дифракционный набор разрешением до 1.9 Å "подсушанных" кристаллов белка [5]. Он характеризовался меньшими размерами ячейки : $57.38 \times 70.13 \times 115.4$ Å. Расхождение в данных для разных модификаций составило по зоне до 2.5 Å

$$R = 2 \frac{\sum_{\text{g}} |F_{\text{wet}} - F_{\text{dry}}|}{\sum_{\text{g}} |F_{\text{wet}} + F_{\text{dry}}|} = 0.255 .$$

По техническим причинам в первоначально отснятом наборе, отвечающем "сухой" форме, отсутствовала значительная часть

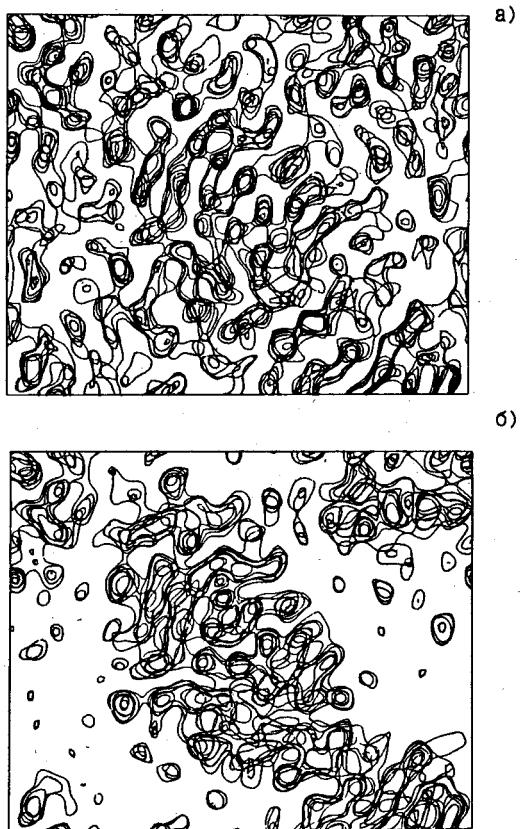


Рис.10. Синтез разрешения 4 \AA для γ -кристаллина III β :
а) синтез с коэффициентами (14); (часть рефлексов отсутствует);
б) результат включения в синтез восстановленных значений для
отсутствовавших значений структурных факторов.

рефлексов (в зоне до 4 \AA отсутствовало 1390 рефлексов из 4224 возможных). Была сделана попытка восстановить часть утерянных данных, используя информацию о гистограмме синтеза электронной плотности.

В качестве стартовой точки был выбран синтез разрешения до 4 \AA , построенный по 2852 рефлексам с коэффициентами

$$F_{dry}(s) \exp[i\varphi_{wet}(s)] , \quad (14)$$

где $F_{dry}(s)$ - модули структурных факторов второй ("сухой") модификации, $\varphi_{wet}(s)$ - фазы, рассчитанные по уточненной атомной модели первой модификации. На Рис.10 приведено несколько сечений синтеза, рассчитанного с такими значениями структурных факторов. Следует подчеркнуть, что качество синтеза определялось не только отсутствием части необходимых рефлексов, но и некоторыми ошибками в фазах, поскольку они отвечали первой модификации, а не второй.

Далее была сделана попытка доопределения неизвестных структурных факторов (и фаз, и модулей), исходя из условия минимума величины (II). Теоретические значения эталонных частот были определены по методике, изложенной в главе 2. Значения эталонных квазичастот, входящих в выражение (II), были пересчитаны из значений частот по формуле (13). На рис.11 приведены "теоретическая" гистограмма и гистограмма, отвечающая синтезу, рассчитанному с восстановленными значениями структурных факторов. Несколько сечений синтеза с восстановленными структурными факторами приведено на рис.10.

Глава 4. ИСПОЛЬЗОВАНИЕ ГИСТОГРАММ В ЗАДАЧЕ УТОЧНЕНИЯ ЗНАЧЕНИЙ ФАЗ СТРУКТУРНЫХ ФАКТОРОВ.

В этой главе устанавливается какое место занимают процедуры уточнения значений фаз структурных факторов за счет дополнительной информации, поставляемой гистограммой - синтеза Фурье, среди других методов уточнения фаз.

Представление дополнительной информации в виде уравнения $\rho(r)=\tau[\rho](r)$.

Проведенный в диссертации анализ показывает, что во многих случаях дополнительная информация о свойствах синтезов электронной плотности может быть выражена математически как свойство функции распределения электронной плотности не меняться при некотором преобразовании этой функции

$$\rho(r) = \tau[\rho](r) , \quad (15)$$

где τ - специальным образом выбранное преобразование, зависящее от того какую именно дополнительную информацию мы пытаемся учесть. Так, например, наличие локальной (некристаллографической) симметрии эквивалентно уравнению (15) с преобразованием $\tau[\rho]$ заключающимся в усреднении электронной плотности в симметрично-связанных точках. Аналогичные представления можно получить и для других видов дополнительной информации (уравнения Сейра, неотрицательность $\rho(r)$, известные границы молекулы,

конечный набор значений, принимаемых функцией $\rho(r)$ и т.п.).

Итерационный подход к определению фаз из уравнения $\rho=\tau[\rho]$.

Уравнение (I5) эквивалентно системе уравнений для структурных факторов функции $\rho(r)$:

$$F(s) = \left| \mathcal{F}_s \left\{ \tau \left[\frac{1}{|v|} \sum_u F(u) e^{i\varphi(u)} e^{-2\pi i(u,r)} \right] \right\} \right| \quad (I6)$$

$$\varphi(s) = \arg \left(\mathcal{F}_s \left\{ \tau \left[\frac{1}{|v|} \sum_u F(u) e^{i\varphi(u)} e^{-2\pi i(u,r)} \right] \right\} \right) \quad (I7)$$

Здесь $\mathcal{F}_s \{ v \}$ обозначает структурный фактор с индексами s , отвечающий функции $v(r)$, $|z|$ - модуль а $\arg(z)$ - фаза комплексного числа z .

Считая модули структурных факторов $\{F(s)\}_s$ известными из рентгеновского эксперимента, мы можем рассматривать систему уравнений (I6)-(I7), как систему уравнений для определения или уточнения фаз $\{\varphi(s)\}_s$.

Основой большого набора методов уточнения значений фаз структурных факторов является итерационная процедура решения фазовой части этих уравнений - уравнений (I7). Радиальная часть (I6) при этом просто игнорируется, а для нахождения фаз решается методом последовательных приближений система уравнений (I7). Многие работы по уточнению значений фаз структурных факторов основаны, явно или неявно, именно на этой итерационной процедуре и различаются лишь тем, какое преобразование $\tau[\rho]$, вводящее ту или иную дополнительную информацию об исследуемом объекте используется.

В диссертации показано, что свойство распределения электронной плотности иметь предписанную гистограмму также может быть представлено в виде (I5). При этом, преобразование $\tau[\rho]$ строится в два этапа. Сначала по имеющейся функции $\rho(r)$ строится модифицирующая функция $\lambda_\rho(t)$ (своя для каждой из возможных функций $\rho(r)$). Эта функция строится как решение уравнения

$$N^{\text{ex}}(\lambda_\rho) = N_\rho(t), \quad (I8)$$

где N^{ex} и N_ρ - кумулятивные функции, отвечающие соответственно точному синтезу Фурье и пробному синтезу $\rho(r)$. Затем осуществляется модификация

$$\rho(r) \rightarrow \rho^m(r) = \lambda_\rho(\rho(r)) = \tau[\rho](r). \quad (I9)$$

Показано, что решение соответствующих этому преобразованию уравнений (I7) методом последовательных приближений составляет основу предложенных недавно методов использования гистограмм для

уточнения значений структурных факторов таких, как histogram specification [2] и histogram matching [3].

В этой главе показано также, что при наличии значительных ошибок в фазах структурных факторов преобразование (18)-(19), восстанавливающее правильную гистограмму, осуществляется с модифицирующей функцией $\lambda_p(t)$ весьма напоминающей широко используемую для модификации электронной плотности функцию $3\rho^2 - 2\rho^3$. Это означает, что "классический" метод модификации электронной плотности можно рассматривать как метод, использующий в неявном виде для работы специфичность формы гистограммы, отвечающей правильному синтезу Фурье.

Глава 5. ПРЯМОЕ РЕШЕНИЕ ФАЗОВОЙ ПРОБЛЕМЫ ДЛЯ НИЗКОУГЛОВЫХ РЕФЛЕКСОВ.

В этой главе предлагается новый подход к прямому (без использования информации по тяжелоатомным производным) решению фазовой проблемы для низкоугловых рефлексов. Подход предполагает использование гистограммы, которой должен обладать искомый синтез электронной плотности, в качестве индикатора правильности определения фаз.

Предложенная процедура может быть разделена на три этапа. На первом этапе генерируется большое число разнообразных наборов фаз (например при помощи датчика случайных чисел) и те из наборов, которые приводят к гистограмме, близкой к предписанному эталону, отбираются как возможные "кандидаты" на правильное решение. На втором этапе методами кластерного анализа изучается множество отобранных "допустимых" наборов фаз. В этом множестве выделяется наиболее компактное подмножество, группирующееся вокруг предполагаемого решения фазовой проблемы. На третьем этапе фазовые наборы усредняются внутри выделенного кластера, что и дает предполагаемое решение фазовой проблемы.

Применение изложенной процедуры может быть проиллюстрировано на следующем тестовом примере.

Модельная структура.

Для проведения теста был искусственно смоделирован и размещен в элементарной ячейке с размерами $76 \times 106 \times 116 \text{ \AA}$ в пространственной группе $P2_12_12_1$ лимер, состоящий из двух молекул карбоксипептидазы. Этот тест предварял работу с белком фактором элонгации C [6], поэтому были взяты параметры элементарной ячейки этого белка и сконструирована модель эквивалентного молекулярного веса. По атомной модели были рассчитаны структурные факторы,

модули которых имитировали далее экспериментально определенные величины $\{F^{\text{ex}}(s)\}$, а фазы использовались лишь для анализа результатов теста. Задачей, решаемой в teste, было определение фаз 29 низкоугловых рефлексов, лежащих в зоне разрешения $d > 30\lambda$. При этом предполагалась известной гистограмма $\{\nu_k^{\text{ex}}\}_{k=1}^K$, отвечающая синтезу, рассчитанному с точными значениями модулей $\{F^{\text{ex}}(s)\}$ и фаз $\{\phi^{\text{ex}}(s)\}$ структурных факторов.

В процессе проведения teste для генерируемых наборов фаз исследовались две характеристики:

а) критерий близости гистограмм, показывающий насколько близки между собой гистограмма синтеза рассчитанного со сгенерированными фазами и гистограмма, отвечающая синтезу с точными значениями фаз;

б) критерий близости синтезов Фурье, показывающий насколько значения сгенерированных фаз близки к правильным.

В этой главе в качестве величины, характеризующей различие в гистограммах, используется величина

$$Q_h = Q_h(\{\nu_k^c\}, \{\nu_k^{\text{ex}}\}) = \sum_{k=1}^K |\nu_k^c - \nu_k^{\text{ex}}| \Delta_k. \quad (20)$$

Мы будем называть эту величину расстоянием между гистограммами $\{\nu_k^c\}$ и $\{\nu_k^{\text{ex}}\}$. Можно, естественно, вводить и другие меры близости гистограмм, например (IO) и т.п. Наши teste не выявили существенных преимуществ использования того или иного способа измерения расстояния между гистограммами перед другими.

Целью решения фазовой проблемы является получение интерпретируемого синтеза. Однаковой величины ошибки в фазах сильных и слабых рефлексов приводят к совершенно различным по величине искажениям синтеза. Особенно это заметно при использовании малого числа рефлексов для расчета синтеза. Поэтому, сравнивая наборы фаз, мы должны учитывать и то, каким по силе рефлексам эти фазы отвечают. Примерами взвешенных критериев близости наборов фаз являются "коэффициент корреляции"

$$\hat{C}(\rho^c, \rho^{\text{ex}}) = \sum_s F^2(s) \cos(\phi^c(s) - \phi^{\text{ex}}(s)) / \sum_s F^2(s)$$

(его максимальное значение равно 1 при $\rho^c = \rho^{\text{ex}}$, минимальное равно -1 при $\rho^c = -\rho^{\text{ex}}$, среднее значение равно 0) или критерий близости соответствующих синтезов

$$\hat{Q}_s(\rho^c, \rho^{\text{ex}}) = \left\{ \int_V [\rho^c(r) - \rho^{\text{ex}}(r)]^2 dv_r / \int_V [\rho^{\text{ex}}(r)]^2 dv_r \right\}^{1/2} =$$

$$= (2 - 2\hat{C})^{1/2}$$

(его минимальное значение равно 0 при $\rho^c = \rho^{ex}$, максимальное равно 2 при $\rho^c = -\rho^{ex}$, среднее значение равно $\sqrt{2}$).

Решая фазовую проблему *ab initio*, нужно, однако, иметь в виду необходимость сведения наборов фаз перед сравнением к одному и тому же началу координат. Дело в том, что все функции вида

$$\rho_{t,\alpha}^c(r) = \rho^c(\alpha r + t),$$

(где t – произвольный вектор, $\alpha = \pm 1$) будут иметь одинаковый набор модулей структурных факторов и одинаковые гистограммы $\{\nu_k^c\}$. Поэтому прежде, чем сравнивать два синтеза Фурье ρ^c и ρ^{ex} , мы должны "сдвинуть" ρ^c в систему координат, в которой он наиболее близок к ρ^{ex} (и, возможно, перейти к энантиоморфу). Мы определим "кристаллографическое" расстояние между ρ^c и ρ^{ex} (или, что то же, взвешенное расстояние между наборами фаз $\{\varphi^c(s)\}$ и $\{\varphi^{ex}(s)\}$) как

$$Q_s = \min_{t \in T} \min_{\alpha=\pm 1} \hat{Q}_s(\rho_{t,\alpha}^c, \rho^{ex}) \quad (21)$$

(здесь T – множество всевозможных сдвигов начала координат). В случае, когда функция $\rho^{ex}(r)$ имеет группу симметрий, отличную от $P1$, множество допустимых сдвигов T может свестись к конечному числу вариантов. Например, для группы $P2_12_12_1$ мы должны, вычисляя Q_s , перебрать 16 вариантов выбора начала координат и энантиоморфа.

Первый этап. Отбор допустимых вариантов.

В таблице 3 показано распределение значений величин Q_h и Q_s для 400 000 наборов фаз (из 29 фаз каждый), сгенерированных при помощи датчика случайных чисел (условия симметрии группы $P2_12_12_1$ были, естественно, учтены при генерации). Первое, что видно из этой таблицы – среди вариантов фазового набора с наименее отличающимися от точной ($Q_h < 0.10$) гистограммами есть как близкие к точному набору, так и достаточно далекие от него ($Q_s \sim 1.0$). Это означает, в частности, что хорошая гистограмма еще не гарантирует правильности синтеза.

Более внимательный взгляд на таблицу 3 позволяет увидеть, что варианты с хорошими гистограммами делятся на две группы: в одну попадают варианты с $Q_s \sim 0.5$, в другую с $Q_s \sim 1.0$. По мере снижения требований к качеству гистограмм (увеличение Q_s) число вариантов в этих группах увеличивается, разброс значений критерия Q_s внутри этих групп возрастает и, наконец, эти группы сливаются. Такая картина позволяет предположить, что существуют по крайней мере два набора фаз, приводящих к предписанной гистограмме при

Таблица 3. Распределение значений величин Q_h и Q_s , отвечающих пробным наборам фаз (в таблице даны количества пробных вариантов, для которых значения Q_h и Q_s лежат в соответствующих интервалах, значения Q_h и Q_s вычисляются согласно (20) и (21) соответственно).

Q_h	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2
Q_s	0	3	10	30	100	300	1000	3000	10000	30000
0.075	1	0	0	0	0	0	0	0	0	0
0.100	1	7	14	6	4	0	5	27	10	0
0.125	5	29	86	103	42	0	44	305	119	0
0.150	1	70	284	359	160	3	138	1290	689	2
0.175	4	84	531	846	428	24	451	3384	2146	9
0.200	1	93	701	1523	872	110	848	6301	4556	71
0.225	1	52	727	2139	1680	269	1502	9732	7571	191
0.250	0	32	685	2812	2852	569	2123	12326	10144	312
0.275	0	27	610	3265	4398	1220	2746	13828	11593	508
0.300	0	6	486	3594	5114	2235	2876	13752	11107	553
0.325	0	2	311	3645	7310	3298	2740	12483	9511	477
0.350	0	2	186	3186	8633	4763	2385	10359	7000	303
0.375	0	0	78	2504	8842	5995	1929	7784	4781	149
0.400	0	0	52	1723	8476	7495	1465	5157	2887	49
0.425	0	0	26	1145	7424	8358	1383	2950	1716	26
0.450	0	0	8	651	6212	8566	1520	1492	891	10
0.475	0	0	3	377	4527	8204	2001	658	453	1

заданных модулях структурных факторов.

Второй этап. Кластерный анализ множества допустимых вариантов.

Поскольку в реальной ситуации эталонная гистограмма предсказывается с некоторой ошибкой, все варианты фазового набора, обеспечивающие не слишком большое значение критерия Q_s , должны рассматриваться как не противоречие эталонной гистограмме. В данном тесте такими допустимыми наборами можно считать, например, 39 вариантов, обеспечивающих значение $Q_h(p_j^c, p_{j'}^{\text{ex}}) < 0.1$. В реальной ситуации точные значения фаз неизвестны и значения критерия близости $Q_s(p_j^c, p_{j'}^{\text{ex}})$ сгенерированных фаз к точным не могут быть вычислены. Однако, могут быть рассчитаны попарные расстояния между допустимыми наборами фаз $Q_s(p_j^c, p_k^c)$. Процедуры кластерного анализа позволяют на основе анализа матрицы попарных расстояний между допустимыми вариантами составить представление о том, как эти варианты распределены в многомерном "конфигурационном" пространстве: образуют ли они одну или несколько компактных групп или разбросаны равномерно по всему пространству. Суть процедуры кластерного анализа заключается в том, что мы пытаемся объединить в кластеры варианты, находящиеся достаточно близко друг от друга

(варианты с $Q_s(\rho_j^c, \rho_k^c) < \varepsilon$). Понятно, что по мере увеличения порогового значения ε число вариантов в кластерах возрастает, а число кластеров убывает. Ход процесса объединения вариантов в кластеры представлен на рис. II (порядок, в котором изображены варианты на рисунке, выбран исходя из удобства представления дерева; он не соответствует, конечно, порядку, в котором эти варианты были генерированы). Анализ проводился при помощи программы РIM пакета программ BMDP [7].

Рис. II показывает, что 39 допустимых вариантов распадаются на два кластера: кластер А из 21 варианта и кластер В из 18 вариантов. При этом оказалось (таблица 4), что в кластер А попали варианты фазового набора, находящиеся на расстоянии Q_s от 0.23 до 0.66 от точного, а в кластер В варианты, находящиеся на расстоянии $Q_s > 0.89$ от точного. Подчеркнем, что при разбиении вариантов на кластеры использовалась только матрица попарных

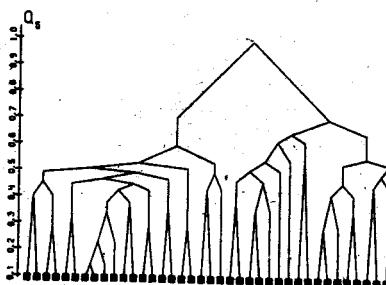


Рис. II. Схема разделения на кластеры вариантов в тесте с модельным белком.

расстояний между вариантами и никак не использовалась информация о том, насколько эти варианты на самом деле близки к точному набору.

Третий этап. Усреднение вариантов внутри кластера.

В качестве решения фазовой проблемы, представляющейся кластером, для каждого из кластеров выбирался его "центр тяжести". Более точно, в каждом из кластеров определялись для каждого из рефлексов показатель достоверности $m(s)$ и "наилучшая" фаза $\varphi^{best}(s)$ по формуле:

$$m(s) e^{i\varphi^{best}(s)} = \frac{1}{M} \sum_{j=1}^M e^{i\varphi_j(s)}$$

Здесь M – число вариантов в кластере (оно равно 21 для кластера А). $\varphi_j(s)$ – значение соответствующей фазы в j -ом наборе. Естественно, перед усреднением все варианты были приведены к

единому началу координат и энантиоморфу. Для этого один из вариантов кластера был принят за реперный, а остальные приводились каждый в ту систему координат, в которой обеспечивалось минимальное расстояние \hat{Q}_s от репера. С модулями $\{F^{\text{ex}}(s)\}$ и полученными таким образом значениями фаз $\{\varphi^{\text{best}}(s)\}$ был рассчитан синтез $\rho_A(r)$. Аналогичная процедура усреднения была применена к 18 вариантам кластера В и был рассчитан аналогичным образом синтез $\rho_B(r)$. На рис.12 приведены карты распределения электронной плотности в одном из сечений элементарной ячейки, отвечающие синтезам $\rho_A(r)$ и $\rho_B(r)$ и синтезу $\rho^{\text{ex}}(r)$, рассчитанному с точными значениями фаз.

В таблице 4 приведены средние значения показателей достоверности и фазовой ошибки для фаз $\{\varphi^{\text{best}}(s)\}$, полученных усреднением по кластерам А и В. Из таблицы видно, что кластер, соответствующий правильному решению (кластер А), характеризуется большим средним показателем достоверности и меньшим разбросом вариантов вокруг среднего, нежели кластер В, отвечающий постороннему решению фазовой проблемы.

Тестовое определение фаз для цитохрома b5.

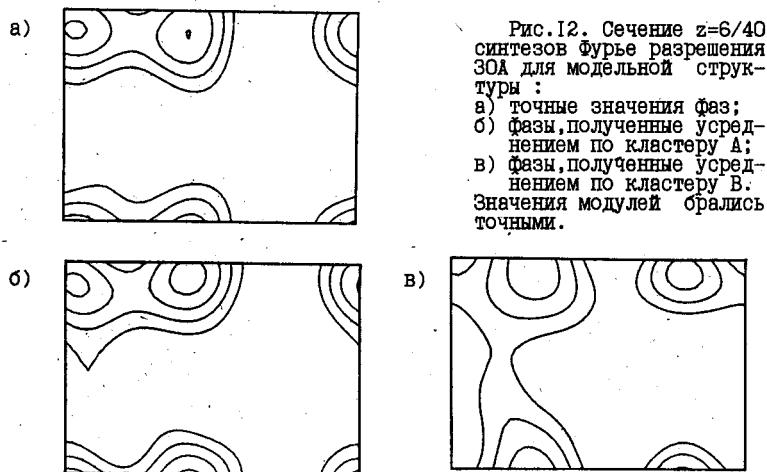
В качестве следующего объекта для тестирования метода был выбран белок цитохром b5 [8] (пространственная группа P2₁2₁2₁, размеры элементарной ячейки 65x46x30 Å). По координатам атомов, взятым из банка белковых молекул, были рассчитаны точные значения

Таблица 4. Характеристики кластеров, выделенных в тесте с модельной структурой.

	Кластер А	Кластер В
Число вариантов в кластере	21	18
Расстояние Q_s между элементами кластера и точным решением фазовой проблемы		
min	0.23	0.89
max	0.66	1.12
average	0.45	0.97
$\langle m \rangle_s$	0.52	0.41
$\langle Q_s(\rho^{\text{best}}, \rho_j) \rangle_j$	0.42	0.54
$Q_s(\rho^{\text{best}}, \rho^{\text{ex}})$	0.34	0.95
C ($\rho^{\text{best}}, \rho^{\text{ex}}$)	0.94	0.55
$\langle \varphi^{\text{best}} - \varphi^{\text{ex}} \rangle_s$ (deg.)	40	71

структурных факторов, по ним рассчитан синтез Фурье разрешения 13.6 Å (29 рефлексов) и построена эталонная гистограмма. Далее была поставлена задача определения значений фаз структурных факторов зоны 13.6 Å, используя для этой цели лишь значения модулей структурных факторов и эталонную гистограмму.

Для этой цели было сгенерировано 500 000 случайных наборов фаз и из них отобрано 49 вариантов, дающих близкую к эталону гистограмму ($Q_h < 0.1$). Процесс разбики этих 49 вариантов на кластеры иллюстрируется рис.13. В лучшем из них было получено решение фазовой проблемы со средней фазовой ошибкой 32° и коэффициентом корреляции с точными фазами $C=0.92$. На рис.14



показаны карты распределения электронной плотности для точного синтеза и синтеза с определенными указанным методом фазами.

Тестовое определение фаз для белка Бена-Джонса.

Этот белок кристаллизуется в пространственной группе $P2_12_12$ в ячейке с размерами $55 \times 52 \times 43$ Å, определялись фазы 25 низкоугловых рефлексов (разрешения до 16 Å). Ход вычислительного эксперимента был аналогичен тесту с цитохромом. Было сгенерировано 100 000 случайных вариантов фазового набора и отобрано для дальнейшего анализа 488 вариантов. Процесс разбиения этих вариантов на кластеры показан на рис.15. Выделение кластера и усреднение вариантов внутри кластера позволило получить решение фазовой проблемы со средней фазовой ошибкой 41° и коэффициентом корреляции с правильным решением $C=0.93$. На рис.16 показаны карты

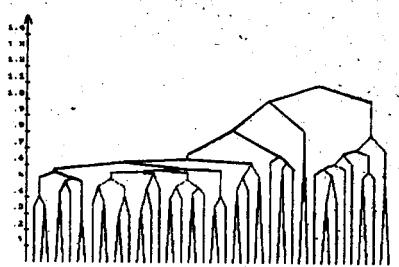


Рис.13. Схема разделения на кластеры вариантов в теоте с цинтохромом.

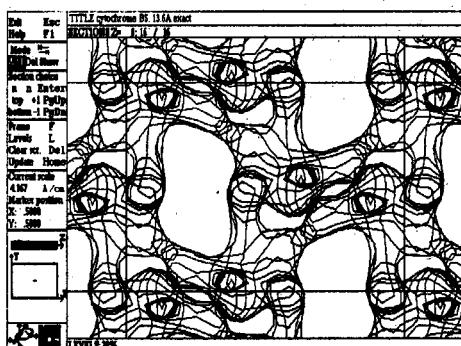
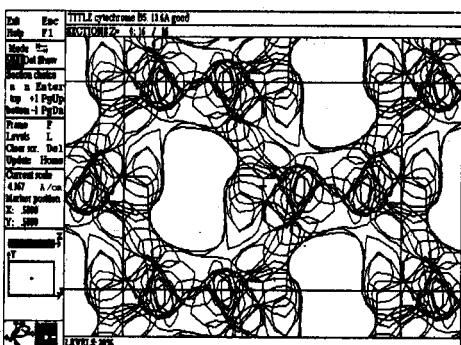


Рис.14. Точный синтез Фурье для цинтохрома и синтез с фазами, определенными *ab initio*.

Показана проекция вдоль кристаллографической оси z.



электронной плотности, рассчитанные с фазами, определенными предложенным методом.

Таким образом, результаты тестов показывают, что следующая процедура :

- генерация случайных наборов фаз и отбор вариантов

фазового набора, приводящих к правильной гистограмме;
 б) разбиение отобранных вариантов на кластеры, исходя из матрицы попарных расстояний Q_S между вариантами;
 в) усреднение вариантов внутри кластеров, приводит к получению небольшого числа возможных решений фазовой проблемы, в число которых входит и решение достаточно близкое к правильному.



Рис.15. Схема разделения на кластеры вариантов в тесте с белком Бен-Джонса.



Рис.16. Сечения синтеза Фурье для белка Бен-Джонса, рассчитанного с фазами, определенными *ad initio*. Показаны положения атомов в модели белка. (На синтезе показаны линии уровня, выделяющие 10, 30 и 50 % объема элементарной ячейки).

Определение фаз при разрешении 30 Å для фактора элонгации G.

Работа по определению структуры фактора элонгации G из *Tetmus Thermophilus* ведется под руководством Ю.Н.Чиргадзе в Институте белка АН СССР и в НИВЦ АН СССР. Белок кристаллизован в пространственной группе $P2_12_12$, и имеет параметры ячейки около $76 \times 106 \times 116 \text{ Å}$ [6].

Для предсказания гистограммы, которой должен обладать "правильный" синтез электронной плотности разрешения 30 Å, для G-фактора был применен подход, изложенный в главе 2.

Описанная выше модельная структура из двух молекул карбоксипептидазы допускает различные варианты размещения в элементарной ячейке. Нами были взяты три различных варианта допустимой упаковки и рассчитаны соответствующие гистограммы, отвечающие разрешению 30 Å. Эти гистограммы оказались близки между собой (значения попарных расстояний Q_{ij} между гистограммами не превышали 0.1). Далее работа по определению фаз проводилась в нескольких вариантах: с использованием поочередно каждой из этих трех гистограмм и с использованием усредненной гистограммы. Во всех четырех вариантах работы были получены сходные результаты. Ниже мы опишем кратко ход работ с использованием усредненной гистограммы.

Было сгенерировано 500 000 вариантов фазового набора и из них отобрано 44 варианта, дающих наилучшее соответствие с смоделированной гистограммой ($Q_{ij} < 0.125$). Рис.17 иллюстрирует процесс выделения кластеров. В результате усреднения вариантов по выделенному кластеру были получены фазы, обладающие средним показателем достоверности 0.54. Разброс вариантов в кластере относительно среднего характеризовался величиной $Q_s = 0.46$.

Рис.18 представляет проекцию элементарной ячейки вдоль кристаллографической оси x. Полученный синтез находится в хорошем соответствии с результатами, полученными методами электронной микроскопии.

Глава 6. Вычислительные проблемы.

В этой главе рассматриваются некоторые вычислительные проблемы, возникающие при практической реализации подходов, изложенных в предыдущих главах.

Алгоритм быстрого дифференцирования.

Центральная тема этой главы связана с задачами минимизации функций, зависящих от большого числа переменных. Многие из

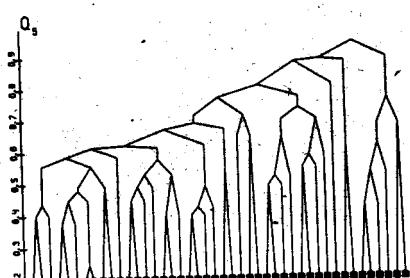


Рис.17. Схема разделения на кластеры вариантов при решении фазовой проблемы для фактора элонгации G.

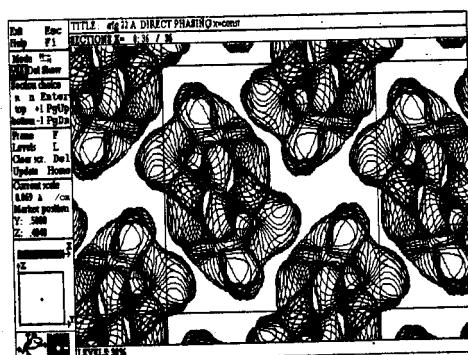


Рис.18. Синтез Fourier разрешения 22А для фактора элонгации G с фазами, определенными *ab initio*. Показана проекция элементарной ячейки вдоль оси x.

разобранных выше подходов сводились в конечном счете к минимизации некоторого сложным образом определенного критерия качества пробного набора фаз. При этом вычисление каждого значения такого критерия требует обычно ощутимых затрат процессорного времени компьютера. До последнего времени считалось, что наиболее трудноразрешимой проблемой при минимизации такого рода функций является расчет градиента минимизируемой функции (необходимого при минимизации, поскольку именно он определяет направление, в котором надо сдвигать варьируемые параметры для того, чтобы уменьшить значение целевой функции). Так, при вычислении частных производных по разностным формулам для расчета градиента требуется в n раз больше времени, нежели для расчета одного значения функции (n – число варьируемых параметров). При времени расчета одного значения критерия, исчисляемого минутами, и значительном числе переменных (в задаче уточнения атомной структуры это число может достигать десятков

тысяч) задача расчета градиента может выглядеть неразрешимой. Однако, оказывается [10], что для любой функции $f(x)$ произвольного числа переменных n можно построить алгоритм, требующий для расчета значений всех компонент градиента практически такое же время, которое нужно для расчета одного значения функции $f(x)$. Этот факт имеет чрезвычайно важное методологическое значение для задач рентгеноструктурного анализа. Из него следует, что при локальном уточнении структуры исследуемого объекта в качестве критерия, контролирующего правильность модели, может использоваться любая характеристика объекта, доступная для вычисления на имеющемся у исследователя компьютере. В главе 6 излагается реализация этой общей идеи для задач, возникающих при определении пространственной структуры вещества методами рентгеноструктурного анализа. В частности, разработанные алгоритмы позволили практически реализовать подходы, изложенные в предыдущих главах.

Кристаллы исследуемого вещества обладают, как правило, нетривиальной симметрией. Это обстоятельство позволяет, в частности, существенно экономить время компьютерных расчетов за счет использования этой симметрии. В диссертации показано, как симметрия может быть учтена в процессе быстрого расчета градиента минимизируемой функции.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ.

1. Гистограммы синтезов Фурье конечного разрешения для функций распределения электронной плотности в кристаллах белков представляют новый вид дополнительной информации об исследуемом объекте. Эти гистограммы чувствительны к ошибкам в значениях фаз структурных факторов и к пропуску отдельных рефлексов и могут использоваться как индикатор правильности значений структурных факторов, использованных для расчета синтеза.

2. Анализ гистограмм, отвечающих синтезам электронной плотности в белках с известной пространственной структурой, показывает, что при среднем и высоком разрешении синтезов гистограммы определяются, по существу, одним параметром – средней электронной плотностью в кристалле. Получены формулы, позволяющие предсказывать гистограммы для белков с неизвестной пространственной структурой. Предложен подход к моделированию гистограмм, отвечающим синтезам низкого разрешения.

3. Знание эталонной гистограммы для искомого синтеза

позволяет существенно повышать интерпретируемость синтеза электронной плотности в ситуации, когда для части структурных факторов значения фаз (а иногда и модулей) неизвестны.

4. Предложенный в диссертации подход к уточнению значений фаз структурных факторов, а также предложенные недавно такие методы уточнения значений фаз, как Histogram Specification [2] и Histogram Matching [3], могут рассматриваться как разные подходы к решению одной и той же задачи - нахождению функции с заданными модулями структурных факторов, обладающей предписанной гистограммой. Широко распространенный на практике Density Modification метод может рассматриваться как упрощенный подход к решению сформулированной выше задачи.

5. Знание эталонной гистограммы для исследуемого объекта позволяет решать задачу прямого (только по данным рентгеновского эксперимента с нативным белком) определения значений фаз низкоугловых рефлексов.

6. Использование алгоритмов быстрого преобразования Фурье и быстрого дифференцирования позволяет строить эффективные алгоритмы для практического решения задач, связанных с определением и уточнением значений фаз структурных факторов исходя из информации, содержащейся в гистограммах синтезов Фурье.

Результаты диссертации опубликованы в следующих работах :

1. Чиргадзе Ю.Н., Сергеев Ю.В., Фоменкова Н.П., Орешин В.Д., Никонов С.В., Лунин В.Ю. "Пространственная структура γ -криSTALLинов II и III_b из глазной линзы теленка при разрешении 5 \AA ". Тезисы V Всесоюзного симпозиума по химии и физике белков и пептидов, Баку, 1980, стр.138.

2. Chirgadze Yu.N., Oreshin V.D., Sergeev Yu.V., Nikonov S.V., Lunin V.Yu. "Structure of γ -crystallin. II_b from calf lens at 5 \AA resolution". FEBS Letters, 1980, 118, p.2, 296-298.

3. Chirgadze Yu.N., Sergeev Yu.V., Fomenkova N.P., Lunin V.Yu., Urzhumtsev A.G. "The structure of γ -crystallin. III_b from calf lens"., 1981, Acta Cryst., A37, S C-40.

4. Фоменкова Н.П., Чиргадзе Ю.Н., Сергеев Ю.В., Невская Н.А., Никонов С.В., Лунин В.Ю., Уржумцев А.Г. "Структура γ -криSTALLина: ход цепи, пространственная гомология доменов, расположение функциональных групп". I Всесоюзный биофизический съезд. Тезисы докладов пленарных лекций и симпозиальных заседаний, Москва, 1982, стр. 17.

5. Чиргадзе Ю.Н., Невская Н.А., Лунин В.Ю., Уржумцев А.Г. "Молекулярная структура белка глазной линзы γ -кристаллина при разрешении 2.7 \AA ". 16 конференция Европейских биохимических обществ., Тезисы докладов, Москва, 1984, стр.232.
6. Lunin V.Yu., Urzhumtsev A.G. "Program construction for refinement of the atomic structure of macromolecules based on the fast differentiation algorithm", 1984, Acta Cryst., A40, S C-18.
7. Chirgadze Yu.N., Fomenkova N.P., Nevskaya N.A., Lunin V.Yu., Urzhumtsev A.G. "Crystal structure of the calf lens protein γ -crystallin III", 1984, Acta Cryst., A40, S C-32&
8. Lunin V.Yu., Urzhumtsev A.G. "Improvement of Protein Phases by Coarse Model Modification", 1984, Acta Cryst., A40, 269-277.
9. Lunin V.Yu., Urzhumtsev A.G., Vernoslova E.A., Chirgadze Yu.N., Nevskaya N.A., Fomenkova N.P. "Phase Improvement in Protein Crystallography Using a Mixed Electron Density Model", 1985, Acta Cryst., A41, 166-171.
10. Lunin V.Yu., Urzhumtsev A.G. "Program Construction for Macromolecule Atomic Model Refinement Based on the Fast Fourier Transform and Fast Differentiation Algorithms", 1985, Acta Cryst., A41, 327-333.
- II. Lunin V.Yu. "Use of the Fast Differentiation Algorithm for Phase Refinement in Protein Crystallography", 1985, Acta Cryst., A41, 551-556.
12. Лунин В.Ю. "Использование информации о распределении значений электронной плотности в белках. I. Восстановление недостающих структурных факторов.", Препринт, ОНТИ НЦБИ, Пущино, 1986.
13. Чиргадзе Ю.Н., Невская Н.А., Фоменкова Н.П., Никонов С.В., Сергеев Д.В., Бражников Е.В., Гарбер М.Б., Лунин В.Ю., Уржумцев А.Г., Вернослова Е.А. "Пространственная структура гамма-кристаллина III β из хрусталика глаза теленка при разрешении 2.5 \AA ", 1986, Доклады АН СССР, 290, в.2, 492-495.
14. Urzhumtsev A.G., Lunin V.Yu., Luzyanina T.B. "Macromolecular position determination with a noisy synthesis at low resolution". 1986, Tenth European Crystallographic Meeting, Collected Abstracts, Wroclaw, Poland, 1A-01, 51-52.
15. Lunin V.Yu. "Use of the Information on Electron Density Distribution in Macromolecules", 1988, Acta Cryst., A44, 144-150.
16. Лунин В.Ю. "Восстановление недостающих структурных факторов при рентгеноструктурном исследовании макромолекул", 1988, Доклады АН СССР, 299, в.2, 363-366.
17. Urzhumtsev A.G., Lunin V.Yu., Luzyanina T.B. "Bounding a

- Molecule in a Noisy Synthesis", 1989, *Acta Cryst.*, A45, 34-39.
18. Chirgadze Yu.N., Nevskaia N.A., Vernoslova E.A., Nikonorov S.V., Sergeev Yu.V., Brazhnikov E.V., Fomenkova N.P., Lunin V.Yu., Urzhumtsev A.G. "Refined Structure of calf lens γ -crystallin IIIb at 2.5 \AA resolution", 1989, Twelfth European Crystallographic meeting, Collected Abstracts, Vol.2, p.362, Moscow.
19. Lunin V.Yu., Skovoroda T.P. "Frequency analysis - new approach to phase improvement in protein crystallography", 1989, Twelfth European Crystallographic meeting, Collected Abstracts, Vol.3, p.146, Moscow.
20. Urzhumtsev A.G., Lunin V.Yu., Vernoslova E.A. "FROG : new refinement program", 1989, Twelfth European Crystallographic meeting, Collected Abstracts, Vol.3, p.193, Moscow.
21. Уржумцев А.Г., Лунин В.Ю., Вернослова Е.А. "FROG - комплекс программ для локального уточнения атомной структуры макромолекул", 1989, Программное обеспечение ЭВМ, вып.86 "Компьютерные методы белковой инженерии", стр.86-97, Минск.
22. Urzhumtsev A.G., Lunin V.Yu., Vernoslova E.A. "New refinement program FROG gives new possibilities to study macromolecular models", 1989, Proceedings of the first international meeting on Structure and Chemistry of Ribonucleases, 410-415, Moscow.
23. Lunin V.Yu. "The Calculation of the Second Moments for the Values of Fourier Syntheses with Random Structure Factors", 1989, *Acta Cryst.*, A45, 501-505.
24. Urzhumtsev A.G., Lunin V.Yu., Vernoslova E.A. "FROG - high-speed restraint-constraint refinement program", 1989, *J. Appl. Cryst.*, 22, 500-506.
25. Lunin V.Yu., Urzhumtsev A.G., Skovoroda T.P. "Direct low-resolution phasing in protein crystallography", 1990, XV International Crystallographic Congress, Collected Abstracts, C 113, Bordeaux, France.
26. Lunin V.Yu., Urzhumtsev A.G., Skovoroda T.P. "Direct Low-Resolution Phasing from Electron-Density Histograms in Protein Crystallography", *Acta Cryst.*, 1990, A46, 540-544.
27. Lunin V.Yu., Skovoroda T.P. "Frequencies-Restrained Structure Factor Refinement. I. Histogram simulation ", 1991, *Acta Cryst.*, A47, 45-52.
28. Lunin V.Yu., Vernoslova E.A. "Frequencies-Restrained Structure Factor Refinement. II. Comparison of Methods", 1991, *Acta Cryst.*, A47, 238-243.

ЧИТАЕМАЯ ЛИТЕРАТУРА

1. Luzzati V., Mariani P. & Delacroix H. "X-ray crystallography at macromolecular resolution : a solution of the phase problem", Macromol.Chem.,Macromol.Symp., 1988, 15, 1-17.
2. Harrison R.W. "Histogram Specification as a Method of Density Modification", J.Appl.Cryst., 1988, 21, 949-952.
3. Zhang K.Y.J. & Main P. "Histogram Matching as a New Density Modification Technique for Phase Refinement and Extension of Protein Molecules", Acta Cryst., 1990, A46, 41-46.
4. Wright C.S., Alden R.A. & Kraut J. "Structure of subtilisin BPN' at 2.5 \AA resolution", Nature, 1969, 221, 235-242.
5. Chirgadze Yu.N., Nevskaya N.A., Vernoslova E.A., Urzhumtsev A.G., Lindley P. & Bibby M. "Structure refinement of "dry" crystal form of calf eye lens- γ -crystallin IIIb at 1.9 \AA resolution", Twelfth European Crystallographic Meeting, 1989, Moscow, Collected Abstracts, vol.2, 363.
6. Chirgadze Yu.N., Nikonov S.V., Brazhnikov E.V., Garber M.B. & Reshetnikova L.S. "Crystallographic study of elongation factor G from *Thermus thermophilus HB8*", J.Mol.Biol., 1983, 168, 449-450.
7. Biomedical Computer Programs P-Series, 1977, Ed. by W.J.Dixon.
8. Mathews F.S., Levine M., Argos P. "The structure of calf liver cytochrome b5 at 2.8 \AA resolution", 1971, Nature New Biol., v.233, 15-16.
9. Furey W.J., Wang B.G., Yoo C.S., Sax M. "Phase Extension and Refinement of Bence-Jones Protein RHE (1.9 \AA)", Acta Cryst., 1979, A35, 810-817.
10. Ким К.В., Нестеров Ю.Е., Черкасский Б.В. "Оценка трудоемкости вычисления градиента", Доклады АН СССР, 1984, Т.275, I308-I309.



21.1.92 г. Зак. 3821Р Тир. 150 экз. уч.-изд.л. 2,0
Отпечатано на ротапринте в ОНТИ ЛИЦ