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.I. Общее описание метода.

2.I.I. 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 (2.5-2.0 Å), на эта-

пе построения карт распределения электронной плотности используется лишь часть этого набора данных, имеющая более низкое разрешение (3.5-3.0 Å). Это связано с тем, что обычно используемые методы определения набора фаз (изоморфного замещения, молекулярного замещения и т.п.) не позволяют порой получить значение фазы для отражений относительно высокого разрешения. Отсутствие этих данных при расчете синтеза электронной плотности может существенно усложнить последующее построение атомной модели по синтезу. Наличие ошибок в имеющихся значениях фаз дополнительно снижает качество карт распределения электронной плотности. Экспериментально определенные значения модулей структурных факторов также содержат некоторые ошибки, однако в данной ситуации, при наличии более сильного влияния на качество синтеза ошибок определения фаз, можно считать значения модулей известными точно.

Таким образом, из-за указанных выше причин, построение атомной модели по соответствующим картам распределения электронной плотности затрудняется, а в ряде случаев становится просто невозможным. В такой ситуации дальнейшее продвижение в исследовании структуры возможно путем решения задачи улучшения набора фаз, то есть

I) определения значений фаз для тех отражений, где эти значения ранее отсутствовали, и

2) уточнения известных значений фаз.

Карты распределения электронной плотности, построенные ис использованием более полного и точного набора фаз, будут иметь более высокое качество и могут дать возможность построения или корректировки атомной модели структуры либо ее

части.

Исходной информацией для решения такой задачи улучшения набора фаз являются значения модулей и приближенные значения фаз для некоторого набора отражений и значения модулей для дополнительного набора отражений, вообще говоря, более высокого разрешения, чем первый. Далее, на протяжении данной главы, кроме специально оговоренных случаев, будет рассматриваться ситуация, когда отсутствует какая-либо дополнительная информация о структуре, кроме той, что структура состоит из отдельных атомов. Вопросы использования дополнительной информации в рамках описываемой методики разобраны отдельно в п.п. 2.7, 2.8.

2.I.2. Принцип использования грубой модели.

Для расширения набора фаз в описанной выше ситуации Агарвал и Айзекс [6] предложили использовать следующий подход.

I) По функции распределения электронной плотности в кристалле, рассчитанной по исходному набору фаз, строится модель структуры, состоящая из фиктивных атомов одного и того же типа. Основное требование к данной модели - удовлетворительное воспроизведение исходной функции распределения электронной плотности. Такая модель строится автоматически и не претендует на поатомное соответствие исследуемой структуре.

2) Параметры фиктивных атомов (координаты центров и температурные факторы) изменяются согласно требованию минимальности критерия типа R -фактора.

3) По параметрам модифицированной модели рассчитывается набор структурных факторов.

4) Рассчитывается новая функция распределения электронной плотности, вообще говоря, уже более высокого разрешения, чем исходная. При расчете этой функции используются полный набор экспериментально определенных модулей структурных факторов и набор соответствующих значений фаз структурных факторов, рассчитанных по модифицированной модели.

В качестве обоснования работоспособности метода Агарвал и Айзекс приводили следующий довод:

благодаря большому радиусу сходимости метода уточнения параметров атомной модели в обратном пространстве значительная часть фиктивных атомов будет двигаться по направлению к позициям атомов точной модели; такое уточнение позволяет существенно улучшить соответствие модулей структурных факторов (экспериментально определенных и рассчитанных по модели) и, по-видимому, уточнить значения фаз.

Предложенный метод был опробован Агарвалом и Айзексом на данных по инсулину. Набор фаз структурных факторов среднего разрешения (до 3.0 Å), определенных методом изоморфного замещения, был расширен до I.9 Å и до I.5 Å. В качестве контрольного использовался набор фаз, рассчитанных по атомной модели инсулина, уточненной до значения кристаллографического фактора достоверности R = 0.II3 по данным разрешением до I.5 Å. Использованная методика позволила определить фазы структурных факторов с точностью 70° для I.5-ангстремного набора и 65° - для I.9-ангстремного набора.

Мы не будем описывать здесь способы решения Агарвалом и Айзексом некоторых практических проблем, а коснемся их ниже,

в сравнении с нашей реализацией данной идеи. Отметим лишь, что наша методика использования грубой модели содержит ряд существенных отличий от методики Агарвала и Айзекса и, фактически, общей в этих методиках является лишь основная идея метода.

2.1.3. Теоретическое рассмотрение метода.

Рассмотрим ситуацию, описанную выше в п.2.1.1.

Заметим, прежде всего, что, действуя предложенным Агарвалом и Айзексом способом, можно не только расширять набор фаз, но и уточнять значения имеющихся фаз. Поэтому далее мы будем говорить об улучшении набора фаз.

Рассматривая метод теоретически, необходимо обратить внимание на следующие важные с нашей точки зрения моменты.

Прежде всего, использование атомной модели позволяет учесть важную дополнительную информацию - атомное строение молекулы. Именно внесение этой дополнительной информации и позволяет надеяться на улучшение набора фаз.

Дополнительно, переход к атомной модели позволяет существенно понизить размерность исходной задачи улучшения функции распределения электронной плотности.

Укажем далее на связь этого метода с другими методами улучшения наборов фаз.

Данный подход может рассматриваться как вариант метода модификации электронной плотности, где на функцию распределения электронной плотности накладываются требования неотрицательности и наилучшего соответствия рентгеновскому эксперименту. Введение фиктивных атомов можно здесь рассматривать как удобное средство для модификации функции.

Поскольку метод Сэйра использует атомность структуры

(в качестве гипотезы), то рассматриваемый подход, в котором эта гипотеза заведомо выполняется, можно трактовать и как один из путей решения системы уравнений Сэйра для точечных атомов. Здесь, в отличие от обычных методов решения системы методом наименьших квадратов либо путем использования тангенс-формулы, уравнения выполнены точно, а задача состоит в определении структурных факторов, наилучшим образом соответствующих эксперименту.

Показанная таким образом связь данного метода с другими методами улучшения наборов фаз дает дополнительную уверенность в его эффективности.

Можно дать некоторые количественные оценки качества фаз, рассчитанных по грубой атомной модели [201].

Предположим, что точная и приближенная модели состоят из одинакового числа N атомов, координаты центров которых отличаются на случайные величины $\vec{\Delta}_j$, независимые и имеющие одинаковое сферически симметричное распределение вероятностей. Обозначим через $F(\vec{S})e^{i\varphi(\vec{S})}$ и $F_c(\vec{S})e^{i\varphi_c(\vec{S})}$ структурные факторы, рассчитанные соответственно по точной и приближенной моделям. Повторяя выкладки работы [224], основанные на применении центральной предельной теоремы, можно получить формулу для совместного распределения величин $P(F,\varphi)$ в предположении, что координаты центров атомов грубой модели есть некоторые известные величины, а $F(\vec{S})$ и $\varphi(\vec{S})$ - случайные функции, зависящие от случайных переменных $\vec{\Delta}_j$. Отсюда, при некоторых дополнительных предположениях и рассмотрении экспериментально определенных величин $F_0(\vec{S})$ как некоторой конкретной реализации значений $\vec{\Delta}_j$, вытекает упрощенное представление условной вероятности

$$P(\varphi|F_0) \sim \exp\{2\mu F_0 F_c \cos(\varphi-\varphi_c)/\sigma_N^2(1-\mu^2)\}, \qquad (62)$$

где $\sigma_N^2 = \sum_{j=1}^{N} f_j^2(|\vec{S}|)$,

$f_j(|\vec{S}|)$ – фактор рассеяния j-го атома,

$\mu(\vec{S}) = \langle \cos 2\pi(\vec{S},\Delta_j)\rangle$ – величина, характеризующая ошибку в определении координат атомов.

Из определения $\mu(\vec{S})$ видно, что при уменьшении ошибок $\vec{\Delta}_j$ величины $\mu(\vec{S})$ возрастают и соответствующие распределения (62) обостряются. В свою очередь, обострение распределений означает уменьшение ошибок в определении значений фаз. В предельном случае, когда $\mu(\vec{S})$ стремится к единице, функция $P(\varphi|F_0)$ становится близкой к δ-функции Дирака, то есть фаза определяется однозначно. В промежуточных случаях в качестве характеристики, определяющей надежность значения фазы $\varphi_c(\vec{S})$, можно использовать, например, среднюю ошибку

$$\langle|\Delta\varphi|\rangle = \langle|\varphi-\varphi_c|\rangle = \int_{\varphi_c-\pi}^{\varphi_c+\pi} |\varphi-\varphi_c|\cdot P(\varphi|F_0)\,d\varphi = \frac{1}{\pi I_0(\alpha)}\int_0^{\pi} \varphi e^{\alpha\cos\varphi}\,d\varphi, \qquad (63)$$

где

$$\alpha = \frac{2\mu F_0 F_c}{\sigma_N^2(1-\mu^2)} \qquad (64)$$

Здесь также видно, что при уменьшении ошибок $\vec{\Delta}_j$ уменьшается и средняя ошибка определения фаз $\langle|\Delta\varphi|\rangle$. Достаточно хорошая аппроксимация стартовой грубой моделью исходного синтеза позволяет надеяться, что значения фаз, определенные по модифицированной модели, будут точнее исходных.

2.2. Построение грубой модели.

2.2.I. Построение исходного синтеза.

Для успешного применения метода необходимо, чтобы функция распределения электронной плотности, по которой строится грубая модель, достаточно хорошо соответствовала истинной функции. Блоу и Криком [216] было показано, что в ситуации, когда известны набор модулей структурных факторов $\{F_0(\vec{s})\}$ и набор функций распределения вероятностей фаз $\{P_{\vec{s}}(\varphi)\}$, наилучшим синтезом (в смысле минимального среднеквадратичного отклонения от правильной функции) является "взвешенный" синтез с коэффициентами Фурье

$$m(\vec{s})F_0(\vec{s})e^{i\varphi_0(\vec{s})}, \tag{65}$$

где

$$m(\vec{s})e^{i\varphi_0(\vec{s})} = \langle e^{i\varphi}\rangle_{P_{\vec{s}}(\varphi)} = \int_0^{2\pi} e^{i\varphi} P_{\vec{s}}(\varphi)\, d\varphi \tag{66}$$

Для описываемой методики существенно, что синтез должен быть рассчитан в абсолютной шкале. Результатом неправильного шкалирования будет набор фиктивных атомов, слишком грубо описывающих правильное распределение электронной плотности. Таким образом, необходима по возможности точная оценка величины заряда элементарной ячейки F_{000} и шкального коэффициента для приведения значений модулей структурных факторов к абсолютной шкале. Для определения шкального коэффициента можно воспользоваться любой из известных методик, например, определением его из графика Вильсона [I,226].

Выбор оптимального числа делений при построении синтеза нами отдельно не исследовался. Отметим лишь, что стандартно рекомендуемый выбор числа делений - шаг по каждой координат-

ной оси приблизительно равен трети верхнего предела разрешения (см., например, [63]) - вполне подходит для наших целей. При выборе шага сетки полезны также соображения о точности исходных данных [227].

2.2.2. Схема построения грубой модели.

Прежде всего, подчеркнем еще раз, что в нашей постановке вопроса основным требованием к грубой атомной модели является по возможности более точное воспроизведение моделью имеющегося синтеза. Произвольные стереохимические требования и поатомное соответствие исследуемой структуре из рассмотрения исключаются.

Грубая атомная модель строится из фиктивных атомов одного и того же типа, но с разными значениями температурного фактора. Обычно при работе на среднем разрешении таким атомам приписывается рассеивающая способность атомов азота. Это приводит к ситуациям, когда сгущения электронной плотности могут моделироваться несколькими фиктивными атомами, расположенными близко один от другого. В принципе, можно было бы в больших сгустках электронной плотности располагать какие-то более тяжелые атомы. Мы специально не делаем этого потому, что при работе на среднем разрешении такие сгустки образуются из-за наложения изображений нескольких атомов, и мы не должны пренебрегать этой информацией. Тяжелые атомы структуры, как правило, локализуются заранее, и их вклад может быть предварительно вычтен из синтеза. Кроме того, представление тяжелого атома совокупностью более легких в пределах данной методики является вполне допустимым.

Эффективность описанного подхода продемонстрирована на рис. 4. В численном эксперименте, выполненном на модельной

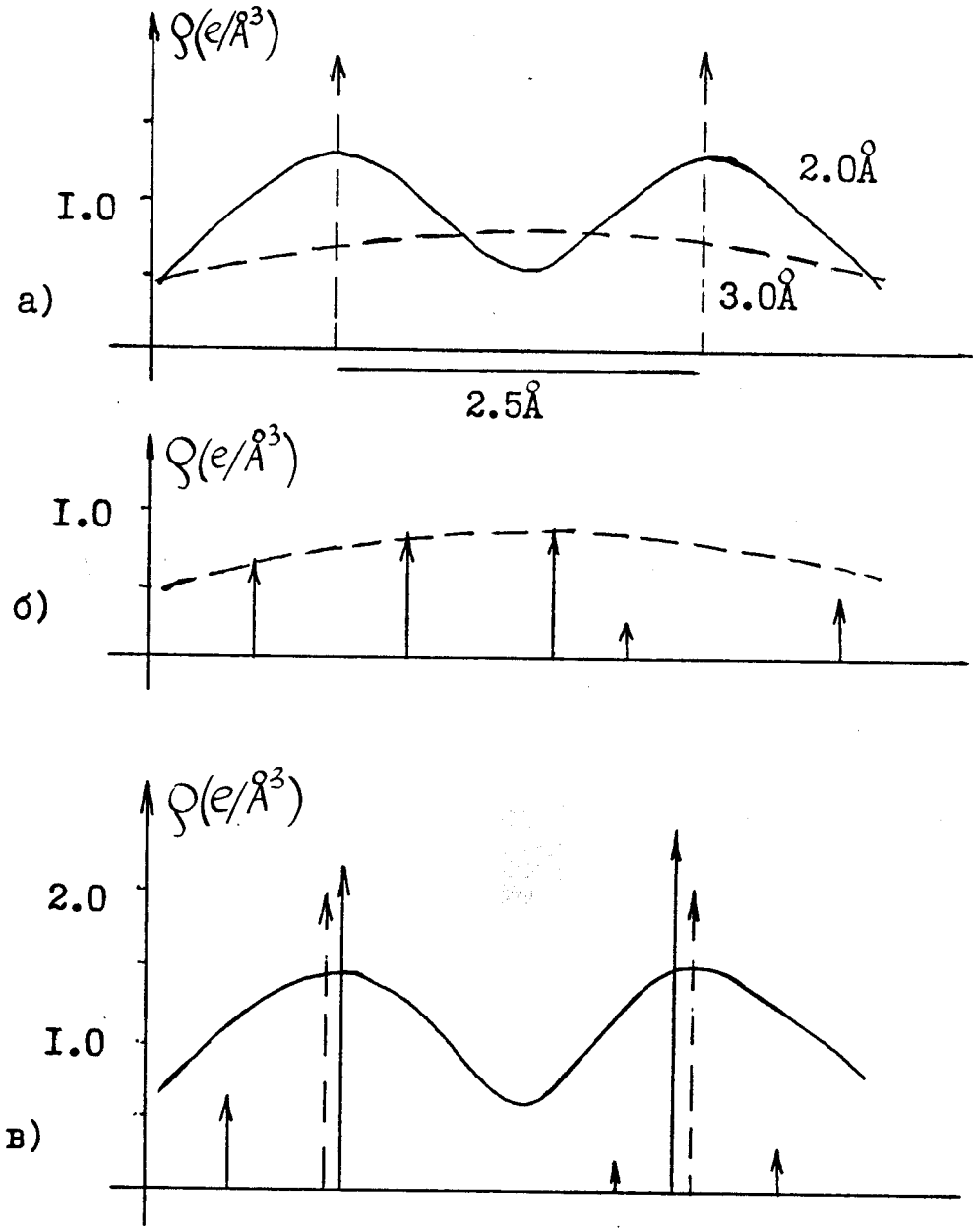

Рис. 4. Расщепление пика электронной плотности на синтезе среднего разрешения.

а) Изображение двух атомов азота, расположенных на расстоянии 2.5 Å, на 2- и 3-ангстремных синтезах.

б) Исходная грубая модель, построенная по 3-ангстремному синтезу.

в) Модифицированная грубая модель и соответствующий 2-ангстремный синтез (модули структурных факторов - точные, фазы - рассчитаны по модели).

малой структуре, исходный трехангстремный синтез имел в заданной области один широкий пик, образованный изображениями двух соседних атомов азота. Построение и модификация грубой атомной модели, в которой исходно данный пик моделировался изображениями пяти фиктивных атомов типа азота, позволило расщепить пик на два со значениями центров и высотой, близкими к истинным.

Агарвал и Айзекс в предложенной ими методике рекомендовали накладывать некоторые стереохимические требования на фиктивные атомы модели при ее построении, в частности, невозможность нахождения пары атомов на расстоянии меньше заданного, невозможность соседства атома более чем с тремя другими и т.п. Мы считаем более правильным отказ от этих либо аналогичных требований потому, что только в этом случае появляется возможность при фиксированном типе атомов достаточно хорошо моделировать большие "вздутия" функции распределения электронной плотности.

В соответствии с принятым принципом, нами была предложена следующая схема построения грубой модели:

I) определение узла сетки $\vec{r}_*$, в котором достигается значение абсолютного максимума функции распределения электронной плотности $\rho(\vec{r})$; координаты $\vec{r}_j$ центра очередного, j-го фиктивного атома полагаются равными координатам этой точки;

2) определение температурного фактора $B_j \geq 0$ этого атома таким образом, чтобы значение функции распределения электронной плотности атома в его центре $\rho_f(\vec{0}, B_j)$ равнялось соответствующей величине $\rho(\vec{r}_*)$ синтеза в точке его максимума; при невозможности задания такого температур-

ного фактора (значение $\rho(\vec{r}_*)$ очень велико) полагается $B_j = 0$;

3) вычитание из синтеза $\rho(\vec{r})$ вклада определенного таким образом атома внутри сферы с центром в точке $\vec{r}_*$ и некоторым фиксированным радиусом d - радиусом атома;

4) возврат к первому пункту уже с новой функцией распределения электронной плотности, полученной в предыдущем пункте, до тех пор, пока максимальное значение на синтезе не станет меньше некоторой заданной величины ρ_{CRIT} .

При реализации данной схемы было обнаружено, что непосредственное ее использование требует большого количества машинных ресурсов, причем основное время тратится на многократный поиск точки абсолютного максимума. Нами был предложен следующий алгоритм, эквивалентный предыдущему по результату (построенной грубой модели), но существенно более быстрый [205].

Заметим, что если точка $\vec{r}_*$ является точкой локального максимума функции $\rho(\vec{r})$ и если в окрестности радиуса d точки $\vec{r}_*$ нет значений $\rho(\vec{r})$ выше, чем $\rho(\vec{r}_*)$, то при последовательной выборке глобальных максимумов функции $\rho(\vec{r})$ и вычитании вкладов соответствующих атомов значение функции в точке $\vec{r}_*$ меняться не будет до тех пор, пока точка $\vec{r}_*$ не станет точкой уже глобального максимума. Следовательно, если имеется такая точка локального максимума, то можно проинтерпретировать ее как центр атома и вычесть вклад этого атома сразу, а не дожидаться того, когда она станет точкой глобального максимума. Поэтому практический алгоритм построения грубой модели рекомендуется следующий:

I) просматриваются поочередно все узлы сетки, пока не встретится точка $\vec{r}_*$ такая, что $\rho(\vec{r}_*) > \rho(\vec{r})$ для

всех $|\vec{r}-\vec{r}_*|\leq d$;

2) определяется температурный фактор $B_j \geq 0$ фиктивного атома с центром в точке $\vec{r}_*$;
3) вклад построенного фиктивного атома вычитается из имеющейся функции $\rho(\vec{r})$;
4) выполняется просмотр следующих узлов сетки;
5) после просмотра всех узлов сетки осуществляется возврат к первому узлу, так как после вычитания вкладов атомов могли образоваться новые локальные максимумы; такие циклические просмотры повторяются до тех пор, пока максимальное значение функции $\rho(\vec{r})$ с вычтенными вкладами фиктивных атомов не станет меньше заданной величины.

2.2.3. Определение вклада отдельного атома.

Поскольку нашей целью является аппроксимация функцией распределения электронной плотности, рассчитанной по атомной модели, некоторого взвешенного синтеза конечного разрешения, то возникает вопрос об изображении на этом синтезе отдельного атома или, что то же самое, об определении вклада отдельного атома в такой синтез. Из-за ограниченности разрешения изображение атома будет расплываться, и появятся волны обрыва. Для целей построения грубой модели важен вопрос аппроксимации центрального пика изображения атома. Ясно, что вопрос сводится к рассмотрению атома с центром в начале координат.

Для атома с фактором рассеяния $f(|\vec{s}|)$ и изотропным температурным фактором B функция распределения электронной плотности имеет вид

$$\rho_f(\vec{r},B)=\frac{1}{V}\sum_{\vec{s}\in S_0} m(\vec{s})f(|\vec{s}|)e^{-\frac{B|\vec{s}|^2}{4}}e^{-2\pi i(\vec{s},\vec{r})}, \qquad (67)$$

где V - объем элементарной ячейки,

S_0 - набор структурных факторов, включенных в расчет анализируемого синтеза,

$m(\vec{s})$ - весовые коэффициенты ("показатели достоверности определения значений фаз"), использовавшиеся при расчете синтеза.

Распределение (67) можно переписать как

$$\rho_f(\vec{r},B)=\frac{1}{V}\sum_{\vec{s}\in\mathbb{R}'}\hat{m}(\vec{s})f(|\vec{s}|)e^{-\frac{B|\vec{s}|^2}{4}}e^{-2\pi i(\vec{s},\vec{r})}, \qquad (68)$$

где $\mathbb{R}'$ - кристаллическая решетка в обратном пространстве, а

$$\hat{m}(\vec{s}) = \begin{cases} m(\vec{s}), & \vec{s}\in S_0 \\ 0 & \text{для остальных } \vec{s} \end{cases} \qquad (69)$$

Тогда

$$\rho_f(\vec{r},B)=\left[\frac{1}{V}\sum_{\vec{s}\in\mathbb{R}'}\hat{m}(\vec{s})f(|\vec{s}|)e^{-2\pi i(\vec{s},\vec{r})}\right]*\left[\sum_{\vec{u}\in\mathbb{R}}\left(\frac{4\pi}{B}\right)^{3/2}e^{-\frac{4\pi^2|\vec{r}+\vec{u}|^2}{B}}\right], \qquad (70)$$

где $|\vec{r}|$ - расстояние от начала координат до точки с координатами $\vec{r}$,

$\mathbb{R}$ - кристаллическая решетка в прямом пространстве,

$*$ - операция свертки.

Учитывая малость поправки от слагаемых с ненулевыми $\vec{u}$, можно приближенно записать

$$\rho_f(\vec{r},B)\approx\left[\frac{1}{V}\sum_{\vec{s}\in S_0}m(\vec{s})f(|\vec{s}|)e^{-2\pi i(\vec{s},\vec{r})}\right]*\left[\left(\frac{4\pi}{B}\right)^{3/2}e^{-\frac{4\pi^2|\vec{r}|^2}{B}}\right] \qquad (71)$$

Функция, заключенная в первые квадратные скобки в правой части (71) соответствует изображению покоящегося атома с

$B = 0$ на исходном синтезе. Для фиксированного типа атомов с фактором рассеяния $f(|\vec{s}|)$ эта функция - обозначим ее $q_0(\vec{r})$ - определяется только весовой функцией $m(\vec{s})$ на наборе отражений S_0. Следовательно, ее можно рассчитать до начала построения модели. Аппроксимируем $q_0(\vec{r})$ в области ее центрального пика гауссовой функцией

$$q(\vec{r}) = C_0 \left(\frac{4\pi}{B_0}\right)^{3/2} e^{-\frac{4\pi^2|\vec{r}|^2}{B}}, \tag{72}$$

где C_0 и B_0 - константы для данного типа атомов и данного синтеза. При переходе к синтезу с другим набором отражений или с другой весовой схемой значения C_0 и B_0, вообще говоря, изменятся. В результате из (71) и (72) следует аппроксимация функции распределения электронной плотности атома на данном синтезе функцией

$$\hat{\rho}_f(r,B) = C_0 \left(\frac{4\pi}{B+B_0}\right)^{3/2} e^{-\frac{4\pi^2|\vec{r}|^2}{B+B_0}} \tag{73}$$

Аппроксимация (73), в которой известны C_0 и B_0, достаточно проста и в вычислительном плане нетрудоемка для построения грубой модели.

Рассмотрим более подробно вопрос определения констант C_0 и B_0. После логарифмирования функций $q(\vec{r})$ и $q_0(\vec{r})$ и введения параметров

$$\alpha = \ln\left[V C_0 \left(\frac{4\pi}{B_0}\right)^{3/2}\right]$$
$$\beta = -\frac{4\pi^2}{B_0} \tag{74}$$

задача сводится к линейной задаче нахождения параметров ап-

проксимации

$$\alpha+\beta|\vec{r}|^2 \approx q_\ell(\vec{r}) \tag{75}$$

где

$$q_\ell(\vec{r})=\ln\left[Vq_0(\vec{r})\right]=\ln\sum_{\vec{s}\in S_0} m(\vec{s})f(|\vec{s}|)e^{-2\pi i(\vec{s},\vec{r})} \tag{76}$$

Параметры α и β могут быть определены из условия минимума величины

$$Q(\alpha,\beta)=\sum_{\vec{r}:|\vec{r}|\le d}\left[q_\ell(\vec{r})-(\alpha+\beta|\vec{r}|^2)\right]^2 \Rightarrow min \tag{77}$$

где d - радиус центрального пика изображения атома (при среднем разрешении величина d - порядка I.5 - 2.0 Å).

Условие минимума по α и β выражения (77) может быть записано как

$$\begin{cases} \frac{1}{2}\frac{\partial Q(\alpha,\beta)}{\partial\alpha}=\sum_{|\vec{r}|\le d}\left[q_\ell(\vec{r})-(\alpha+\beta|\vec{r}|^2)\right]=0 \\ \frac{1}{2}\frac{\partial Q(\alpha,\beta)}{\partial\beta}=\sum_{|\vec{r}|\le d}\left[q_\ell(\vec{r})-(\alpha+\beta|\vec{r}|^2)\right]|\vec{r}|^2=0 \end{cases}, \tag{78}$$

то есть α и β есть решение системы уравнений

$$\begin{cases} \alpha+a_1\beta=b_1 \\ a_1\alpha+a_2\beta=b_2 \end{cases}, \tag{79}$$

где $N=\sum_{|\vec{r}|\le d} 1$

$$a_1=\frac{1}{N}\sum_{|\vec{r}|\le d}|\vec{r}|^2, \qquad b_1=\frac{1}{N}\sum_{|\vec{r}|\le d} q_\ell(\vec{r}),$$
$$a_2=\frac{1}{N}\sum_{|\vec{r}|\le d}|\vec{r}|^4, \qquad b_2=\frac{1}{N}\sum_{|\vec{r}|\le d} q_\ell(\vec{r})|\vec{r}|^2 \tag{80}$$

Окончательно,

$$\begin{cases} \alpha = \dfrac{a_1 b_2 - a_2 b_1}{a_1^2 - a_2} \\ \beta = \dfrac{a_1 b_1 - a_2}{a_1^2 - a_2} \end{cases} \tag{81}$$

и

$$B_0 = -\frac{4\pi^2}{\beta}$$
$$C_0 = e^{\alpha} \cdot \frac{1}{V} \cdot \left(-\frac{\pi}{\beta}\right)^{3/2} \tag{82}$$

Как уже отмечалось выше, конкретное значение B_j для j -го фиктивного атома определяется из равенства значений функции $\hat{\rho}_f(\vec{r},B)$ (73) в центре атома и $\rho(\vec{r})$ в точке локального максимума:

$$\rho(\vec{r}_*) = \hat{\rho}_f(\vec{0},B_j) = C_0 \left(\frac{4\pi}{B_0+B_j}\right)^{3/2} \tag{83}$$

При

$$B_j = 4\pi\left(\frac{C_0}{\rho(\vec{r}_*)}\right)^{2/3} - B_0 < 0 \tag{84}$$

величина B_j полагается равной нулю (имеющийся максимум функции не описывается функцией $\hat{\rho}_f(\vec{r},B)$ полностью), и возможно, вообще говоря, расположение еще одного или нескольких фиктивных атомов в той же точке $\vec{r}_*$.

2.2.4. Критерии качества построенной грубой модели.

Как уже указывалось в п.2.2.2, основным критерием качества соответствия грубой модели исходному синтезу является максимальное значение ρ^R_{MAX} остаточной функции $\rho^R(\vec{r})$, получен-

ной вычитанием из исходной функции распределения электронной плотности $\rho(\vec{r})$ вкладов $\rho_f(\vec{r}, B_j)$ всех атомов грубой модели. "Исчерпывание" достаточно высоких пиков приводит к тому, что ρ^R_{MAX} становится меньше некоторой заданной величины ρ_{CRIT}, являющейся параметром построения модели. Понижая ρ_{CRIT}, можно улучшать качество аппроксимации исходного синтеза за счет увеличения числа фиктивных атомов. Однако, при излишнем снижении ρ_{CRIT}, с одной стороны, начинает "моделироваться" шум на исходном синтезе, а с другой стороны, чрезмерное увеличение количества атомов грубой модели осложняет ее последующую модификацию. В практической работе мы обычно использовали значения

$$\rho_{CRIT} = \rho_{AV} + \lambda(\rho_{MAX} - \rho_{AV}), \tag{85}$$

где $\rho_{MAX} = \max\limits_{\vec{r}} \rho(\vec{r})$,

$$\rho_{AV} = \sum_{\vec{r}} \rho(\vec{r}) / M(\{\vec{r}\}),$$

$M(\{\vec{r}\})$ - количество узлов сетки,

$\lambda \sim$ 0.I - 0.3 в зависимости от обостренности синтеза и конкретной ситуации.

Такой выбор ρ_{CRIT} основывается на дополнительном критерии - количестве атомов грубой модели. Основным принципом здесь является приблизительное соответствие суммарной величины заряда грубой модели истинному значению. Поскольку грубая модель состоит обычно из фиктивных атомов типа азота, то их количество может несколько превосходить общее количество неводородных атомов структуры. Желательно еще некоторое уве-

личение числа фиктивных атомов потому, что при построении модели часть из них может оказаться "совсем не на месте", и такие атомы в дальнейшем будут удалены из модели.

Еще одним критерием качества построенной грубой модели может служить распределение значений температурного фактора фиктивных атомов. Установление температурного фактора большинства атомов близким к нулю может быть объяснено плохим качеством исходного синтеза (наличие сильных "размазанных" пиков, ошибки в построении синтеза, неправильное шкалирование и т.п.). Соответственно, построенная грубая модель вряд ли является пригодной для дальнейшей работы. Наличие же значительного количества атомов с чрезмерно большими значениями температурного фактора (порядка 100 $Å^2$ и выше) означает либо избыточность модели, вызванную заниженным значением ρ_{CRIT} , либо наличие ошибок при расчете синтеза.

2.3. Модификация грубой модели.

2.3.1. Критерии модификации.

Поскольку исходная грубая модель может лишь воспроизвести имеющееся распределение электронной плотности, то фазы структурных факторов, рассчитанных по ней, по качеству, вообще говоря, не превосходят фаз, использованных для построения синтеза. Для улучшения значений фаз и расширения их набора необходимо провести модификацию грубой модели.

Пусть имеется набор экспериментально определенных модулей структурных факторов $F_o(\vec{s})$, $\vec{s} \in S_o$, про которые мы будем предполагать, что их значения известны точно. Под модификацией модели мы будем понимать такое изменение параметров q имеющейся модели, при котором минимизируется расхождение

модельных величин с экспериментом:

$$R\left(\{F_o(\vec{s})\},\{F_c(\vec{q};\vec{s})\}\right)\Rightarrow \min_{\vec{q}} \qquad (86)$$

В качестве функционала R обычно используется один из квадратичных критериев [199]

$$R_{12}=\sum_{\vec{s}\in S_o} w_1(\vec{s})\left[F_c(\vec{q};\vec{s})-F_o(\vec{s})\right]^2 \qquad (87)$$

либо [228]

$$R_{22}=\sum_{\vec{s}\in S_o} w_2(\vec{s})\left[F_c^2(\vec{q};\vec{s})-F_o^2(\vec{s})\right]^2 \qquad (88)$$

Агарвалом и Айзексом при проверке метода использовалась программа [199], минимизирующая критерий (87).

Луниным и Уржумцевым [53] был предложен более общий критерий соответствия модельных данных рентгеновскому эксперименту

$$R_x=\sum_{\vec{s}\in S_o}\Big\{w_x(\vec{s})\left[F_c^2(\vec{q};\vec{s})-F_o^2(\vec{s})\right]^2-$$
$$-\left[A_{\vec{s}}\cos\varphi_c(\vec{q};\vec{s})+B_{\vec{s}}\sin\varphi_c(\vec{q};\vec{s})+C_{\vec{s}}\cos 2\varphi_c(\vec{q};\vec{s})+D_{\vec{s}}\sin 2\varphi_c(\vec{q};\vec{s})\right]\Big\} \qquad (89)$$

Здесь $A_{\vec{s}}, B_{\vec{s}}, C_{\vec{s}}, D_{\vec{s}}$ - параметры полученного каким-либо образом распределения вероятностей для фазы, записанного в унифицированном виде (58) (для структурных факторов с неизвестными значениями фаз полагается $A_{\vec{s}}=B_{\vec{s}}=C_{\vec{s}}=D_{\vec{s}}=0$). Использование критерия R_x может, по-видимому, дать лучший результат, чем использование R_{12} либо R_{22} , потому, что позволяет включить в работу дополнительную информацию, например, данные об экспериментах с изоморфными производными (путем соответствующего задания величин $\{A_{\vec{s}}, B_{\vec{s}}, C_{\vec{s}}, D_{\vec{s}}\}$).

Агарвал и Айзекс не обсуждали принципы проведения моди-

фикации. Нами при практической реализации метода были обнаружены некоторые моменты, потребовавшие особой тщательности в разработке и заслуживающие отдельного разбора.

2.3.2. Общая стратегия модификации.

Модификация грубой модели начинается, как правило, с предварительного уточнения температурного фактора. Необходимость данного этапа следует из того, что в программах уточнения, в частности - в [199], обычно, как более точная, используется двухгауссовая аппроксимация фактора рассеяния атома, а нами в программе построения грубой модели для быстроты вычислений и удобства расчетов используется аналогичная одногауссовая аппроксимация. Следовательно, возникает некоторое несоответствие между функциями распределения электронной плотности, рассчитанными по одной и той же модели, но с несколько отличающимися факторами рассеяния атомов. Предварительное уточнение значений температурного фактора снимает это несоответствие и улучшает качество воспроизведения моделью исходного синтеза.

Выбор набора структурных факторов, по которому ведется модификация, может существенно влиять на результат улучшения набора фаз. Нижняя граница зоны разрешения должна выбираться так, чтобы исключить влияние растворителя на значения экспериментально определенных модулей. В качестве такой границы может быть использована, например, величина d_{MAX} = 10Å [229]. Верхняя граница зоны разрешения постепенно повышается в ходе модификации таким образом, чтобы на первых циклах модификации использовались лишь те отражения, по которым строился исходный синтез, а на последних циклах - все отражения, для которых мы желаем получить улучшенный набор фаз. Начальные

циклы модификации по отражениям, включенным в расчет исходного синтеза, нам кажутся необходимыми для улучшения соответствия грубой модели исходному синтезу. При построении грубой модели фиктивные атомы располагались лишь в узлах сетки, что могло несколько ухудшить качество воспроизведения исходного синтеза.

Включение в последние циклы модификации полного набора модулей является необходимым потому, что, как будет показано ниже в главе 3, фазы рассчитанных структурных факторов, не включенных ранее в критерий модификации, "в среднем" равномерно распределены вокруг точных значений. Следовательно, исключение отражения из критерия не дает возможности определить для него значение фазы даже приближенно. Плавный переход от исходного к полному набору данных мы считаем оправданным из тех соображений, что резкое изменение функционала, зависящего от большого числа переменных, может нарушить сходимость процесса минимизации. Отметим еще, что включение в критерий отдельных отражений из зон разрешения с малым процентом собранных данных вряд ли является целесообразным.

Выбор параметров модели, для которых одновременно производится модификация их значений, существенно зависит от возможностей используемой программы уточнения. Возможны ограничения как на тип параметров, так и на количество однотипных параметров, уточняемых одновременно. Так, использовавшаяся Агарвалом и Айзексом и нами программа [199] позволяет проводить уточнение координат атомов и температурного фактора лишь поочередно.

При очень большом количестве фиктивных атомов, когда одновременная модификация их параметров становится невозмож-

ной либо из-за ограничений программы, либо из-за ограниченности машинных ресурсов, либо из-за каких-то вычислительных соображений, возможна разбивка модели на компоненты и поочередная их модификация. Крайне желательной является разделенность в пространстве этих компонентов между собой. Игнорирование этого требования может привести к "размазыванию" фиктивных атомов в областях пересечения компонентов (за счет вытеснения атомов из максимумов функции распределения электронной плотности атомами другого компонента) и к ненужному увеличению числа параметров, усложняющему сходимость процесса уточнения.

2.3.3. Автоматические корректировки грубой модели.

Основной причиной необходимости выполнения корректировок грубой модели является аномальное поведение величины температурного фактора B_j ряда фиктивных атомов.

В случае, когда какой-либо атом оказывается "совсем не на месте", программа уточнения пытается удалить его, "размазав" соответствующую ему функцию распределения электронной плотности. Этому соответствует приписывание атому очень больших значений температурного фактора B_j - порядка 50 - 100 $Å^2$. Таким образом, если значение B_j температурного фактора атома превышает некоторую величину B_{LIM}, то имеет смысл удалить данный атом из модели. Удаление атомов с $B_j > B_{LIM}$ нами было названо "срезкой модели". Общая тенденция в выборе B_{LIM} состоит в постепенном уменьшении этой величины от модели к модели с тем, чтобы суммарный заряд последней модели был приблизительно равен заряду элементарной ячейки. "Постепенность" понимается в смысле приблизительного равенства количества удаляемых атомов за срезку. Для опреде-

ления значений B_{LIM} полезно после первых циклов модификации температурного фактора собрать статистику его распределения. Как показывает опыт, значения B_j стабилизируются довольно быстро, и несложно составить план изменения B_{LIM} для последующих моделей. Приемлемыми являются значения B_{LIM} порядка 20 - 50 $Å^2$.

При выходе значений температурного фактора фиктивных атомов на нижнюю границу $B=0$ области допустимых значений возможно два подхода к корректировке модели. Во-первых, можно предупреждать выход B_j на границу, для атомов с $B_j<B_{CRIT}$ принудительно устанавливая $B_j=B_{CRIT}$, где B_{CRIT} - некоторое небольшое положительное число порядка единицы. Во-вторых, можно изменить тип атома. Поскольку попытка установить значение B_j отрицательным означает, что данный фиктивный атом должен моделировать сильный максимум электронной плотности, но его заряда не хватает для такого моделирования, то увеличение заряда (замена фиктивного атома на атом другого типа с большим зарядом) может привести к успеху. Такой подход был опробован нами, но применения не получил потому, что замена атома азота на атом близкого типа (кислород) эффекта практически не дает, а замена на существенно более тяжелый атом, как уже отмечалось, нецелесообразна.

2.3.4. Контроль процесса модификации модели.

Основным средством контроля процесса модификации является характер поведения значений минимизируемого критерия (86). Типичным является резкий спад значений на первых циклах модификации и медленное снижение критерия дальше. При добавлении в критерий новых отражений величина его обычно повышается, но не очень сильно. Сильное повышение означает

слишком резкое изменение критерия, которое может привести к отрицательным последствиям при минимизации.

Аналогичным образом меняются значения кристаллографического фактора достоверности

$$R = \sum_{\vec{s}}' \left| F_c(\vec{s}) - F_o(\vec{s}) \right| / \sum' F_o(\vec{s}) \qquad (90)$$

Отметим, что при модификации грубой модели важны изменения R-фактора, а не его абсолютное значение, которое, как правило, бывает ниже значений, получающихся при обычном кристаллографическом уточнении атомных моделей макромолекул. Этот факт объясняется существенно большим числом степеней свободы у грубой модели как при ее построении, так и при последующей модификации.

Еще одним важным средством контроля выполнения модификации является средняя величина сдвигов фиктивных атомов за цикл. Как показывает практика, нормальное поведение заключается в средней подвижке атомов на 0.4 - 0.6 Å на первых циклах и постепенном уменьшении сдвига до величин порядка 0.01 Å. Установление таких сдвигов означает, что возможности модификации по данной зоне разрешения практически исчерпаны. Включение в критерий дополнительных отражений обычно увеличивает средний сдвиг до 0.1 - 0.3 Å за цикл с его последующим уменьшением.

В ряде ситуаций приходится искусственно ограничивать сдвиги атомов потому, что достаточно большие сдвиги могут привести к "разваливанию" модели, что, в свою очередь, ставит под сомнение успех всей работы.

2.4. Частичная реконструкция грубой модели.

2.4.I. Способы использования модифицированной модели.

После того, как завершена модификация грубой модели, возникает вопрос использования полученной информации. Агарвал и Айзекс предложили рассчитать по модифицированной модели структурные факторы $\{F_c(\vec{s})e^{i\varphi_c(\vec{s})}\}_{\vec{s}}$ и построить новый синтез, вообще говоря, более высокого разрешения, чем исходный, с коэффициентами $F_o(\vec{s})e^{i\varphi_c(\vec{s})}$. В случае, когда вновь полученные карты распределения электронной плотности оказывались все еще недостаточно высокого качества, предлагалось повторить процедуру, взяв новый синтез в качестве исходного.

Как показали наши эксперименты (глава 3), значительно более сильный эффект дальнейшего улучшения фаз дает частичная реконструкция модифицированной модели с сохранением наиболее достоверной ее части и заменой остальных фиктивных атомов на новые.

Рассматривая процесс модификации грубой модели как процесс минимизации некоторого функционала, частичную реконструкцию модели можно трактовать как средство борьбы с "застреванием" в локальных минимумах функционала. С этой точки зрения, предложение Агарвала и Айзекса строить грубую модель заново есть выбор новой исходной точки для поиска минимума с отбрасыванием значительной части ранее полученной информации. В связи со сделанным замечанием мы считаем, что частичная реконструкция обязательно должна использоваться в ходе модификации грубой модели, а не подключаться только как вспомогательное средство в случае не очень удачного завершения работы. Более того, при значительном продвижении по разрешению частичная реконструкция должна использоваться не

только на полном наборе данных, но и на промежуточных наборах.

2.4.2. Частичная реконструкция грубой модели с помощью разностного синтеза.

Критерием для отбора наиболее достоверно расположенных фиктивных атомов модифицированной модели M_R являются значения температурного фактора. Выполняя "срезку" модели с соответственно выбранным значением B_{LIM}, можно отобрать наиболее достоверную часть модели - от I/3 до 2/3 общего числа атомов. Обозначим эту неполную модель через M_P, а рассчитанные по ней структурные факторы через $F_P(\vec{S})e^{i\varphi_P(\vec{S})}$. Построим новую грубую модель путем добавления к атомам модели M_P новых фиктивных атомов, которые определим путем построения грубой модели по синтезу с коэффициентами

$$m_c(\vec{S})\left[F_0(\vec{S})e^{i\varphi_c(\vec{S})} - kF_P(\vec{S})e^{i\varphi_P(\vec{S})}\right], \qquad (9I)$$

то есть по улучшенному синтезу, из которого вычтен вклад уже имеющейся части грубой модели. Здесь $k \geqslant 1$ - некоторый коэффициент, а $m_c(\vec{S})$ и $\varphi_c(\vec{S})$ определены из (66) с $P_{\vec{S}}(\varphi) = P^c_{\vec{S}}(\varphi)$, где $P^c_{\vec{S}}(\varphi)$ - наиболее правильное распределение вероятностей фаз, обычно получаемое как результат комбинирования исходной фазовой информации и информации по модифицированной модели.

Отметим, что, по-прежнему, синтез должен быть построен в абсолютной шкале.

Существенно использование коэффициента k. Наша основная цель на данном этапе - автоматическая "интерпретация" тех участков синтеза, где еще отсутствует грубая модель. Поскольку модель M_P имеет суммарный заряд меньше, чем исследуемая структура, то при значении $k=1$ на уже "проинтерпре-

тированных" участках может сохраниться остаточная электронная плотность, которая будет мешать правильному дополнению модели M_P . Из этих рассуждений становится ясной возможная оценка k :

$$k = \arg\min_{k} \sum_{\vec{s}} \left(F_0(\vec{s}) - kF_P(\vec{s})\right)^2 \qquad (92)$$

Построение дополнительной части M_A модели, состоящей из заново определяемых фиктивных атомов, ведется с помощью рассмотренного выше алгоритма построения грубой модели. Коэффициенты C_0 и B_0 аппроксимации (73) должны быть определены заново, так как для разностного синтеза используется новая весовая функция и, как правило, новый, более полный набор отражений. Количество фиктивных атомов в модели M_A оценивается разностью количества атомов в моделях M_R и M_P .

Дальнейшая работа с грубой моделью, полученной объединением фиктивных атомов моделей M_P и M_A , ведется рассмотренным выше (п. 2.3) образом.

2.5. Построение заключительного синтеза.

Заключительный синтез электронной плотности является результатом всей работы и решающим критерием ее успеха или неудачи. Поэтому сформулируем здесь наши рекомендации по вопросу построения заключительного синтеза, хотя большая часть их уже была рассмотрена в предыдущих параграфах.

Так как синтез, построенный по структурным факторам, взвешенным показателями достоверности определения фаз, имеет наименьшую среднеквадратичную ошибку среди других синтезов, то общий вид коэффициентов заключительного синтеза должен быть следующим:

$$m_F(\vec{s})F_0(\vec{s})e^{i\varphi_F(\vec{s})}, \tag{93}$$

где

$$m_F(\vec{s})e^{i\varphi_F(\vec{s})} = \int_0^{2\pi} e^{i\varphi} P^F_{\vec{s}}(\varphi)\,d\varphi \tag{94}$$

Заключительное распределение $P^F_{\vec{s}}(\varphi)$ вероятностей фаз на первых итерациях метода построения и модификации грубой атомной модели должно определяться путем комбинирования информации по модели со стартовой информацией, чтобы избежать появления различного рода артефактов. На более поздних этапах работы комбинирование становится уже нецелесообразным, т.к. определяемая по модели фазовая информация становится значительно более достоверной, чем исходная.

Для отражений малоугловой зоны, не включавшихся в минимизируемый критерий во избежание неучтенного влияния растворителя, фазовая информация, рассчитанная по модели, не используется. Фазы для этих отражений должны определяться по исходно имевшейся информации. Например, в качестве фаз следует брать значения, полученные методом изоморфного замещения.

Выбор разрешения синтеза определяется верхней границей зоны разрешения, по которой проводилась модификация. Включение в синтез отражений из зон с малым процентом собранных данных заметного положительного эффекта не дает. Возможен, скорее всего, обратный эффект.

2.6. Анализ качества полученных данных.

2.6.I. Интерпретируемость синтеза.

Основная цель всей выполняемой работы по улучшению набора фаз - это, в конечном счете, повышение интерпретируемости исходного синтеза электронной плотности. Поэтому сугубо субъективное решение исследователя об улучшении либо неулучшении качества карт распределения электронной плотности является решающим критерием. В настоящее время какие-либо количественные (следовательно, объективные) характеристики именно интерпретируемости синтеза отсутствуют, хотя можно использовать ряд вспомогательных критериев.

2.6.2. Обострение синтеза.

Одной из количественных характеристик синтеза является его обостренность

$$S = \rho_{MAX} / \rho_{AV} , \tag{95}$$

где

$$\rho_{MAX} = \max_{\vec{r}} \rho(\vec{r}) ,$$
$$\rho_{AV} = \sum_{\vec{r}} \rho(\vec{r}) / M(\{\vec{r}\}) , \tag{96}$$

либо само значение ρ_{MAX} , поскольку величина $\rho_{AV} \approx F_{000}/V$ практически не изменяется при построении синтезов в одинаковой шкале. Имеет смысл сравнение величин ρ_{MAX} лишь для невзвешенных синтезов, так как на величину ρ_{MAX} взвешенного синтеза существенно влияет выбор весовой функции.

По мере улучшения набора фаз величины S и ρ_{MAX} должны возрастать. Особенно резкое увеличение наблюдается при реальном повышении разрешения синтеза.

2.6.3. Уровень зашумленности.

Близким к описанному выше критерию является контроль уровня зашумленности синтеза. В качестве характеристики уровня зашумленности можно использовать величину

$$N = \rho_{MAX} / \left(\frac{1}{2}\rho_{MOL}\right) , \qquad (97)$$

где ρ_{MOL} - значение функции распределения электронной плотности, соответствующее нижней линии уровня при разрисовке карт распределения электронной плотности. Данный критерий, естественно, зависит от субъективного решения исследователя в выборе ρ_{MOL}. Однако, учитывая определенную стандартность таких решений при подготовке карт для одного и того же объекта и реально существующие границы выбора ρ_{MOL}, использование данного критерия в качестве вспомогательного представляется полезным.

2.6.4. Отклонение значений фаз от исходных величин.

Общая идея данного критерия состоит в том, что, если исходное значение фазы является достаточно надежным, то отклонение его от нового, уточненного значения должно быть небольшим. Отклонение фазы φ от величины φ_0 мы будем характеризовать величиной

$$\Delta\varphi = \min\left\{|\varphi-\varphi_0| ; 360-|\varphi-\varphi_0|\right\} \qquad (98)$$

при условии $0^\circ \leq \varphi < 360^\circ$, $0^\circ \leq \varphi_0 < 360^\circ$. Понятно, что имеют смысл лишь характеристики, усредненные по некоторому набору отражений. Критерием надежности исходного значения фазы может служить величина $m(\vec{S})$ показателя достоверности ее определения.

Рассмотрим поведение функции

$$\delta(m) = \langle \Delta\varphi(\vec{s}) \rangle_{\vec{s}:\, m(\vec{s})=m} \qquad (99)$$

в трех идеализированных ситуациях.

I) Фазы $\{\varphi_c(\vec{s})\}$, рассчитанные по модели, являются точными. Тогда среднее расхождение их со стартовыми значениями является убывающей по m функцией, $\delta(0) \approx 90^\circ$, $\delta(1) \approx 0^\circ$.

2) $\{\varphi_c(\vec{s})\}$ – случайный набор фаз. Тогда $\delta(m)$ не зависит от m, $\delta(m) \approx 90^\circ$.

3) $\{\varphi_c(\vec{s})\}$ совпадает со стартовым набором. Тогда $\delta(m) \approx 0^\circ$. Следовательно, анализ функции $\delta(m)$, построенной по конкретным данным, позволяет дать некоторую характеристику полученного набора фаз. Заметим, что данный критерий может использоваться скорее как способ обнаружения ошибок, чем как способ оценки точности значений.

Предложенный контроль качества фаз возможен лишь при правильности оценок показателей достоверности исходного набора. Поэтому, в частности, когда исходный набор получен методом изоморфного замещения, рекомендуется для оценки значений использовать отражения среднего разрешения (3 – 5 Å), принадлежащие центросимметричным зонам, так как показатель достоверности для них определяется наиболее правильно. Естественно, в случае недостаточного количества этих данных приходится включать в оценку и другие отражения.

2.6.5. Оценка средних ошибок.

Зная распределение вероятностей $P^F_\zeta(\varphi)$ для полученного набора фаз, можно оценить среднюю погрешность определения фазы величиной

$$\tilde{\Delta}(\zeta) = \int_0^{360} \Delta\varphi \cdot P^F_\zeta(\varphi)\, d\varphi \qquad (100)$$

где $\Delta\varphi$ определена из (98) с $\varphi_0 = \varphi_F(\vec{s})$, $\varphi_F(\vec{s})$ - значение фазы, использованное для построения заключительного синтеза (93).

Для некоторого шарового слоя в обратном пространстве $S_K^1 = \{\vec{s}: s_1^K \leq |\vec{s}| < s_2^K\}$ можно оценить среднюю ошибку значений фаз величиной

$$\bar{\Delta}_K = \sum_{\vec{s} \in S_K^1}{}' \tilde{\Delta}(\vec{s}) / M(S_K^1) \qquad (101)$$

Сравнение $\bar{\Delta}_K$ с аналогичными величинами, определенными для исходного и других наборов фаз, дает представление о сравнительной надежности этих наборов.

Заметим, что наиболее сложным моментом здесь является правильная оценка параметров распределений вероятностей (п. I.8).

2.6.6. Анализ показателей достоверности.

Зная распределение вероятностей фаз для исходного, заключительного и, возможно, еще для каких-нибудь интересующих нас наборов фаз, можно для всех таких наборов оценить значения показателей достоверности определения значений фаз и построить функцию

$$\bar{m}(s) = \langle m(\vec{s}) \rangle_{\vec{s}: |\vec{s}| = s} \qquad (102)$$

Сравнительный анализ функций $\bar{m}(s)$ также дает возможность выбора наилучшего набора фаз. Подчеркнем, что, в принципе, в разных зонах разрешения в качестве наилучших могут быть выбраны различные наборы фаз, исходя из значений $\bar{m}(s)$ и $\bar{\Delta}_K(s)$.

2.7. Использование дополнительной информации.

2.7.1. Основные типы используемой информации.

В отличие от рассматриваемой на протяжении данной главы упрощенной ситуации, при практических исследованиях макромолекул часто имеется возможность использования дополнительной информации, специфичной для исследуемой структуры. Наиболее простыми примерами являются

1) известная локализация молекулы,
2) наличие локальной симметрии,
3) наличие проинтерпретированной части структуры (частичной модели).

Последняя ситуация, когда часть структуры уже проинтерпретирована, дает возможность применения метода построения и модификации смешанной атомной модели и будет отдельно разобрана ниже в п. 2.8.

2.7.2. Использование данных о локализации молекулы.

Наиболее простым подходом, позволяющим использовать сведения о локализации молекулы, является требование расположения фиктивных атомов грубой модели внутри заданной области. На этапе построения грубой модели это требование может быть реализовано, например, путем зануления функции распределения электронной плотности вне границ молекулы. Правда, тем самым мы не учитываем наличия молекул воды вне области молекулы, однако при работе на среднем разрешении это не вносит большой ошибки. Естественно, что при определении границ молекулы последние нужно проводить с некоторым "запасом"; только в таком случае можно ожидать заметного улучшения синтеза вблизи поверхности молекулы, где обычно и присутствуют наиболее сложные для интерпретации участки.

На этапе модификации грубой модели знание границ молекулы может использоваться в качестве средства контроля. С одной стороны, выход каких-либо атомов с достаточно малыми значениями температурного фактора за границу области означает, скорее всего, неточность проведения границы на данном участке. С другой стороны, при достаточно надежном проведении границ факт выхода атома из области может трактоваться как ошибочность расположения данного атома, и атом может быть удален из модели.

Анализ уровня шума вне границ молекулы на заключительном синтезе является очевидным способом оценки качества этого синтеза.

Более тонким подходом использования сведений о локализации молекулы является учет влияния растворителя на рассеяние рентгеновских лучей. Зная область, занятую исследуемой молекулой, исходя из соответствия рассчитанных модулей структурных факторов экспериментально определенным величинам можно подобрать оптимальное значение электронной плотности растворителя, считая ее постоянной вне области молекулы [229]. По определенной таким образом функции распределения электронной плотности растворителя вычисляются соответствующие структурные факторы $F_S(\vec{S})e^{i\varphi_S(\vec{S})}$, которые должны учитываться при расчете различных используемых синтезов. Наиболее значительная поправка путем учета растворителя вносится в структурные факторы низкого разрешения - от 7 - 10 Å и ниже.

Зная структурные факторы $F_S(\vec{S})e^{i\varphi_S(\vec{S})}$, можно на этапе модификации грубой модели включить с соответствующей поправкой данные по отражениям малоугловой зоны в минимизируемый

критерий. В результате такой модификации при построении синтезов может быть использована новая, уточненная фазовая информация и для отражений низкого разрешения.

2.7.3. Использование локальной симметрии.

Наличие локальной симметрии в структуре является источником существенной дополнительной информации, которая может быть использована на большинстве этапов работы по улучшению наборов фаз данным методом.

При построении грубой модели локальная симметрия может быть использована в двух, в некотором смысле противоположных ситуациях.

В ситуации, когда участки синтеза, связанные локальной симметрией, достаточно идентичны, можно строить грубую модель лишь для одного такого участка (например, предварительно занулив функцию распределения электронной плотности на другом участке), а потом размножить эту модель согласно локальной симметрии. Тем самым, кроме внесения дополнительной информации, достигается и значительная экономия машинного времени.

Противоположной является ситуация, когда один из участков, связанных локальной симметрией, проявляется существенно лучше другого участка (например, в силу влияния каких-либо низкочастотных гармоник). В такой ситуации непосредственное построение грубой модели во всей независимой области элементарной ячейки приведет к еще большей диспропорции между этими участками, так как на участке, где молекула проявлена слабее, будет расположено меньшее число фиктивных атомов. Предлагается в такой ситуации занулить функцию $\rho(\vec{r})$ на втором участке, а построение грубой модели вести по лучшему

участку, размножив затем модель по локальной симметрии.

Заметим, что довольно часто участки синтеза, связанные локальной симметрией, соприкасаются друг с другом. В этом случае более эффективным будет сознательное небольшое расширение этих участков. В случае возникновения двойного набора фиктивных атомов на каком-либо (небольшом) участке в зависимости от дальнейшей модификации либо атомы одного из наборов будут "размазаны", либо для атомов обоих наборов установится "половинный" температурный фактор.

При наличии соответствующего программного обеспечения локальная симметрия может быть использована и на этапе модификации грубой модели. Требование строгого выполнения соответствующих уравнений локальной симметрии, связывающих попарно атомы грубой модели, накладывает дополнительное ограничения на возможные наборы фаз и позволяет, кроме того, добиться заметной экономии машинного времени.

В ситуации, когда наличие локальной симметрии по каким-либо причинам не использовалось ни при построении, ни при модификации грубой модели, симметрия может рассматриваться как средство контроля. В данном случае, когда информация о локальной симметрии в грубую модель не закладывалась, аналогичные изменения функции распределения электронной плотности на симметрично связанных участках свидетельствуют, скорее всего, о достоверности этих изменений, чем об их случайности.

2.8. Использование смешанной атомной модели.

2.8.I. Общее описание методики.

В ситуации, когда метод использования грубой модели уже

не позволяет существенно улучшить имеющиеся карты распределения электронной плотности, а частичная модель еще недостаточно полна для ее обычного кристаллографического уточнения, нами было предложено использовать данные о частичной модели для дальнейшего улучшения фаз. Общая схема действий в данной ситуации близка к схеме использования грубой модели, однако ряд методических особенностей и другая область использования дают возможность выделения такого подхода в качестве отдельной методики. Заметим, что в ситуации, когда частичная модель мала, эти методики сближаются. В противоположной ситуации, когда частичная модель становится достаточно полной, метод использования смешанной атомной модели переходит в обычное кристаллографическое уточнение.

Общий алгоритм метода использования смешанной атомной модели следующий:

1) построение (корректировка, расширение) частичной модели;
2) построение смешанной модели;
3) модификация смешанной модели;
4) расчет новой функции распределения электронной плотности с использованием фазовой информации, полученной по модифицированной смешанной модели.

Этап построения и корректировки частичной модели здесь обсуждаться не будет. Остальные этапы, так же, как и общая стратегия, будут рассмотрены более подробно.

2.8.2. Построение смешанной модели.

Смешанная атомная модель строится путем объединения атомов частичной модели и фиктивных атомов грубой модели. Атомы грубой модели должны располагаться на тех участках

синтеза, где отсутствует частичная модель. Существование же фиктивных атомов вблизи атомов частичной модели нежелательно по двум причинам.

Во-первых, возникает "конкуренция" атомов частичной и грубой моделей за максимумы функции распределения электронной плотности. Наиболее ожидаемым исходом будет ликвидация атомов частичной модели, так как они расставляются вручную с учетом геометрических требований и обычно не попадают в точки максимумов $\rho(\vec{r})$ в отличие от атомов грубой модели, которые в дальнейшем и будут моделировать эти максимумы.

Во-вторых, расположение фиктивных атомов на уже проинтерпретированных участках синтеза при условии приблизительного равенства суммарных зарядов модели и структуры означает нехватку этих фиктивных атомов на остальных участках, куда эти атомы и предназначались.

Следовательно, при неаккуратном построении грубой модели не будут улучшаться ни частичная модель, ни функция распределения электронной плотности как на проинтерпретированных, так и на остальных участках. Для предупреждения возникновения данной ситуации следует тщательно удалить вклад атомов частичной модели из имеющегося синтеза перед построением грубой модели.

Первым шагом здесь является предварительное уточнение температурного фактора атомов частичной модели. Этот шаг приводит к более точному соответствию функций распределения электронной плотности $\rho(\vec{r})$ и $\rho_P(\vec{r})$ - исходного синтеза и частичной модели - в местах существования атомов частичной модели. В качестве исходного значения температурного фактора может быть взята величина B_{AV} среднего температурного фак-

тора, определяемая из графика Вильсона. Вторым способом определения температурного фактора атомов может быть его определение исходя из значений функции распределения электронной плотности в точках элементарной ячейки, соответствующих центрам атомов - в этом случае уточнение температурного фактора может не проводиться.

Следующим шагом является построение синтеза с коэффициентами

$$m(\vec{s})\left[F_o(\vec{s})e^{i\varphi_o(\vec{s})} - kF_{PB}(\vec{s})e^{i\varphi_{PB}(\vec{s})}\right], \qquad (103)$$

где $m(\vec{s})$ и $\varphi_o(\vec{s})$ определены согласно (66) из наилучшей доступной функции распределения вероятностей фаз,

$F_o(\vec{s})$ - экспериментально определенные значения модулей структурных факторов,

$F_{PB}(\vec{s})e^{i\varphi_{PB}(\vec{s})}$ - структурные факторы, рассчитанные по частичной модели с предварительно уточненными значениями температурного фактора,

$k \geq 1$ - шкальный коэффициент.

Здесь коэффициент k служит той же цели более полного вычитания функции распределения электронной плотности модели, что и ранее в п. 2.4. Величина коэффициента может быть взята равной величине оптимального шкального коэффициента для F_{PB}, полученной при предварительном уточнении значений температурного фактора атомов частичной модели.

Окончательный синтез для построения грубой модели получается из описанного выше синтеза с коэффициентами (103) занулением функции распределения электронной плотности внутри окрестностей заданного радиуса R_{AT} центров атомов частичной

модели. Этот этап является дополнительной мерой для отделения атомов частичной модели от фиктивных атомов. Значение радиуса атома R_{AT} определяется радиусом сходимости программы уточнения и степенью достоверности частичной модели. При нескольких итерациях метода использования смешанной атомной модели значение R_{AT} может несколько увеличиваться, например, с 0.7 Å до 1.2 - 1.5 Å.

2.8.3. Модификация смешанной модели.

С момента дополнения частичной модели грубой различие между ними до некоторой степени стирается. Смешанная модель служит для воспроизведения распределения электронной плотности и для улучшения его соответствия экспериментальным данным, и с этой точки зрения все атомы смешанной модели равноправны. В частности, должны модифицироваться параметры атомов одновременно обоих компонентов, а не поочередно атомов грубой и частичной моделей. Поочередная модификация, более того, может привести к разрушению частичной модели, так как при таком подходе фиктивные атомы смогут "захватить" максимумы функции распределения электронной плотности.

Тем не менее, различие между двумя компонентами смешанной модели не должно игнорироваться полностью. Атомы частичной модели, как определенные относительно достоверно, в принципе, не должны сильно уходить от исходных позиций. Для этого необходимо ограничивать максимальный сдвиг этих атомов при модификации. На более поздних итерациях использования метода, по-видимому, полезным является наложение на расположение атомов частичной модели различных стереохимических ограничений и контроль контактов типа "атом - атом" и "атом-фиктивный атом". Взаимодействие фиктивных атомов между со-

бой, по-прежнему, учитываться не должно.

В методе использования грубой атомной модели мы предлагали в первую очередь проводить модификацию температурного фактора. Здесь мы рекомендуем, наоборот, вначале модифицировать координаты атомов, чтобы атомы частичной модели преимущественно сразу же сдвинулись к максимумам функции распределения электронной плотности.

В остальном процесс модификации смешанной атомной модели совпадает с процессом модификации грубой атомной модели.

2.8.4. Корректировки смешанной модели.

При корректировках модели, особенно на поздних этапах, по-видимому, нецелесообразно удалять атомы частичной модели даже при установлении для них достаточно больших значений температурного фактора. Вообще говоря, при достаточно корректно построенной частичной модели и правильных построении и модификации смешанной модели такая ситуация практически не должна встречаться. При ее возникновении следует при последующем анализе частичной модели обратить внимание на атомы с большими значениями температурного фактора.

Важным в методическом плане является вопрос использования частичной реконструкции смешанной модели. В данном методе, в противоположность рекомендациям для метода использования грубой модели, мы считаем не нужным выполнение этой процедуры. Здесь наиболее достоверной частью структуры следует считать уже не атомы с наименьшими значениями температурного фактора, а, в первую очередь, атомы частичной модели. Как правило, даже самая плохая частичная модель содержит около половины общего числа неводородных атомов структуры (атомы главной цепи). Следовательно, при перестроении смешанной

модели грубая модель должна заменяться полностью или практически полностью. В такой ситуации более правильным будет предварительно провести корректировку и дополнение частичной модели, то есть, фактически, начать новую итерацию с использованием новой частичной модели.

2.8.5. Использование модифицированной частичной модели.

В методе использования грубой атомной модели исходным принципом является отказ от попытки установления поатомного соответствия фиктивных атомов атомам структуры, поэтому единственным результатом выполненной работы является улучшенный набор фаз. В методике, использующей смешанную модель, возникает вопрос об использовании такой модифицированной модели.

При корректном построении и модификации модели фиктивные атомы не должны нарушать частичную модель. Напротив, они играют некоторую стабилизирующую роль при модификации и не позволяют атомам частичной модели уходить далеко от исходных позиций. Таким образом, атомы модифицированной частичной модели по меньшей мере демонстрируют тенденцию движения того или иного атома по направлению к правильному положению.

Важен поатомный анализ значений температурного фактора B_j. Наличие у атома частичной модели низких либо высоких значений B_j свидетельствует о большей либо меньшей надежности его расположения. Одновременное увеличение B_j и вывод атома из максимума функции распределения электронной плотности свидетельствуют о некорректности определения этого атома.

2.9. Программное обеспечение.

Основное программное обеспечение разрабатывалось в НИВЦ АН СССР на вычислительных машинах серии ЕС ЭВМ в процессе постановки и использования описанных методов. Некоторые из программ являются адаптациями зарубежных вариантов. Постановке таких программ сопутствовали обнаружение и исправление ряда ошибок, в том числе и достаточно серьезных.

В данной разработке программного обеспечения, кроме автора, принимали также участие В.Ю.Лунин и Е.А.Вернослова.

Имеющиеся программы могут быть условно разделены, согласно общей схеме методов, на следующие группы.

1) Программы для построения синтезов электронной плотности - программа быстрого преобразования Фурье [230], созданная Тен Эйком [231], и программа зануления синтеза внутри большого набора шаров.
2) Программы для построения грубой модели - программа определения параметров изображения атома на синтезе электронной плотности и программа собственно построения грубой модели.
3) Программа уточнения атомной модели по методике Агарвала [199].
4) Программы для обработки фазовой информации. Наиболее крупной программой здесь является фазовая программа [232], созданная Тен Эйком. Программа была усовершенствована нами для возможности определения коэффициентов $A_{\vec{s}}, B_{\vec{s}}, C_{\vec{s}}, D_{\vec{s}}$ унифицированного распределения вероятностей фаз (58), определяемого по данным изоморфного замещения и аномального рассеяния. К этой группе относятся также программа определения параметров распределения вероятностей фаз из

принципа максимального правдоподобия, программа для комбинирования фазовой информации, программа для определения показателей достоверности и "наилучших" фаз.

5) Программа визуализации результатов - программа для изображения на графопостроителях ЕС-7052 и ЕС-7054 карт распределения электронной плотности и (или) скелетной модели структуры в виде отдельных сечений либо в виде стереоизображений.

Кроме указанных основных программ, имеется также значительное количество относительно небольших программ для проверки и подготовки данных, корректировок моделей, контроля, сбора различных статистических характеристик и т.п.

Более подробная информация о наиболее крупных программах представлена в таблице 3.

Большая часть программ являются универсальными по отношению к пространственным группам симметрии. Остальные программы являются специфическими для использовавшейся нами пространственной группы $P2_12_12_1$, более или менее просто модифицируемыми при переходе к другой пространственной группе.

Программное обеспечение оформлено в виде пакета программ с возможностью его дальнейшего расширения. Некоторые из программ, в частности, программы уточнения и визуализации, могут использоваться вне применения данной методики.

Таблица 3. Характеристики основных используемых программ.

программа	объем текста (тыс.опер.языка FORTRAN)	объем выполняемого модуля, Кб	характерное время работы на ЕС-I040,мин (белок средних размеров)
фазовая программа (Тен Эйк)	2.5	I00	I00/цикл
программа быстрого преобразования Фурье (Тен Эйк)	5.0	I00-300	I0
программа зануления синтеза	0.3	200	20
программа определения параметров изображения атома	0.2	I00	3
программа построения грубой модели	0.6	700	I0
программа уточнения (Агарвал)	4.2	500	40/цикл
программа оценки параметров распред. вероятн. фаз	0.2	I00	2
программа комбинир. фазов. информации	0.2	I00	2
программа определения "наилучш." фазы и показ. достоверн.	0.2	I00	2
программа рисования	4.0	400	5

ГЛАВА 3
ПРОВЕРКА МЕТОДА НА МОДЕЛЬНЫХ ДАННЫХ

3.I. Общая организация численных экспериментов.

Разработка методики использования грубой модели и проверка эффективности метода проводились на данных по структуре белка актинидина. Необходимые для выполнения численных экспериментов данные были взяты из Банка белковых молекул [9]. Выбор конкретной структуры для эксперимента определялся лишь заранее фиксированной пространственной группой симметрий и средними характерными размерами белка. Какая-либо специфика данного белка при работе не использовалась.

Кристаллы актинидина принадлежат пространственной группе $P2_12_12_1$, параметры элементарной ячейки a=78.2 Å, b=81.8 Å, c =33.05 Å. В независимой части элементарной ячейки содержится одна молекула (218 аминокислотных остатков). Молекула состоит из 1653 атомов: 1039 атомов углерода, 268 - азота, 337 - кислорода, 9 - серы.

Значения координат атомов актинидина, взятые из Банка белковых молекул, рассматривались как точные значения. Всем атомам было приписано значение температурного фактора B =20 Å^2. По полученной таким образом модели M_0 был рассчитан набор структурных факторов $\{F_0(\vec{s})e^{i\varphi_0(\vec{s})}\}$ разрешением до 2.0 Å, состоящий из 15009 отражений.

В соответствии с поставленным вопросом об эффективности метода грубой атомной модели для расширения набора фаз и для уточнения значений фаз было проведено два численных эксперимента.

Первый эксперимент заключался в определении наиболее

эффективной методики для расширения набора фаз среднего разрешения. Набор $\{F_0(\vec{s})\}$ модулей структурных факторов разрешением $d=1/|\vec{s}| \geqslant 2$ Å использовался в качестве набора экспериментально определенных данных. Набор фаз $\{\varphi_0(\vec{s})\}$ структурных факторов разрешением $d \geqslant 3$ Å моделировал исходный набор фаз, определенных каким-либо методом. Считалось, что исходные данные - значения модулей и фаз - известны абсолютно точно. Фазы структурных факторов разрешением 3 Å $> d \geqslant$ 2 Å считались неизвестными, и требовалось определить их значения. В качестве контрольных использовались значения фаз $\varphi_0(\vec{s})$, 3 Å $> d = 1/|\vec{s}| \geqslant 2$ Å.

Во втором эксперименте анализировались возможности метода для снижения ошибок имеющихся значений фаз. В значения фаз $\varphi_0(\vec{s})$ исходного трехангстремного набора структурных факторов были внесены случайные ошибки, одинаково распределенные, с параметрами распределения, зависящими от "силы" отражения и от разрешения $d=1/|\vec{s}|$. Требовалось, исходя из этих значений и значений модулей $F_0(\vec{s})$, уточнить значения фаз, используя только данные разрешением до 3 Å. По-прежнему, контроль уточнения проводился по точным значениям фаз $\varphi_0(\vec{s})$.

Подробное описание численных экспериментов и полученных результатов приведено ниже.

3.2. Численный эксперимент по расширению набора фаз.

3.2.I. Постановка задачи.

Исходными данными для эксперимента явились структурные факторы $F_0(\vec{s})e^{i\varphi_0(\vec{s})}$, рассчитанные по известным значениям параметров атомов модели M_0 белка актинидина. Значения моду-

лей $F_0(\vec{s})$ структурных факторов разрешением до 2.0 Å считались известными точно. Набор Φ_0 точных значений фаз $\varphi_0(\vec{s})$ был разделен на две части. Фазы 464I отражения разрешением до 3.0 Å считались известными точно и равными $\varphi_0(\vec{s})$, а фазы остальных I0368 отражений разрешением 3.0Å $> d \geq$ 2.0Å могли использоваться лишь в качестве контрольных величин. Требовалось без привлечения дополнительной информации о структуре определить считавшиеся неизвестными значения фаз структурных факторов разрешением 3.0 Å $> d \geq$ 2.0 Å.

3.2.2. Построение грубой модели.

Для построения грубой модели предварительно по имеющимся данным была рассчитана функция распределения электронной плотности разрешением 3.0 Å. Функция была рассчитана на сетке с числом делений II2×I20×48, то есть шаг сетки, приблизительно равный I/3×2Å, соответствовал требуемым рекомендациям для последующей работы с 2-ангстремными данными. Максимальное значение ρ_{MAX} функции распределения электронной плотности, рассчитанной в абсолютной шкале, равнялось 2.I7 e/Å^3.

Построение грубой модели велось по всему объему независимой части элементарной ячейки. Параметры изображения атома (73) равнялись B_0 =7I.4 Å^2, C_0 =I2.I e. Минимальное значение функции для построения атома равнялось ρ_{CRIT} =0.30 e/Å^3.

Построенная исходная грубая модель M_I состояла из 2I68 фиктивных атомов со средним значением температурного фактора B_{AV} =2I.7 Å^2. По модели M_I был рассчитан набор структурных факторов $\{F_1(\vec{s})e^{i\varphi_1(\vec{s})}\}_{1/|\vec{s}| \geq 2Å}$. Для набора фаз $\Phi_I = \{\varphi_1(\vec{s})\}$ среднее отклонение значений от соответствующих $\varphi_0(\vec{s})$ составило 57° по всему набору и 22° по исходному набору отражений, включенных в расчет синтеза. Здесь и в дальнейшем, кро-

ме специально оговоренных случаев, все статистические данные приводятся по наборам отражений, не принадлежащих центросимметричным зонам.

3.2.3. Модификация грубой модели.

Модификация параметров модели M_I проводилась последовательно по пяти различным зонам с разрешением, постепенно повышавшимся от 3.0 до 2.0 Å. Различные характеристики модификации приведены в таблице 4. Заметим, что обычно в каждой зоне выполнялось по два цикла модификации температурного фактора (B), а затем по 3 цикла модификации координат (X). Дополнительные 5 циклов (2B+3X) модификации по исходному набору структурных факторов предназначались для корректировки построения грубой модели.

Полученная в результате модификации модель M_7 состояла из 2022 фиктивных атомов. Набор фаз Φ_7 рассчитанных по M_7 структурных факторов $F_7(\vec{s})e^{i\varphi_7(\vec{s})}$ имел среднюю ошибку 42° по полному набору отражений и 17° и 50° соответственно по наборам $d \geqslant 3.0$ Å и 3.0 Å $> d \geqslant 2.0$ Å.

3.2.4. Построение и модификация новой грубой модели.

Согласно рекомендациям Агарвала и Айзекса [6], для дальнейшего улучшения фаз был рассчитан новый синтез электронной плотности более высокого разрешения, нежели исходный, с коэффициентами

$$F(\vec{s})e^{i\varphi(\vec{s})} = \begin{cases} F_0(\vec{s})e^{i\varphi_0(\vec{s})}, & 1/|\vec{s}| \geqslant 3.0 \text{ Å} \\ F_0(\vec{s})e^{i\varphi_7(\vec{s})}, & 3.0 \text{ Å} > 1/|\vec{s}| \geqslant 2.5 \text{ Å} \end{cases} \qquad (104)$$

Максимальное значение ρ_{MAX} равнялось 2.54 e/Å^3.

По этому синтезу была сгенерирована новая модель M_8,

состоявшая из 2182 фиктивных атомов с параметрами изображения B_o =42.3 Å^2 и C_o =II.I е и средней величиной температурного фактора B_{AV} =33.5 Å^2. Критическое значение ρ_{CRIT} равнялось 0.32 е/Å^3.

С моделью M_8 была повторена процедура модификации по расширяющемуся набору структурных факторов. Модифицированная модель M_{I2} состояла из I868 фиктивных атомов. Сравнение соответствующих моделям M_8 и M_{I2} наборов фаз Φ_8 и Φ_{I2} с наборами Φ_I и Φ_7 показало неэффективность подхода, предложенного Агарвалом и Айзексом. Несмотря на повышение разрешения синтеза для построения модели, погрешности фаз обоих полученных наборов Φ_7 и Φ_{I2} были практически одинаковыми (таблица 5).

3.2.5. Частичная реконструкция грубой модели с помощью разностного синтеза.

Нам удалось добиться дальнейшего улучшения фаз, используя частичную реконструкцию грубой модели с помощью разностного синтеза.

По модели M_{I2} были определены фиктивные атомы (II40 штук) со значениями температурного фактора, не превосходящими B_{LIM} =30 Å^2. Координаты этих атомов, образовавших модель M_{I3}, уточнялись по данным разрешением до 2.5 Å. Затем был рассчитан синтез электронной плотности с коэффициентами

$$\begin{cases} F_0(\vec{s})e^{i\varphi_0(\vec{s})} - F_{14}(\vec{s})e^{i\varphi_{14}(\vec{s})}, & 1/|\vec{s}| \geq 3.0 \text{ Å} \\ F_0(\vec{s})e^{i\varphi_{12}(\vec{s})} - F_{14}(\vec{s})e^{i\varphi_{14}(\vec{s})}, & 3.0 \text{ Å} > 1/|\vec{s}| \geq 2.5 \text{ Å} \end{cases} \quad (105)$$

где $F_{14}(\vec{s})e^{i\varphi_{14}(\vec{s})}$ – структурные факторы, рассчитанные по модифицированной модели M_{I4}. Величина ρ_{MAX} составила 2.32 е/Å^3.

Таблица 4. План модификации грубой модели актинидина (расширение набора фаз).

модифицируемая модель*)	к-во атомов	зона модификации, d_{MIN}(Å)**)	уровень срезки, Å^2	R-фактор, %		функционал R_{12} (×I0^{-7})	
				исх.	фин.	исх.	фин.
M_I	2I68	3.0	-	29.5	I2.9	3I.9	6.3
M_2	2I68	3.0	80	I2.9	9.4	6.3	3.2
M_3	2I42	2.7	70	I5.5	I2.I	9.I	4.6
M_4	2043	2.5	70	I6.2	I3.3	8.8	5.5
M_5	2033	2.2	70	2I.0	I6.4	I3.7	7.9
M_6	2026	2.0	70	2I.4	I8.2	I3.2	9.5
M_7	2022	-	-	I8.2	-	9.5	-
M_8	2I82	3.0	-	26.0	20.7	23.4	I3.I
M_9	2I82	2.5	70	2I.9	I5.5	I6.0	7.8
M_{I0}	I878	2.2	70	22.5	I7.I	I5.5	8.9
M_{II}	I875	2.0	70	22.0	I9.0	I4.2	I0.4
M_{I2}	I868	-	-	I9.0	-	I0.4	-
M_{I5}	2933	3.0	-	I7.I	7.8	8.8	I.9
M_{I6}	2933	2.5	70	I4.7	I0.I	7.I	3.I
M_{I7}	2872	2.2	50	I6.I	I3.6	8.2	5.4
M_{I8}	2560	2.0	40	I8.I	I6.6	9.5	7.9
M_{I9}	2295	2.0	35	I6.6	I6.8	7.9	8.4
M_{20}	2080	-	-	I6.8	-	8.4	-

Таблица 4. (Продолжение).

M_{2I}	2205	2.0	-	22.6	I6.5	I7.4	8.5
M_{22}	2205	2.0	50	I6.5	I6.9	8.5	8.8
M_{23}	I782	2.0	45	I6.9	I5.4	8.8	7.3
M_{24}	I68I	-	-	I5.4	-	7.3	-

*) Для каждой из моделей проводилось по 5 циклов модификации параметров (2 цикла модификации температурного фактора и 3 цикла модификации координат).

**) Количество отражений в различных зонах следующее:

d_{MIN} = 3.0 - 464I

d_{MIN} = 2.7 - 6294

d_{MIN} = 2.5 - 7874

d_{MIN} = 2.2 - II353

d_{MIN} = 2.0 - I5009

Таблица 5. Зависимость средней фазовой ошибки (в градусах, для нецентросимметричных отражений) от $|\vec{S}|^2$ для наборов фаз, рассчитанных по грубым атомным моделям актинидина.

набор фаз	границы интервалов по $|\vec{S}|^2$										
	.000–.025	.025–.050	.050–.075	.075–.100	.100–.125	.125–.150	.150–.175	.175–.200	.200–.225	.225–.250	.000–.250
Φ_1	21	20	19	27	47	63	68	70	72	76	57
Φ_2	17	15	16	23	43	59	67	70	70	75	54
Φ_3	16	15	15	23	43	59	64	70	69	74	53
Φ_4	18	15	16	24	40	55	62	65	66	71	51
Φ_5	17	14	15	23	37	51	59	64	63	69	49
Φ_6	17	14	15	22	34	46	55	58	58	63	45
Φ_7	17	14	14	22	31	43	51	53	53	58	42
Φ_8	26	24	24	32	42	52	63	66	70	71	54
Φ_9	21	19	19	28	39	52	61	66	70	72	52
Φ_{10}	21	19	19	26	35	47	55	61	64	68	48
Φ_{11}	21	19	18	24	33	44	52	56	59	63	45
Φ_{12}	21	18	18	24	32	41	49	51	55	59	42

Таблица 5. (Продолжение).

Φ_{15}	16	16	16	22	33	47	55	58	59	63	45
Φ_{16}	14	13	14	21	33	45	53	57	59	64	44
Φ_{17}	14	13	14	21	30	42	50	53	53	60	41
Φ_{18}	15	14	15	20	29	38	45	49	49	55	38
Φ_{19}	16	15	16	21	27	35	42	45	44	51	36
Φ_{20}	17	14	17	22	27	35	40	43	43	50	35
Φ_{21}	21	17	16	23	29	37	43	46	46	51	37
Φ_{22}	17	14	12	18	24	33	36	40	42	47	33
Φ_{23}	18	14	14	19	24	32	35	38	40	46	32
Φ_{24}	17	14	13	19	24	29	34	37	39	44	31
кол-во отра-жений	320	681	957	1078	1278	1475	1502	1691	1783	1984	12713

Построение новой грубой модели M_{I5} выполнялось путем дополнения модели M_{I4} фиктивными атомами, определяемыми по разностному синтезу (I05). Значения ρ_{crit} и параметров B_0 и C_0 составляли соответственно 0.40 е/Å^3, 42.8 Å^2 и I2.0 е. Качество фаз $\varphi_{15}(\vec{s})$ набора структурных факторов, соответствующих M_{I5}, подтвердило эффективность данного подхода.

Модификация модели M_{I5} позволила снизить ошибку определения фаз с 45° до 35° для модели M_{20} по полному набору нецентросимметричных отражений.

Ввиду достигнутого эффекта описанный прием частичной реконструкции грубой модели по разностному синтезу был применен также и к модели M_{20}. Из 2080 фиктивных атомов модели M_{20} I067 со значениями температурного фактора, большими 20 Å^2, были заменены на II92 фиктивных атома, определенных по соответствующему разностному синтезу разрешением 2.0 Å. Построенная таким образом модель M_{2I} хорошо воспроизводила исходный набор фаз (средняя ошибка по полному набору 37°). Модификация M_{2I} позволила довести отклонение фаз $\varphi_{24}(\vec{s})$ итогового набора фаз Φ_{24} от фаз $\varphi_0(\vec{s})$ до 3I° по полному набору и 35° по отражениям разрешением 3.0 Å $> d \geq$ 2.0 Å.

3.2.6. Результаты эксперимента.

Результаты сравнения синтезов (рис. 5), построенных по соответствующим наборам данных, аналогично как и высокая точность определения неизвестных значений фаз, подтвердили эффективность предложенной методики.

Данные, приведенные в таблицах 5 и 6, подробно демонстрируют динамику улучшения значений фаз.

Обращают на себя внимание следующие факты.

Во-первых, интересной и не имеющей на настоящий момент

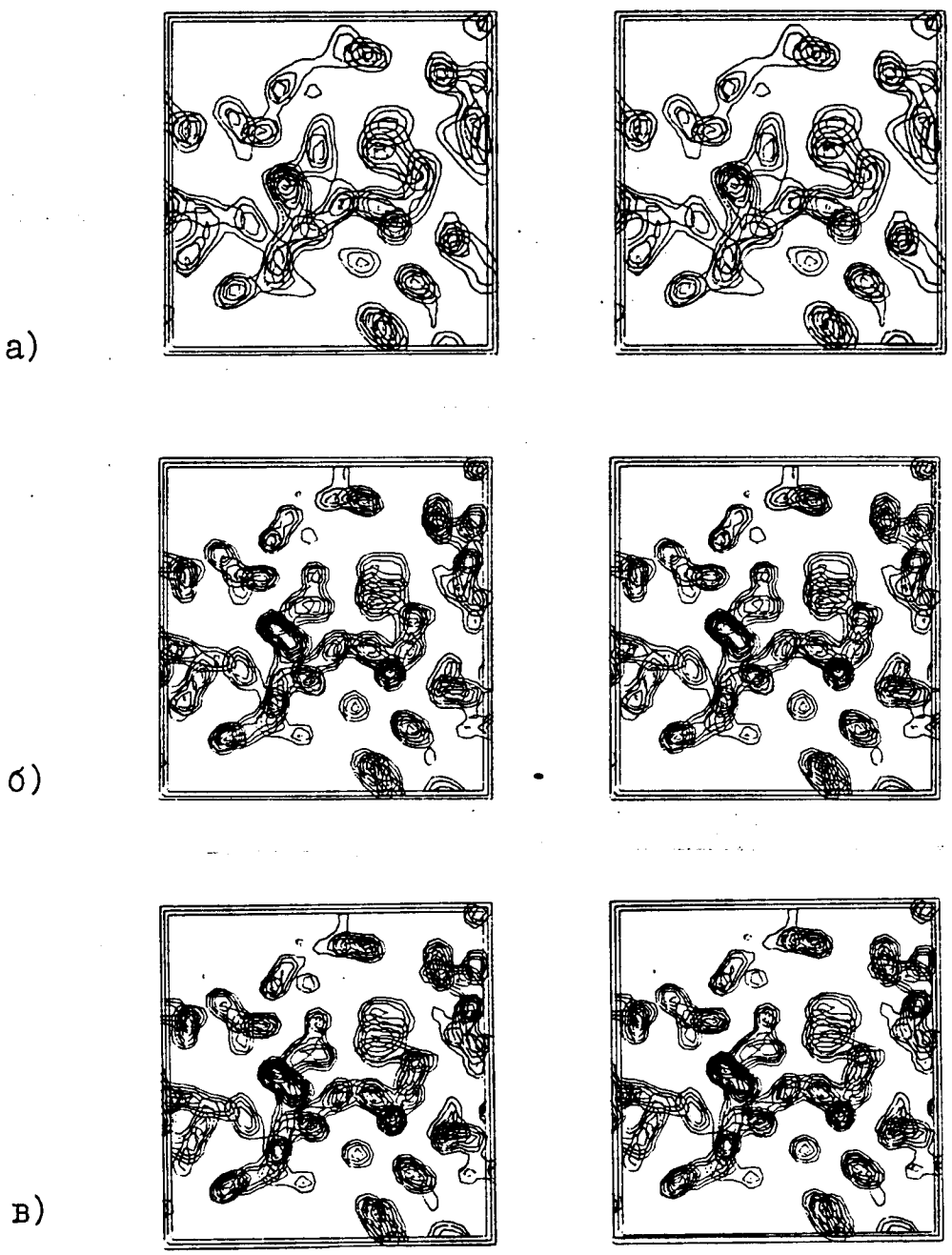

Рис. 5. Стереоизображение участка $0.50 \leq x \leq 0.82$, $0.64 \leq y \leq 0.96$, $0.00 \leq z \leq 0.07$ функции распределения электронной плотности в актинидине:

а) точный синтез, разрешение 3.0 Å;

б) синтез с фазами, рассчитанными по модифицированной грубой модели M_{24}, разрешение 2.0 Å;

в) точный синтез, разрешение 2.0 Å.

Таблица 6. Зависимость средней фазовой ошибки (в градусах, для нецентросимметричных отражений) от относительной силы отражения $t = 2F_oF_c/\sigma_N^2$ для наборов фаз, рассчитанных по грубым атомным моделям актинидина.

набор фаз	границы интервалов по t							
	0.00–0.50	0.50–0.75	0.75–I.00	I.00–I.50	I.50–2.00	2.00–5.00	5.00–I0.00	I0.00–∞
Φ_I	66	6I	59	54	49	43	30	28
Φ_2	66	56	55	48	43	32	I9	2
Φ_3	65	59	52	47	4I	3I	I5	2
Φ_4	64	55	5I	43	38	28	I3	3
Φ_5	63	52	46	40	32	23	9	3
Φ_6	6I	45	4I	3I	25	I7	7	3
Φ_7	59	4I	34	27	20	I4	6	3
Φ_8	67	58	53	47	4I	32	I9	6
Φ_9	65	57	50	46	39	30	2I	4
Φ_{I0}	63	5I	46	4I	32	23	I2	2
Φ_{II}	6I	47	39	32	25	I7	6	I
Φ_{I2}	59	43	34	27	2I	I3	6	I

Таблица 6. (Продолжение).

Φ_{15}	62	48	41	36	27	19	9	3
Φ_{16}	61	46	41	32	26	17	6	1
Φ_{17}	58	42	35	27	22	15	6	1
Φ_{18}	56	37	32	24	19	13	6	3
Φ_{19}	54	35	27	21	17	11	5	3
Φ_{20}	54	34	27	21	17	11	7	2
Φ_{21}	54	35	26	22	16	11	7	5
Φ_{22}	49	30	23	19	14	10	6	3
Φ_{23}	50	30	25	20	15	10	7	1
Φ_{24}	49	28	23	18	14	9	6	1
среднее кол-во отражений	5000	1800	1400	1800	1000	1300	1000	10

строгого объяснения является чувствительность фазовой ошибки к зоне модификации грубой модели. Параметры фиктивных атомов модифицированной модели каким-то образом "запоминают", отражения какого разрешения включались в минимизируемый критерий, и для этих отражений средние ошибки определяемых по модели фаз меньше, чем для оставшихся отражений.

Во-вторых, характер зависимости фазовой ошибки от относительной "силы" отражения $t(\vec{s})=2F_c(\vec{s})F_o(\vec{s})/\sigma_N^2(\vec{s})$ хорошо соответствует теоретически полученным соотношениям (62).

3.3. Численный эксперимент по уточнению значений фаз.

3.3.I. Постановка задачи.

Эксперимент по уточнению зашумленного набора фаз должен был дать ответ на вопрос о возможности использования метода в реальной ситуации, когда исходный набор данных содержит ошибки.

Исходный зашумленный набор фаз был смоделирован следующим образом. В точные значения фаз $\varphi_o(\vec{s})$ отражений исходного набора (разрешением до 3.0 Å), рассчитанных по точной атомной модели M_0 актинидина, были внесены случайные ошибки $\delta(\vec{s})$. Ошибки вносились таким образом, чтобы средняя ошибка $\langle\delta(\vec{s})\rangle$ возрастала с увеличением $|\vec{s}|=2\sin\theta/\lambda=1/d$ и убывала с ростом модуля структурного фактора. Более точно, распределение вероятностей задавалось функцией

$$P_{\vec{s}}(\delta)\sim exp\{t(|\vec{s}|)F_o^2(\vec{s})\cos\delta\}\ , \qquad (106)$$

где $t(s)$, параметр, регулирующий среднее значение ошибки, задавался невозрастающей ступенчатой функцией по s . При выбранных нами значениях $\{t(s)\}$ средняя ошибка $\bar{\delta}=\langle\delta(\vec{s})\rangle_{\vec{s}}$ по

всем нецентросимметричным отражениям разрешением до 3.0 Å составила 37°.

Требовалось, используя значения модулей $F_0(\vec{s})$ структурных факторов разрешением до 3.0 Å, уменьшить ошибку исходного набора фаз Φ_N.

3.3.2. Построение и модификация грубой модели.

Общая схема построения и модификации грубой модели практически совпадала со схемой предыдущего эксперимента за тем лишь исключением, что вся работа велась при разрешении 3.0 Å.

Исходная модель M_{101}, построенная по синтезу электронной плотности с коэффициентами $F_0(\vec{s})e^{i\varphi_N(\vec{s})}$, $1/|\vec{s}| \geqslant$ 3.0 Å, состояла из 2025 фиктивных атомов. Последующая модификация параметров фиктивных атомов модели (всего 12 циклов модификации значений температурного фактора и 18 циклов модификации координат) с дважды выполненной частичной реконструкцией модели по разностным синтезам позволила снизить фазовую ошибку до 21°. Подробно ход эксперимента и изменение фазовой ошибки отражены в таблицах 7 и 8.

3.3.3. Результат эксперимента.

Данный эксперимент показал, что улучшение фаз по описываемой методике сводит ошибки определения фаз до такого же уровня, который достигается в процессе модификации модели, построенной по идеальному, точному синтезу. Таким образом, наличие ошибок в исходных значениях фаз не является препятствием к расширению набора фаз путем использования грубой модели. Предельно допустимая величина исходных ошибок зависит от конкретной ситуации и, в частности, от используемой дополнительной информации. Применение данной методики позво-

Таблица 7. План модификации грубой модели актинидина (уточнение зашумленных значений фаз).

модифицируемая модель*)	кол-во атомов	зона модификации, d_{MIN} (Å) **)	уровень срезки, Å^2	R-фактор, %	R_{12}, функционал ($\times 10^{-7}$)
M_{I01}	2025	3.0	-	32.4	39.5
M_{I02}	2025	3.0	50	I4.8	7.6
M_{I03}	I859	-	-	II.4	4.0
M_{I05}	2469	3.0	-	3I.2	30.2
M_{I06}	2469	3.0	-	I2.6	5.7
M_{I07}	2469	-	-	8.2	2.2
M_{I09}	2040	3.0	-	27.3	22.8
M_{II0}	2040	3.0	-	II.4	4.0
M_{III}	2040	-	-	9.0	2.5

*) Для каждой модели выполнялось 5 циклов модификации (2 цикла модификации значений температурного фактора и 3 цикла модификации координат).

**) В минимизируемый критерий включались данные по 464I отражению разрешением $d \geq d_{MIN}$ = 3.0 Å.

Таблица 8. Зависимость средней фазовой ошибки (в градусах, для нецентросимметричных отражений) от $|\vec{S}|^2$ для наборов фаз, рассчитанных по грубым атомным моделям актинидина.

набор фаз	границы интервалов по $|\vec{S}|^2$				
	.000–.028	.028–.055	.055–.083	.083–.III	.000–.III
Φ_N	24	28	36	47	37
Φ_{I0I}	27	26	3I	43	34
Φ_{I02}	24	24	28	39	3I
Φ_{I03}	25	24	27	39	3I
Φ_{I05}	24	23	28	4I	3I
Φ_{I06}	I9	I6	20	29	22
Φ_{I07}	I6	I4	I9	29	2I
Φ_{I09}	2I	2I	26	37	28
Φ_{II0}	I9	I6	I9	3I	23
Φ_{III}	I7	I5	20	3I	22
кол-во отражений	388	8I6	II05	I308	36I7

ляет не только определить ранее отсутствовавшие значения фаз, но и уменьшить ошибку в имеющихся значениях.

ГЛАВА 4

УЛУЧШЕНИЕ НАБОРА ФАЗ ДЛЯ γ-КРИСТАЛЛИНА IIIb С ИСПОЛЬЗОВАНИЕМ ГРУБОЙ АТОМНОЙ МОДЕЛИ

4.1. Общая характеристика γ-кристаллина IIIb .

Кристаллины - основные структурные белки глаз позвоночных. Кристаллы γ-кристаллина IIIb из глазной линзы теленка принадлежат пространственной группе симметрий $P2_12_12_1$ и имеют размеры элементарной ячейки a=58.7 Å, b=69.5 Å, c=116.9 Å [10]. В независимой части элементарной ячейки находятся две молекулы общей массой около 40 КДа. Полипептидная цепь молекулы содержит около 170 остатков. Имеются предварительные данные по аминокислотной последовательности этого белка [233], причем полностью отсутствуют данные по участку цепи 37 - 55, и возможны ошибки в идентификации других остатков. Известно также, что в γ-кристаллине IIIb отсутствуют α-спиральные участки.

Кристаллы нативного белка γ-кристаллина IIIb имели дифракционное поле приблизительно до 2.3 Å (всего измерена интенсивность 12510 отражений), при этом процент собранных данных более высокого разрешения, начиная приблизительно с 2.7 Å, был достаточно мал.

Методом изоморфного замещения были определены значения фаз 6777 отражений разрешением до 3.0 Å. Использовались 5

тяжелоатомных производных, из них 3 - с аномальным рассеянием. Средний показатель достоверности определения фаз равнялся 0.72 [10]. На соответствующих картах распределения электронной плотности были хорошо различимы обе молекулы в асимметричной части элементарной ячейки, в каждой из молекул выделялось два домена. Функции распределения электронной плотности для обеих молекул были близкими, и детально анализировалась только одна из них. Фрагмент карты распределения электронной плотности приведен на рис. 6.

Ход полипептидной цепи был определен в областях вытянутых сегментов β -слоев, а также в области участка, соединяющего два домена. Остальные участки цепи просматривались плохо. В результате качество карт распределения электронной плотности позволило провести ход полипептидной цепи приблизительно для 80% всех остатков, общее количество которых предполагалось равным 165. Интерпретация боковых групп была проблематичной.

Учитывая, что при расчете синтеза электронной плотности не использовалась почти половина модулей структурных факторов из-за отсутствия для них фазовой информации, можно было надеяться на улучшение синтеза при включении в расчет этих дополнительных данных. С этой целью было проведено улучшение набора фаз для γ-кристаллина Шb . Улучшение выполнялось путем построения и модификации грубой атомной модели.

4.2. Построение грубой модели.

Для построения грубой модели использовался синтез электронной плотности, рассчитанный с коэффициентами Фурье

$$m_{HL}(\vec{s})F_0(\vec{s})e^{i\varphi_{HL}(\vec{s})}, \quad 1/|\vec{s}| \geq 3.0\ \text{Å} \qquad (107)$$

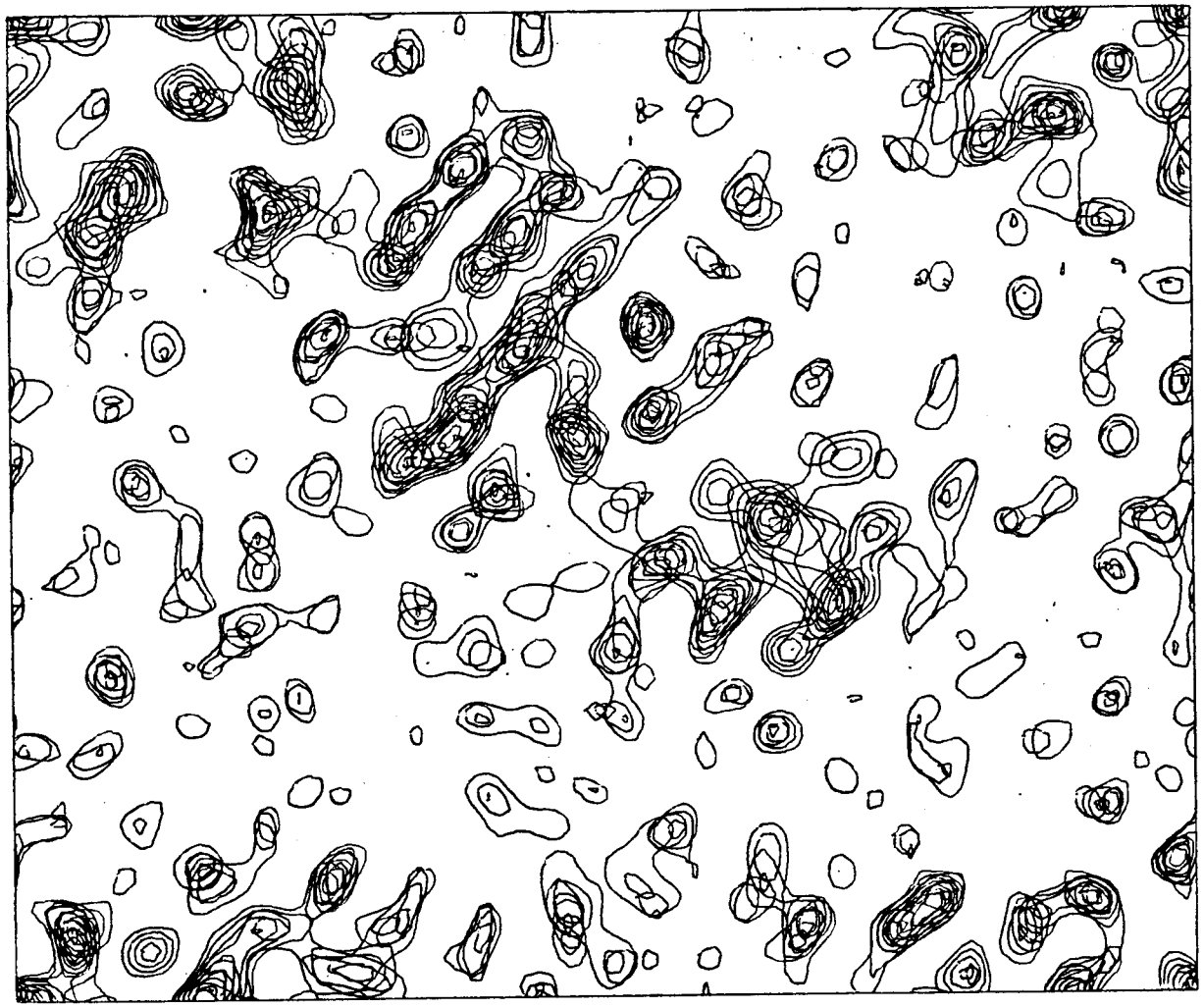

Рис.6. Изображение функции распределения электронной плотности в γ-кристаллине Шb . xy -сечение, z =(27÷30)/144. Набор фаз - исходный (изоморфное замещение, аномальное рассеяние), разрешение 3.0 Å.

где $F_o(\vec{s})$ - экспериментально определенные значения модулей структурных факторов,

$\varphi_{HL}(\vec{s}), m_{HL}(\vec{s})$ - "наилучшая" фаза и ее показатель достоверности (66), вычисленные из распределения $P_{\vec{s}}^{HL}(\varphi)$ типа (58), параметры которого определялись с помощью усовершенствованной фазовой программы Тен Эйка по данным изоморфного замещения.

Приведение модулей структурных факторов к абсолютной шкале осуществлялось по графику Вильсона [1,226]. Значение F_{000}/V было оценено как 0.177 е/Å^3. Максимальное значение функции распределения электронной плотности ρ_{MAX} равнялось 0.91е/Å^3. Шаг сетки, на которой была рассчитана функция $\rho(\vec{r})$, определялся возможностью последующей работы с данными разрешением до 2.3 Å. Соответствующее число разбиений элементарной ячейки по осям Ox, Oy, Oz составило 100, 120, 200.

По значениям показателей достоверности определения фаз исходного набора S_o структурных факторов, состоявшего из 6777 отражений, были определены значения параметров аппроксимации (73) изображения атома азота на рассчитанном синтезе: B_o =83.9 Å^2, C_o =8.4 е. Поскольку проведение полипептидной цепи на ряде участков было затруднено, то решено было отказаться от выделения области молекулы и проводить построение грубой модели во всем объеме независимой части элементарной ячейки. В результате по имевшемуся трехангстремному синтезу при использовании критического значения электронной плотности ρ_{CRIT} =0.28 е/Å^3 была построена исходная грубая модель M_0, состоявшая из 3104 фиктивных атомов.

4.3. Модификация грубой модели.

Модификация грубой модели велась так называемыми "сериями", состоявшими из последовательно выполняемых двух циклов модификации значений температурного фактора и трех циклов модификации координат. Как правило, в серию входила также и автоматическая корректировка грубой модели по значениям температурного фактора.

Первоначально были выполнены 2 серии модификации модели M_0 по данным разрешением до 3.0 Å. Величина кристаллографического фактора достоверности упала с 0.353 до 0.115 по модифицированной модели M_4 (здесь и далее кристаллографический фактор достоверности R (90) рассчитан по отражениям, которые включались в минимизируемый критерий).

Для фиктивных атомов модели M_4, имевших нулевой температурный фактор, их тип был изменен с азота на кислород. Полученная таким образом модель M_5, состоявшая из 467 фиктивных атомов типа кислорода и 2585 фиктивных атомов типа азота, модифицировалась затем последовательно по данным разрешением до 2.7 Å и до 2.5 Å. Результатом модификации явилась модель M_9.

Последняя модель реконструировалась с помощью разностного синтеза. Из модели M_9 было выбрано 2602 фиктивных атома со значениями температурного фактора, не превышавшими B_{LIM} =50 Å^2 (модель M_{10}). Модель M_{10} дополнялась фиктивными атомами, определенными по синтезу электронной плотности, рассчитанному с коэффициентами Фурье

$$F_0(\vec{s})e^{i\varphi_9^c(\vec{s})} - F_{10}(\vec{s})e^{i\varphi_{10}(\vec{s})}, \quad 1/|\vec{s}| \geqslant 3.0 \text{ Å} \qquad (108)$$

Здесь $F_{10}(\vec{s})e^{i\varphi_{10}(\vec{s})}$ - структурные факторы, рассчитанные по модели M_{10}, а фазы $\varphi_9^c(\vec{s})$ - "наилучшие" фазы, определенные согласно (66) из

$$P_{\vec{s}}^{c9}(\varphi)=P_{\vec{s}}^{HL}(\varphi)\cdot P_{\vec{s}}^{9}(\varphi) \qquad (109)$$

Параметры распределения $P_{\vec{s}}^{9}(\varphi)$, построенного по данным модели M_9, определялись из принципа максимального правдоподобия [225]. 1030 заново определенных фиктивных атомов (параметры изображения атома на синтезе равнялись B_0 =64.2 Å^2, C_0 =10.3 e) вместе с атомами грубой модели M_{10} составили новую грубую модель M_{11}.

Параметры атомов модели M_{11} модифицировались последовательно по данным разрешением до 3.0 Å, до 2.7 Å и до 2.5 Å. Для фиктивных атомов полученной в результате модификации модели M_{17} вновь проводилось изменение типов атомов. При установлении значения температурного фактора B_j равным нулю тип атома изменялся с кислорода на серу, а с азота - на кислород. Одновременно из модели удалялись все атомы со значениями B_j> 50 Å^2.

Полученная в результате модель M_{18}, состоявшая из 3455 фиктивных атомов типа азота, кислорода и серы, модифицировалась по полному набору имевшихся данных разрешением до 2.3Å. Набор фаз $\varphi_{22}(\vec{s})$ структурных факторов, рассчитанных по последней модели M_{22}, был взят в качестве улучшенного и послужил объектом дальнейшего анализа.

Подробно ход модификации продемонстрирован в таблице 9.

4.4. Анализ качества наборов фаз.

Контроль процесса модификации и оценка качества полу-

Таблица 9. План модификации грубой модели γ-кристаллина Шb.

модифицируемая модель*)	кол-во атомов	зона модификации, Å	кол-во отражений	уровень срезки, Å^2	R-фактор,% исх.	фин.
M_0	3I04	3.0-30.0	8533	-	35.3	I6.6
M_2	3I04	3.0-30.0	8533	80	I6.6	II.5
M_4	3052	-	8533	-	II.5	-
M_5	3052	2.7-30.0	I0527	80	I6.9	II.9
M_7	3045	2.5-30.0	II727	70	I4.6	I2.3
M_9	2962	-	II727	-	I2.3	-
M_{II}	3632	3.0-30.0	8533	-	24.8	I2.0
M_{I3}	3632	2.7-30.0	I0527	80	I5.2	II.5
M_{I5}	3628	2.5-30.0	II727	70	I3.6	II.4
M_{I7}	36I4	-	II727	-	II.4	-
M_{I8}	3455	2.3-30.0	I25I0	40	I7.2	I4.2
M_{20}	3229	2.3-I0.0	I2I95	-	I3.6	II.6
M_{22}	3229	-	I2I95	-	II.6	-

*) Для каждой модели проводилось по 5 циклов модификации (2 цикла модификации значений температурного фактора и 3 цикла модификации координат).

ченных наборов фаз проводились с помощью критериев, описанных выше в п.п. 2.3, 2.6.

Поведение фактора достоверности

$$R = \sum_{\vec{s}}' |F_M(\vec{s}) - F_0(\vec{s})| \Big/ \sum_{\vec{s}}' F_0(\vec{s}) \qquad (110)$$

отражено в таблице 9.

График зависимости расхождения фаз $\varphi_{22}(\vec{s})$ с исходными значениями $\varphi_{HL}(\vec{s})$ приведен на рис. 7. Полученная зависимость характеризует набор фаз $\{\varphi_{22}(\vec{s})\}$ как непротиворечащий экспериментам с тяжелоатомными производными.

В таблице 10 приведено распределение средней ошибки фаз структурных факторов, рассчитанных по различным моделям. Средняя ошибка оценивалась из распределений $P^M_{\vec{s}}(\varphi)$, параметры которых были определены из принципа максимального правдоподобия. Как следует из рис. 3 (см. п. 1.7), оценка погрешности фаз структурных факторов, включавшихся в минимизируемый критерий, является заниженной приблизительно в два раза. Соответственно, полученные по $P^M_{\vec{s}}(\varphi)$ оценки для ошибок таких фаз перед включением в таблицу были помножены на 2. Оставляя открытым вопрос об абсолютной величине ошибок, можно утверждать, что данные таблицы 10 отражают динамику уточнения значений фаз.

Анализ графиков зависимости среднего значения показателей достоверности от разрешения для различных наборов фаз (рис. 8) приводит к следующим выводам.

1) Комбинированные фазы $\varphi^C_{22}(\vec{s})$, полученные из распределения

$$P^{C22}_{\vec{s}}(\varphi) = P^{HL}_{\vec{s}}(\varphi)\, P^{22}_{\vec{s}}(\varphi) \qquad (111)$$

в интервале от 3.0 Å приблизительно до 10.0 Å имеют " в

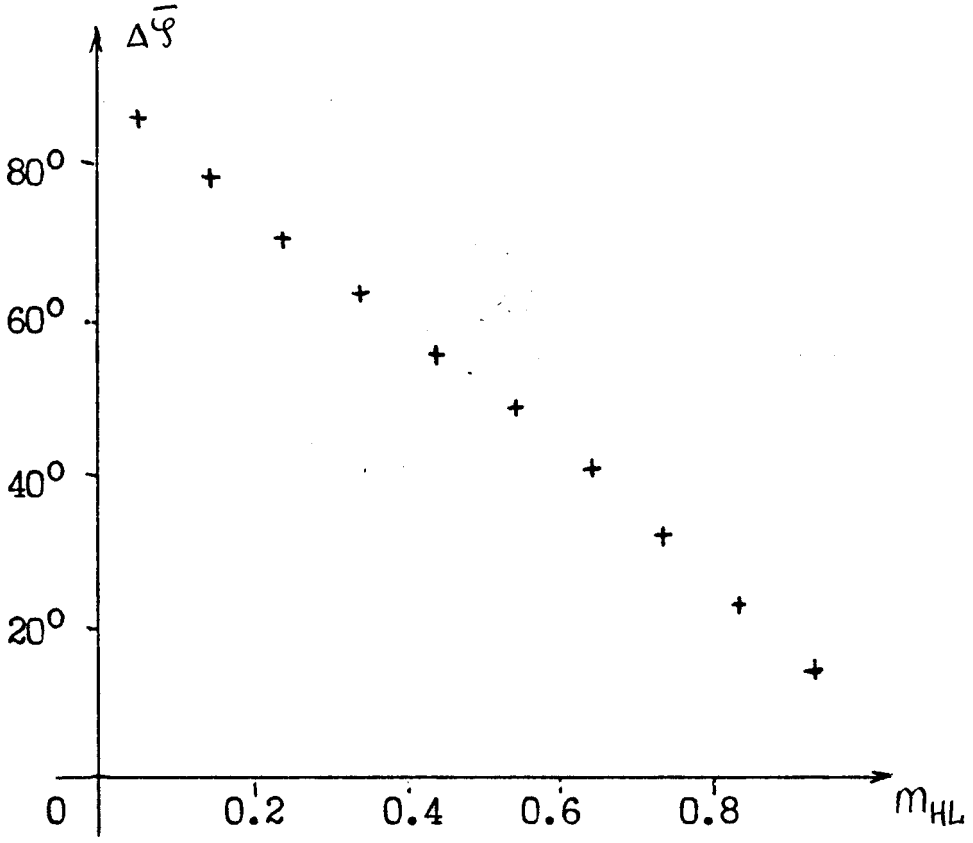

Рис. 7. Зависимость среднего расхождения исходных значений фаз $\varphi_{HL}(\vec{s})$ и фаз $\varphi_{22}(\vec{s})$, рассчитанных по модифицированной грубой модели M_{22} γ-кристаллина Шb, от величины показателя достоверности $m_{HL}(\vec{s})$ исходных фаз.

Таблица I0. Оценка (в градусах) средней ошибки фаз структурных факторов, рассчитанных по грубым моделям γ-кристаллина IIIb (символом $^+$ помечены оценки, умноженные на 2).

модель	границы интервалов по $\lvert\vec{s}\rvert^2$									
	0.00–0.02	0.02–0.04	0.04–0.06	0.06–0.08	0.08–0.I0	0.I0–0.I2	0.I2–0.I4	0.I4–0.I6	0.I6–0.I8	0.00–0.I8
M_0	38	36	3I	38	4I	52	90	84	75	5I
M_2	54^+	26^+	20^+	22^+	24^+	27	90	90	80	4I
M_4	50^+	20^+	$I6^+$	$I6^+$	$I6^+$	24	90	85	77	37
M_7	48^+	$I8^+$	$I4^+$	$I6^+$	$I6^+$	20^+	42^+	90	77	30
M_9	50^+	$I8^+$	$I4^+$	$I4^+$	$I6^+$	20^+	26^+	34^+	76	23
M_{II}	30	24	I7	I9	20	26	37	43	8I	27
M_{I3}	50^+	22^+	$I6^+$	$I6^+$	$I6^+$	I6	40	56	77	27
M_{I5}	48^+	$I8^+$	$I4^+$	$I6^+$	$I6^+$	$I8^+$	28	6I	90	26
M_{I7}	46^+	$I8^+$	$I4^+$	$I4^+$	$I4^+$	$I8^+$	20^+	30	90	22
M_{20}	48^+	22^+	$I6^+$	$I6^+$	$I6^+$	$I8^+$	20^+	26^+	36^+	22
M_{22}	48^+	20^+	$I4^+$	$I4^+$	$I6^+$	$I6^+$	$I8^+$	22^+	28^+	I8
кол-во отраж.	504	I068	I442	I759	I667	I634	I360	872	4I7	I0723

среднем" большие показатели достоверности, чем исходные фазы $\varphi_{HL}(\vec{s})$.

2) В интервале от 3.0 Å приблизительно до 2.7 Å фазы $\varphi^{c}_{22}(\vec{s})$ имеют достаточно высокие показатели достоверности.

3) Для отражений с разрешением приблизительно от 2.5 Å и выше предсказанные значения фаз $\varphi^{c}_{22}(\vec{s})$ имеют низкие показатели достоверности. Мы предполагаем, что такое резкое снижение $m(|\vec{s}|)$ связано с существенным недостатком данных в этой области.

4) Для отражений низкого разрешения (приблизительно от 10 Å и ниже) исходные значения фаз более достоверны, чем значения фаз, определенные по моделям. Скорее всего, это связано с отсутствием учета влияния растворителя на рассеяние рентгеновских лучей.

4.5. Построение и анализ заключительного синтеза.

Нами был построен ряд синтезов электронной плотности с использованием фазовой информации, определенной по грубой модели M_{22}. При этом варьировались разрешение синтезов, значения фаз, весовая схема. При подготовке соответствующих карт распределения электронной плотности менялись значения линий уровня. Наиболее удачным набором линий уровня был признан следующий

$$\rho_j = \rho_{AV} + \ell_j(\rho_{MAX} - \rho_{AV}), \qquad j = 1,2,3,4,5 , \qquad (112)$$

где $\rho_{MAX} = \max_{\vec{r}} \rho(\vec{r})$, $\rho_{AV} = \sum_{\vec{r}} \rho(\vec{r}) / M(\{\vec{r}\})$, а ℓ_j принимает соответственно значения 0.20, 0.35, 0.50, 0.65, 0.80 .

В качестве значений фаз коэффициентов Фурье функции распределения электронной плотности использовались либо фазы

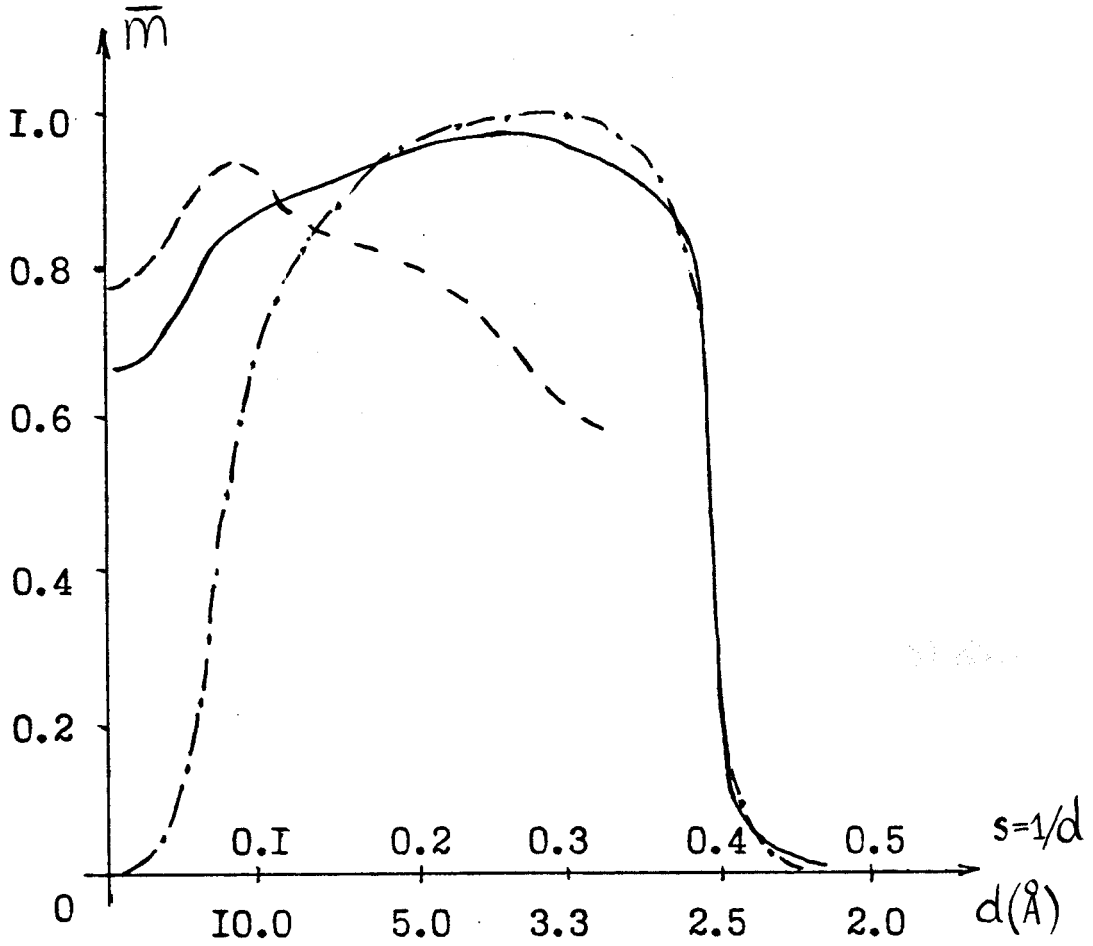

Рис. 8. Зависимость средних значений показателей достоверности определения фаз от разрешения.

— — — — исходный набор $\{\varphi_{HL}(\vec{s})\}$

—·—· — набор фаз $\{\varphi_9(\vec{s})\}$ структурных факторов, рассчитанных по грубой модели M_9

——— — набор фаз $\{\varphi^c_{22}(\vec{s})\}$, полученных из комбинированного распределения $P^{c22}_{\vec{s}}(\varphi)$ (III)

$\varphi_{22}(\vec{s})$, рассчитанные по последней грубой модели M_{22}, либо их комбинации с исходными величинами. Использовались различные способы комбинирования и различные оценки параметров распределения $P^{22}_{\vec{s}}(\varphi)$. Также были рассмотрены синтезы, где для отражений разрешением $d >$ 10.0 Å брались фазы $\varphi_{HL}(\vec{s})$, определенные по данным изоморфного замещения, а для отражений более высокого разрешения брались фазы, определенные каким-либо из описанных выше способов. Анализировались синтезы как невзвешенные, так и взвешенные показателями достоверности соответствующих фаз. Варьировалась также и верхняя граница разрешения структурных факторов, включавшихся в расчет синтеза. Синтезы разрешением 3.0 Å использовались для сравнения с исходным синтезом. По полному набору данных строились синтезы разрешением 2.3 Å. Поскольку в интервале 2.7 Å $\geq d \geq$ 2.3 Å процент собранных данных был мал, то дополнительно анализировались и синтезы разрешением 2.7 Å.

В результате проделанного сравнения карт различных распределений электронной плотности наиболее удачным был признан синтез с коэффициентами

$$F(\vec{s})e^{i\varphi(\vec{s})} = \begin{cases} m_{HL}(\vec{s})F_0(\vec{s})e^{i\varphi_{HL}(\vec{s})}, & 1/|\vec{s}| > 10.0 \text{ Å} \\ m^c_{22}(\vec{s})F_0(\vec{s})e^{i\varphi^c_{22}(\vec{s})}, & 10.0 \text{ Å} \geq 1/|\vec{s}| \geq 2.7 \text{ Å} \end{cases} \qquad (113)$$

где $m^c_{22}(\vec{s})$ и $\varphi^c_{22}(\vec{s})$ определены из (66) с $P_{\vec{s}}(\varphi) = P^{HL}_{\vec{s}}(\varphi)\, P^{22}_{\vec{s}}(\varphi)$, а параметры распределения $P^{22}_{\vec{s}}(\varphi)$ оценивались из принципа максимального правдоподобия [225].

В расчет синтеза вошло 10804 отражения. При расчете синтеза в абсолютной шкале его максимальное значение оказалось равным 2.06 e/Å^3 (для аналогичного невзвешенного

синтеза ρ_{MAX} =2.64 e/Å^3). Контрастность синтеза $S=\rho_{MAX}/\rho_{AV}$ возросла с 5.2 до II.6.

Аналогичный синтез был построен с данными разрешением до 3.0 Å. Для него числовые характеристики оказались равными ρ_{MAX} =I.76 e/Å^3, $S=\rho_{MAX}/\rho_{AV}$ =9.9 (для невзвешенного синтеза ρ_{MAX} =2.40 e/Å^3). Количество отражений, вошедших в расчет - 88I6.

Отметим, что анализ различных синтезов и выбор среди них наилучшего выполнялся независимо от сравнения зависимостей $\bar{m}(|\vec{S}|)$ (рис. 8), но привел фактически к тем же самым выводам.

Сравнивая исходный и заключительный синтезы разрешением 3.0 Å (изображение их фрагментов приведено на рис. 6 и рис. 9), нетрудно заметить превосходство второго над первым.

По сравнению с итоговым трехангстремным синтезом, на синтезе разрешением 2.7 Å более четко проявились отдельные детали структуры (изображение фрагмента синтеза приведено на рис. I0). В то же время включение в расчет синтеза дополнительных слагаемых более высокого разрешения привело к возникновению некоторых разрывов в изображении полипептидной цепи. Таким образом, при последующем построении атомной модели использовался синтез разрешением 2.7 Å в то время, как для анализа хода полипептидной цепи более удобным оказался улучшенный трехангстремный синтез.

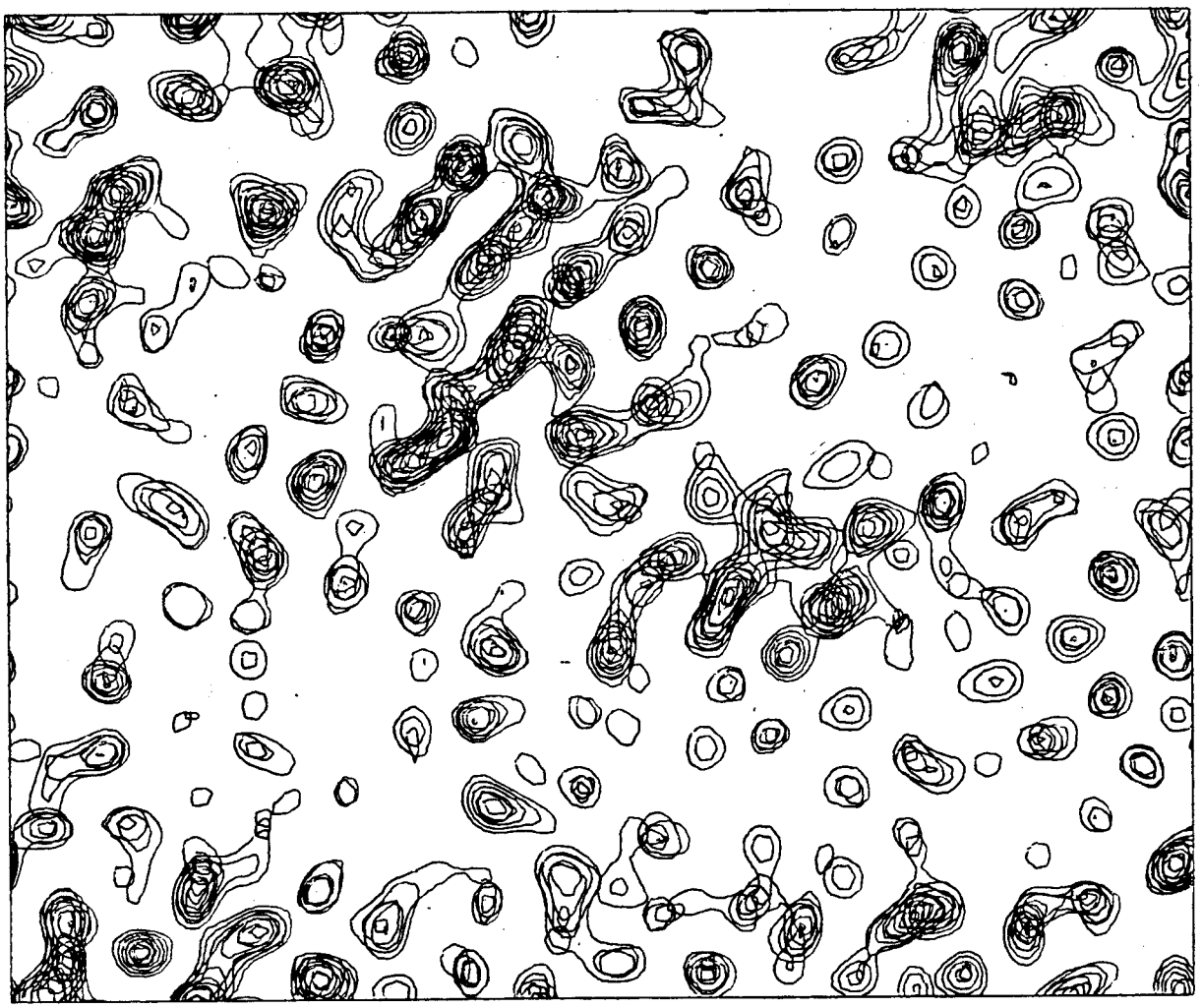

Рис.9. Изображение функции распределения электронной плотности в γ-кристаллине Шb. xy-сечение, z =(27÷30)/I44. Набор фаз определен из комбинированного распределения (III) $P_{\vec{s}}^{C22}(\varphi)=P_{\vec{s}}^{HL}(\varphi)\,P_{\vec{s}}^{22}(\varphi)$, разрешение 3.0 Å.

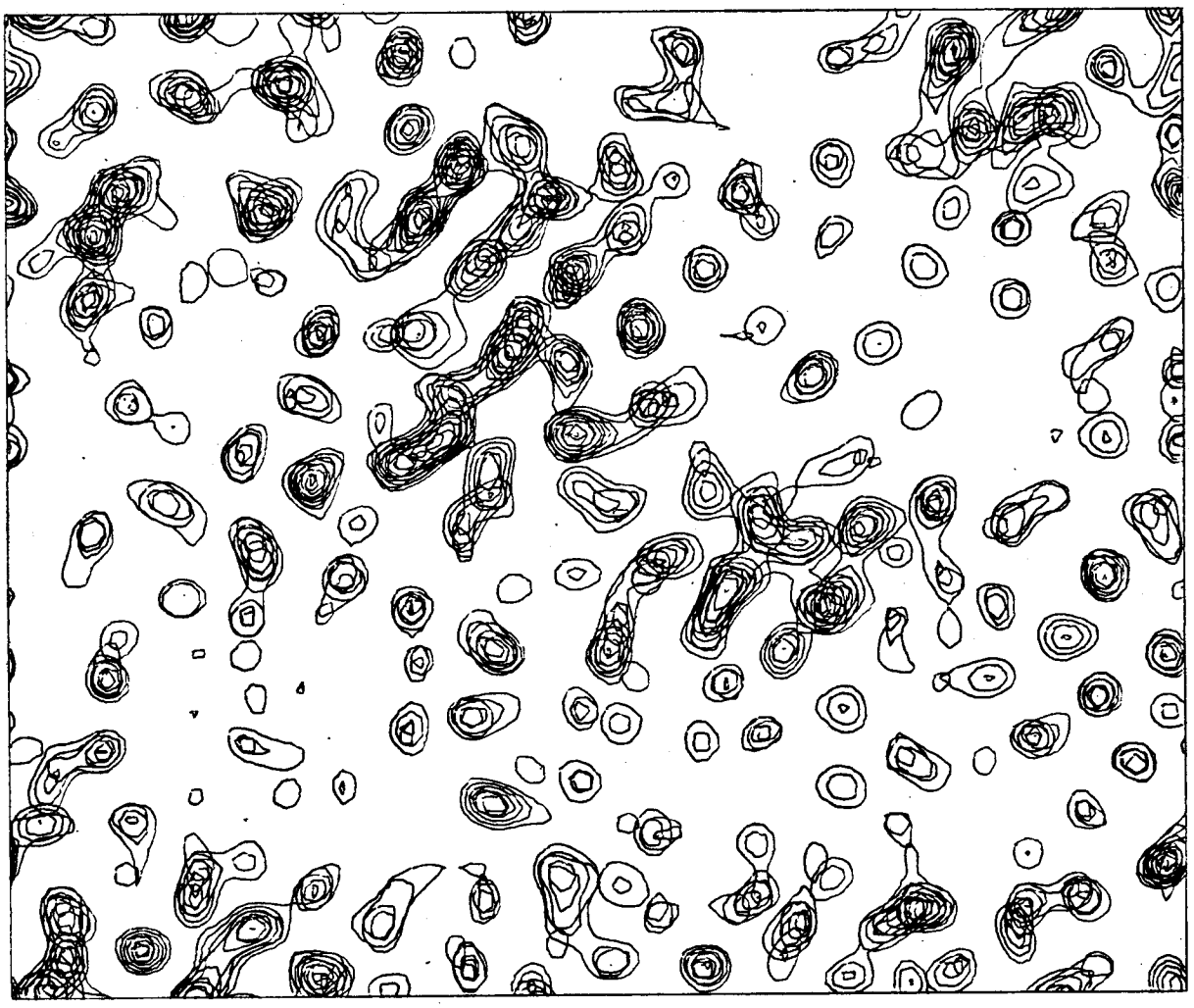

Рис.I0. Изображение функции распределения электронной плотности в γ-кристаллине Шb. xy-сечение, z =(27÷30)/I44. Набор фаз определен из комбинированного распределения (III) $P_{\vec{s}}^{C22}(\varphi)=P_{\vec{s}}^{HL}(\varphi)\,P_{\vec{s}}^{22}(\varphi)$, разрешение 2.7 Å.