

ГЛАВА IV. ОПРЕДЕЛЕНИЕ *AB-INITIO* ЗНАЧЕНИЙ ФАЗ СТРУКТУРНЫХ ФАКТОРОВ ДЛЯ РЕФЛЕКСОВ НИЗКОГО РАЗРЕШЕНИЯ С ПОМОЩЬЮ КРИТЕРИЯ ОБОБЩЕННОГО ПРАВДОПОДОБИЯ

4.1 Постановка задачи.

Данная глава посвящена определению значений фаз $\{\varphi(\mathbf{h})\}_{\mathbf{h} \in \mathbf{S}}$ структурных факторов для некоторого фиксированного набора рефлексов \mathbf{S} . Предполагается, что модули структурных факторов этого набора $\{F^{obs}(\mathbf{h})\}_{\mathbf{h} \in \mathbf{S}}$ известны из эксперимента. Широко распространенной стратегией определения неизвестных фаз *ab-initio* является рассмотрение большого количества пробных значений фаз $\{\varphi(\mathbf{h})\}_{\mathbf{h} \in \mathbf{S}}$, удовлетворяющих ограничениям симметрии для centrosymmetric рефлексов, и выбор из них одного или нескольких возможных решений на основании некоторого критерия. В данной главе исследуется возможность использования критерия обобщенного правдоподобия при определении и расширении значений фаз низкого разрешения.

Каждому возможному фазовому набору можно поставить в соответствие область молекулы в элементарной ячейке и свести выбор между альтернативными фазовыми наборами к выбору между альтернативными областями молекулы. Для любого гипотетического набора фаз низкого разрешения можно вычислить синтез Фурье, используя эти фазы и экспериментальные модули структурных факторов. В качестве области молекулы можно выбрать ту область элементарной ячейки фиксированного объема, которая содержит точки с максимальными значениями рассчитанного синтеза (Рис. 4.1). Задача выбора наилучшего фазового набора из альтернативных наборов может быть переформулирована как задача выбора области молекулы среди альтернативных областей. В свою очередь, как было показано в главе 1, выбор между альтернативными областями

можно рассматривать как выбор между альтернативными априорными распределениями координат атомов в элементарной ячейке.

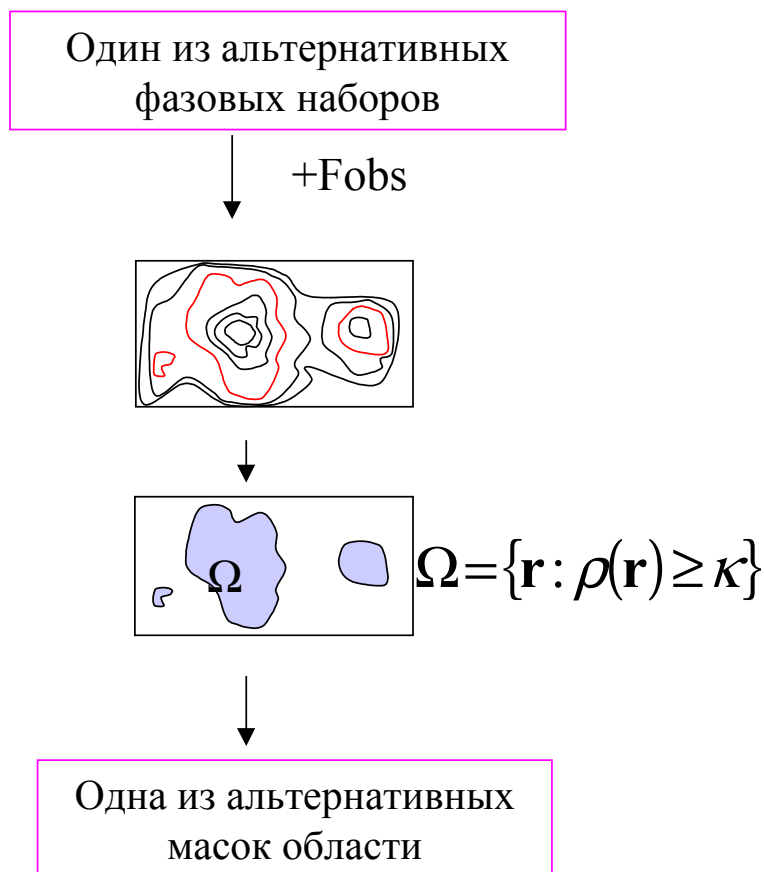


Рисунок 4.1. Схема перехода от задачи выбора наилучшего фазового набора к выбору одной из альтернативных масок области молекулы.

4.2 Выбор наилучшего фазового набора среди нескольких альтернативных фазовых наборов

Критерий обобщенного правдоподобия был применен для выбора одного из нескольких альтернативных фазовых наборов, полученных с помощью FAM-метода. FAM-метод приводит на некотором этапе к нескольким альтернативным кластерам, состоящим из близких фазовых наборов (Podjarny, Urzhumtsev & Lunin, 1998). Для каждого кластера были

рассчитаны центроидные фазы, а затем, в соответствии со схемой, изображенной на рис. 4.1, маска области. Каждой маске области было поставлено в соответствие априорное распределение координат атомов, равное константе внутри этой маски и нулю вне ее. Для каждого такого априорного распределения в ходе компьютерного эксперимента рассчитывалось обобщенное правдоподобие, определенное в главе 2. Ожидалось, что лучшей маске области будет соответствовать максимальное значение обобщенного правдоподобия. Это предположение подтвердилось в тестах с экспериментальными данными для РНКазы SA разрешения 16 Å при сравнении масок, выделяющих 60% объема элементарной ячейки (Рис. 4.2 а), что приблизительно соответствует объему молекулы. При сравнении масок, построенных по тем же фазовым наборам, но выделяющим только 30% элементарной ячейки, максимальное значение обобщенного правдоподобия получалось не для лучшей маски (Рис. 4.2 б). Отметим, однако, что объем, занимаемый молекулой в элементарной ячейке кристалла, определяется, как правило, достаточно точно на основании значения молекулярного веса.

4.3 Использование обобщенного правдоподобия в качестве критерия при *ab-initio* определении фаз низкого разрешения

Процедура определения *ab-initio* значений фаз аналогична процедуре, предложенной Луниным, Уржумцевым и Сковородой (1990). Она состоит из генерирования очень большого числа фазовых наборов, ниже называемых вариантами, отбора вариантов с максимальными

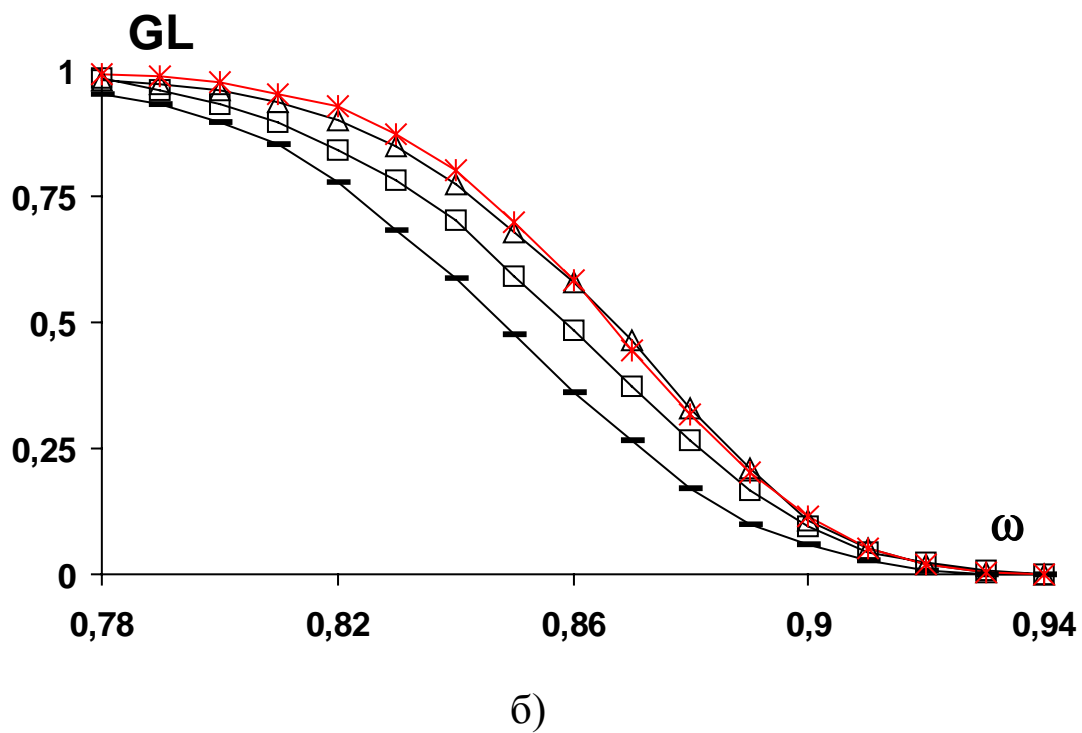
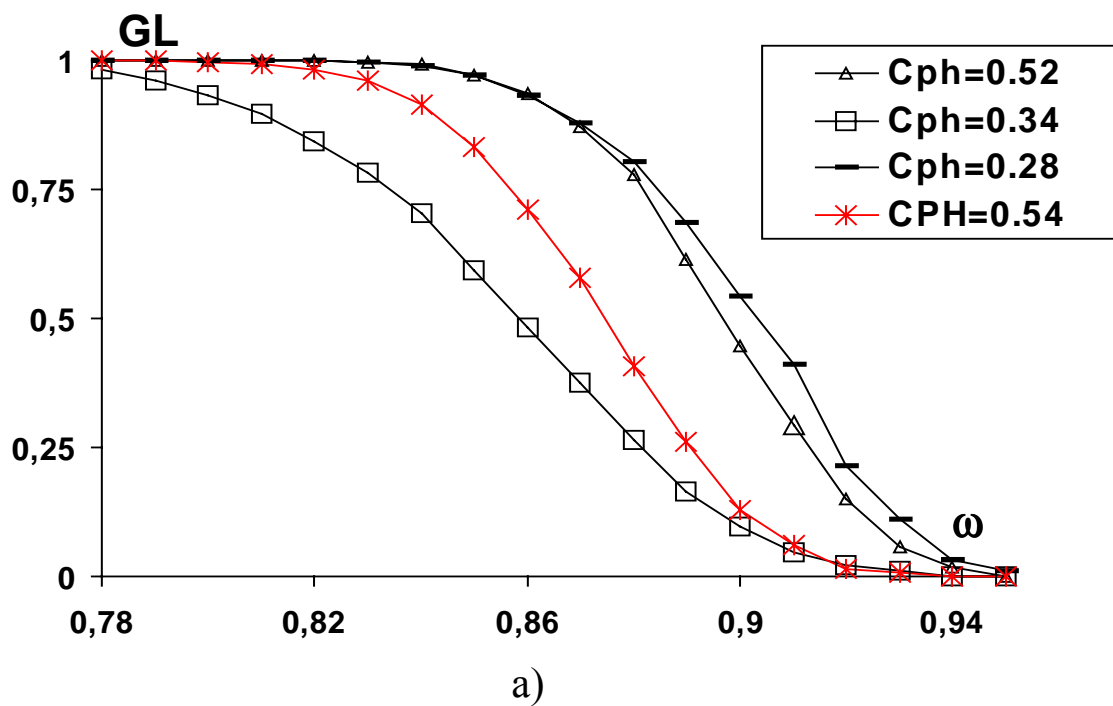


Рисунок 4.2. Зависимость величины критерия обобщенного правдоподобия от уровня срезки для четырех альтернативных масок, полученных с помощью FAM метода. Дифракционные данные для RNase Sa разрешения $d=16-100 \text{ \AA}$. Объем сравниваемых масок а) 30% б) 60% объема элементарной ячейки.

значениями некоторого критерия и дальнейшего анализа и усреднения отобранных вариантов. Ключевым моментом этой процедуры является выбор критерия отбора вариантов. В данной работе в качестве критерия используется значение обобщенного правдоподобия, определенного по формуле (2.5).

В практической работе более эффективно использовать не один, а несколько независимых критериев. Однако, основной задачей данной работы было исследование возможности использования критерия обобщенного правдоподобия самого по себе. Ниже будут кратко рассмотрены несколько примеров использования критерия правдоподобия в комбинации с другими критериями.

4.4 Генерация вариантов

Первым этапом рассматриваемой процедуры является генерация большого числа значений фаз для выбранной группы рефлексов низкого разрешения. Широко используемые способы перебора большого количества фазовых наборов описаны в Гл.1 п.1.22 . Из перечисленных способов в данной работе были использованы два: “полный” перебор и случайная генерация. При “полном” переборе значение фазы каждого centrosymmetric рефлекса принимает одно из двух возможных значений, а фаза каждого неcentrosymmetric рефлекса принимает значения: $\pm\pi/4$, $\pm3\pi/4$. Такой способ перебора был использован, когда рассматриваемый набор рефлексов представлял собой небольшое количество только сильных рефлексов. Заметим, что во многих пространственных группах симметрии среди рефлексов низкого разрешения преобладают centrosymmetric рефлексы. Поэтому на очень низком разрешении для небольшой группы рефлексов в первую очередь представляется целесообразным использовать именно этот способ перебора. В тестах, описанных ниже, при работе со всеми подряд рефлексами из некоторой зоны разрешения использовалась также случайная генерация фазовых наборов.

4.5 Контрольные функции

Все тесты проводились со структурами с уже известными атомными моделями. В качестве контрольной функции был использован коэффициент корреляции фаз (2.7), где $\{\varphi_h^{true}\}$ представляет собой набор правильных фаз, рассчитанных по известной атомной модели, $\{\varphi_h\}$ набор пробных фаз, а $\{F_h^o\}$ набор экспериментальных модулей. “Хорошими” вариантами являются фазовые наборы, обладающие высоким значением коэффициента корреляции фаз (2.7), соответственно, “плохие” варианты - это фазовые наборы с низким значением коэффициента (2.7). Наилучшим фазовым набором является набор, обладающий максимальным значением фазовой корреляции.

4.6 Тесты и результаты

Тесты проводились с тремя наборами экспериментальных данных, описанных выше в разделе 2.5.2 : комплекса аминоксил-тРНК-синтетазы с тРНК, γ -кристаллина Шв и фактора элонгации G. Все расчеты проводились с экспериментальными, а не с рассчитанными данными.

4.6.1 *Ab-initio* определение фаз. Полный перебор

В первом тесте с данными для комплекса аминоксил-тРНК-синтетазы с тРНК было сформировано 4096 фазовых наборов путем полного перебора для 12 сильнейших рефлексов в диапазоне разрешения $68 \text{ \AA} - \text{inf}$. На рис. 4.3 представлено распределение значений обобщенного правдоподобия и фазовой корреляции. Можно видеть, что вариант, наиболее близкий к правильному варианту, имеет одно из самых высоких значений правдоподобия. Однако, существуют плохие варианты с высоким значением правдоподобия и, наоборот, хорошие варианты с низким

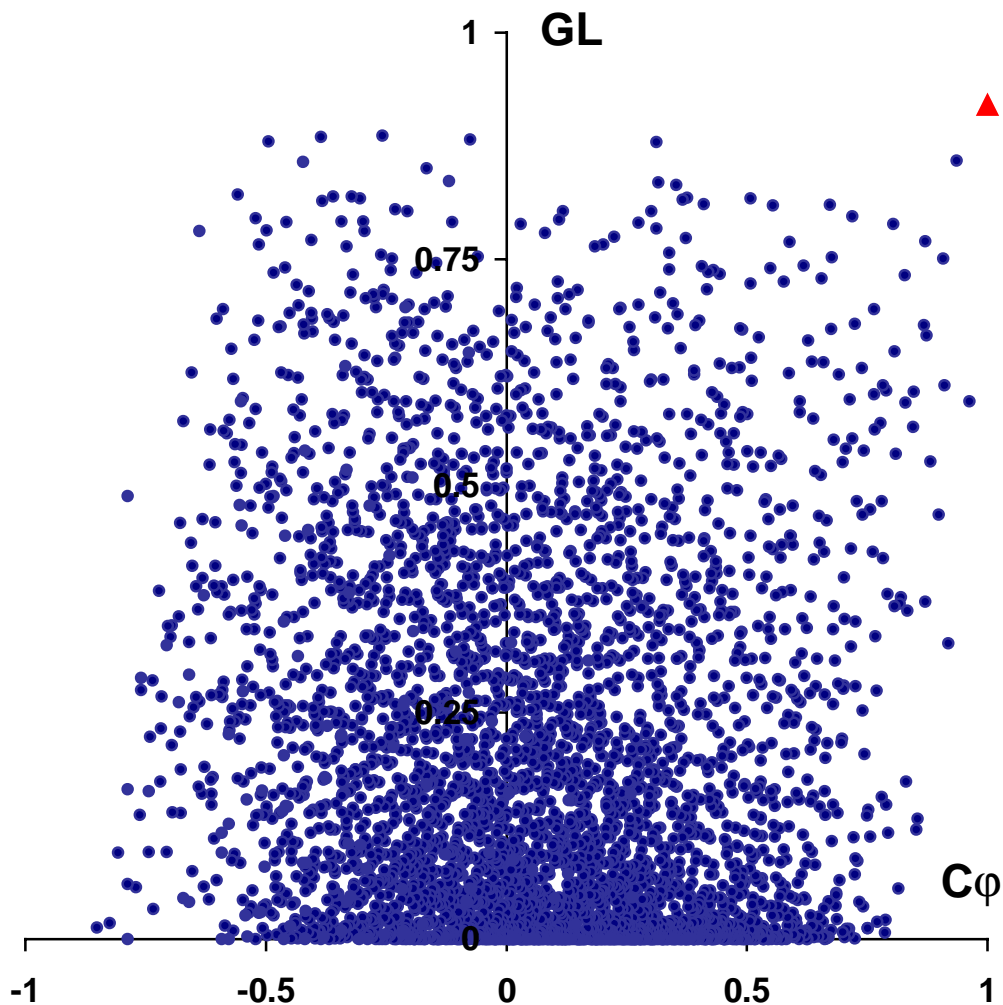


Рисунок 4.3 *Ab-initio* определение фаз. Комплекс аминоксил-тРНК-синтетазы с тРНК. Полный перебор значений фаз для 12-ти самых сильных рефлексов в диапазоне разрешения $d=68-250$ Å. Рассмотрено 4096 фазовых наборов.

Обобщенное правдоподобие вычислялось по всем рефлексам в диапазоне разрешения $d=60-250$ Å. $C\phi$ рассчитывался по рефлексам в диапазоне разрешения $d=68-250$ Å. Значения параметров для вычисления обобщенного правдоподобия: $B_{glob}=8000$, объем тестируемой маски области $V=0.3$ объема элементарной ячейки, число псевдоатомов $N=100$, уровень срезки $w=0.86$. Треугольник соответствует фазовому набору, ближайшему к правильному решению.

значением правдоподобия. Т.о. не наблюдается однозначной зависимости между качеством фазового набора и значением правдоподобия.

В данной ситуации возможны два подхода: либо использовать какую-либо дополнительную информацию об исследуемой структуре, либо менять стратегию определения значений фаз. Примером первого подхода является тест с комплексом аминоксил-тРНК-синтетазы с тРНК, который показал, что при наличии некоторой дополнительной информации об исследуемой структуре критерий обобщенного правдоподобия может однозначно указывать на правильное решение. В случае с данным комплексом из анализа упаковки было ясно, что молекула не является ни тримером, ни тетрамером. Из этого следует, что не должно быть областей высокой плотности на осях вращения 3-го и 4-го порядков. Учитывая этот факт, из всех 4096 фазовых наборов были отобраны только варианты, для которых построенные маски имеют меньше, чем 0.12% точек, лежащих на осях вращения 3-го и 4-го порядков. На рис. 4.4 представлены только эти варианты. Можно видеть, что вариант, ближайший к правильному, имеет значение правдоподобия значительно превышающее правдоподобие для остальных рассмотренных вариантов. Следует подчеркнуть, что наличие структурной информации, которая может быть затем успешно использована в сочетании с критерием правдоподобия является скорее исключением, чем правилом. Тем не менее, даже как частный случай, такой пример представляет интерес.

При формировании фазовых наборов в тестах с гамма-кристаллином Шь и фактором элонгации G необходимо было учесть тот факт, что в пространственной группе $P2_12_12_1$ существует восемь вариантов выбора начала координат, а также неопределенность относительно выбора энантиоморфа. Это дает возможность зафиксировать значения фаз для четырех сильных centrosимметричных рефлексов. (Заметим, что эти четыре centrosимметричных рефлекса выбираются не произвольно. Они должны быть подобраны таким образом, чтобы при всех возможных сдвигах начала координат и смене энантиоморфа были перебраны все

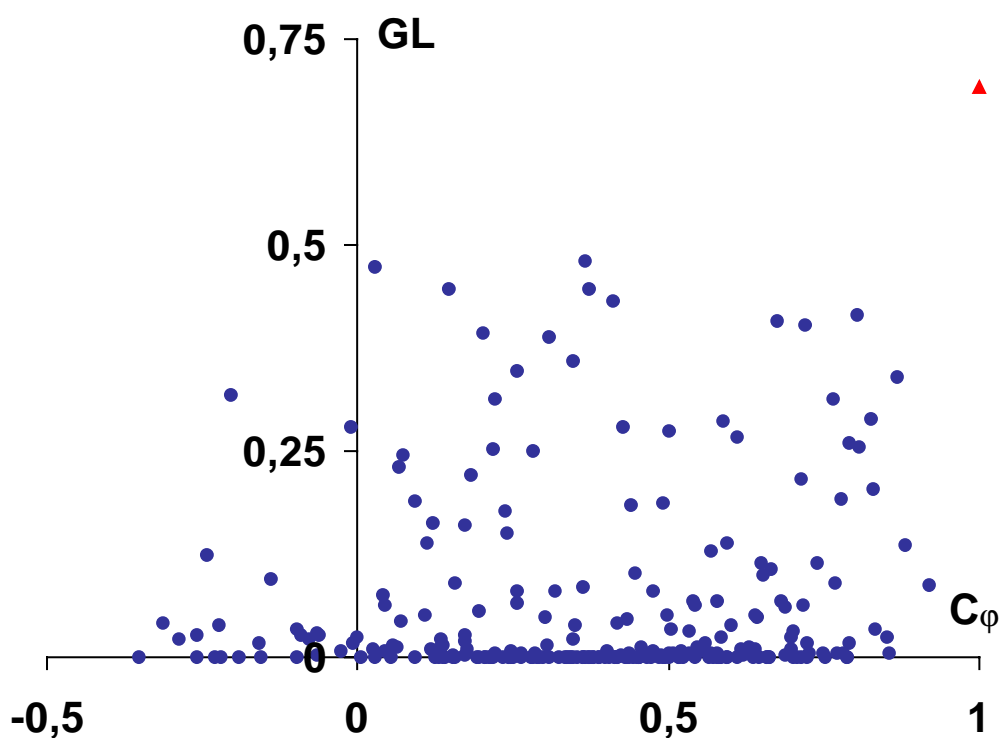


Рисунок 4.4 *Ab-initio* определение фаз. Использование дополнительной информации о структуре. Комплекс аминоксил-тРНК-синтетазы с тРНК. Из 4096 масок, рассчитанных по 4096 фазовым наборам, были отобраны только маски, имеющие меньше, чем 0.12 % точек на осях вращения 3-го и 4-го порядков. C_ϕ рассчитывался по рефлексам в диапазоне разрешения $d=68-250$ Å. Обобщенное правдоподобие вычислялось рефлексам в диапазоне разрешения $d=60-250$ Å.

Значения параметров для вычисления обобщенного правдоподобия: $B_{glob} = 8000$, объем тестируемой маски области $V=0.3$ объема элементарной ячейки, число псевдоатомов $N=100$, уровень срезки $w=0.88$.

Треугольник соответствует фазовому набору, ближайшему к правильному решению.

возможные комбинации фаз для этих рефлексов.) Для γ -кристаллина были выбраны 13 сильнейших рефлексов разрешения 29 \AA^{-1} , фазы 4-х ц/с рефлексов были зафиксированы, а для 9-ти оставшихся рефлексов был проведен полный перебор. Для фактора элонгации G были выбраны 12 сильнейших рефлексов разрешения 34 \AA^{-1} , фазы 4-х ц/с рефлексов были зафиксированы, а для 8-и оставшихся рефлексов был проведен полный перебор. Диаграмма для распределений обобщенного правдоподобия и фазовой корреляции получились аналогичны диаграмме, полученной в тестах с комплексом аминоксил-тРНК-синтетазы с тРНК. Как хорошие, так и плохие варианты имеют высокое значение правдоподобия. Лучшие фазовые наборы не могут быть однозначно определены только с помощью критерия обобщенного правдоподобия. Распределение значений фазовой корреляции и обобщенного правдоподобия γ -кристаллина представлено на рис. 4.5. Заметим, что даже самый плохой вариант имеет фазовую корреляцию выше 0.5. Такая близость всех рассмотренных вариантов объясняется фиксацией 4-х сильных рефлексов.

4.6.2 *Ab-initio* определение фаз. Случайная генерация фазовых наборов

В предыдущих тестах с сильнейшими рефлексами низкого разрешения не удалось выделить лучшие фазовые наборы только с помощью критерия правдоподобия. Тем не менее, дальнейшие тесты показали, что во многих случаях существует следующая статистическая тенденция – хорошие варианты чаще обладают высоким значением правдоподобия, чем плохие. Было сгенерировано случайно большое количество фазовых наборов, рассчитано значение обобщенного правдоподобия для всех вариантов, а затем были отобраны только варианты с высоким значением обобщенного правдоподобия. Сравнение распределений значений фазовой корреляции для всех сгенерированных вариантов и для отобранных вариантов показало, что среди отобранных вариантов относительное число вариантов с высоким

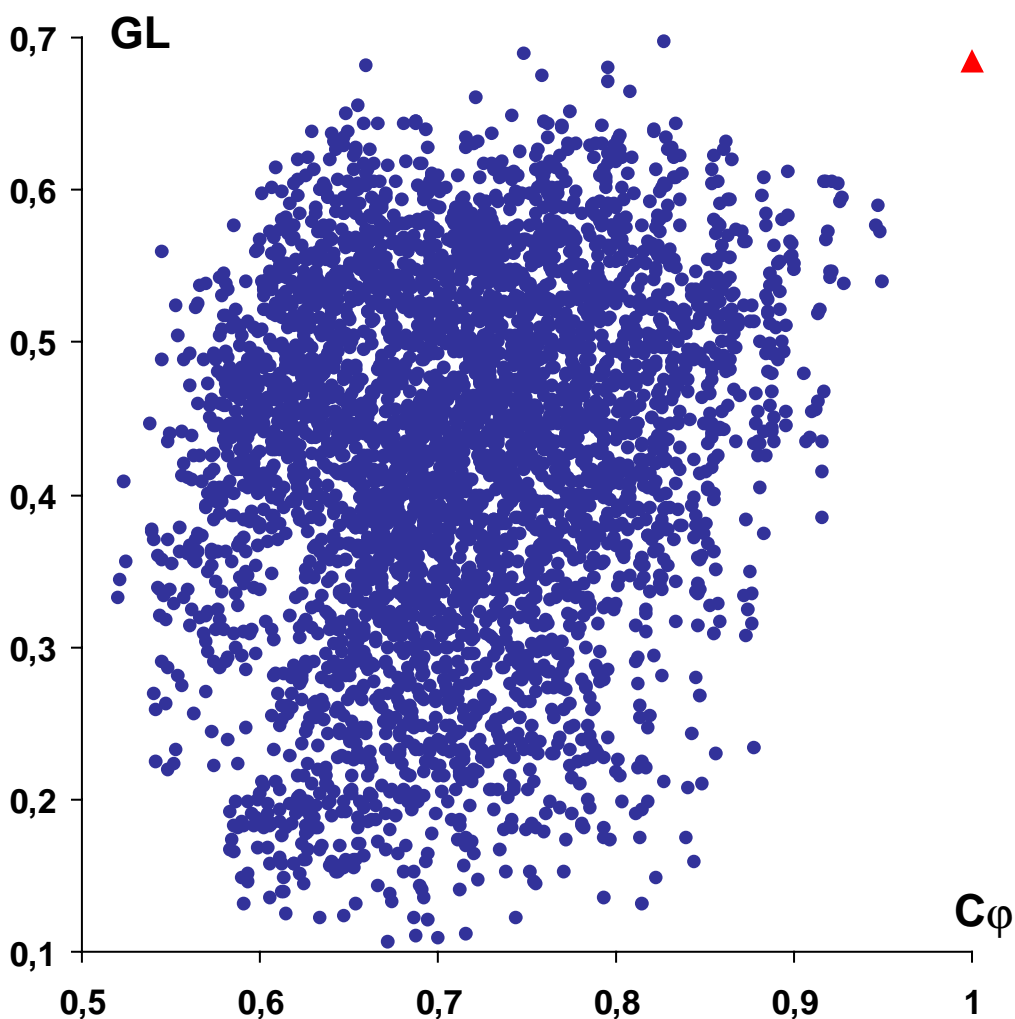


Рисунок 4.5 *Ab-initio* определение фаз. γ -кристаллин IIIb. Фазы 4-х сильных рефлексов были фиксированы. Полный перебор значений фаз для 9-ти самых сильных рефлексов в диапазоне разрешения $d=29-100$ Å. Рассмотрено 4096 фазовых наборов.

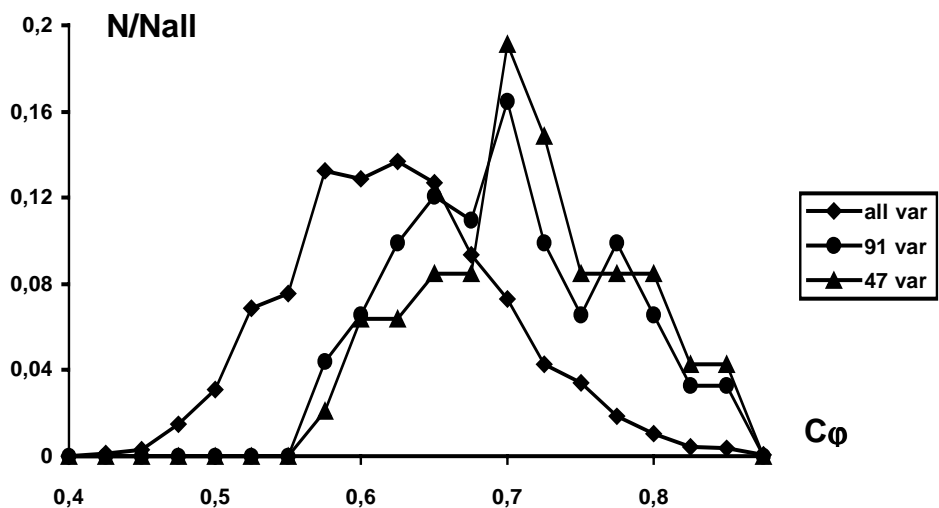
Обобщенное правдоподобие вычислялось по всем рефлексам в диапазоне разрешения $d=23-100$ Å. C_ϕ рассчитывался по рефлексам в диапазоне разрешения $d=29-100$ Å. Значения параметров для вычисления обобщенного правдоподобия: $B_{glob}=100$, объем тестируемой маски области $V=0.4$ объема элементарной ячейки, число псевдоатомов $N=100$, уровень срезки $w=0.80$. Треугольник соответствует фазовому набору, ближайшему к правильному решению.

значением фазовой корреляции намного выше, чем среди всех сгенерированных вариантов (Рис. 4.6). Критерий обобщенного правдоподобия может служить как фильтр для отбора ансамбля вариантов с более высокой концентрацией хороших вариантов. В случае трех рассматриваемых структур эта тенденция наблюдалась для наборов, состоящих из фаз 12-30 рефлексов самого низкого разрешения.

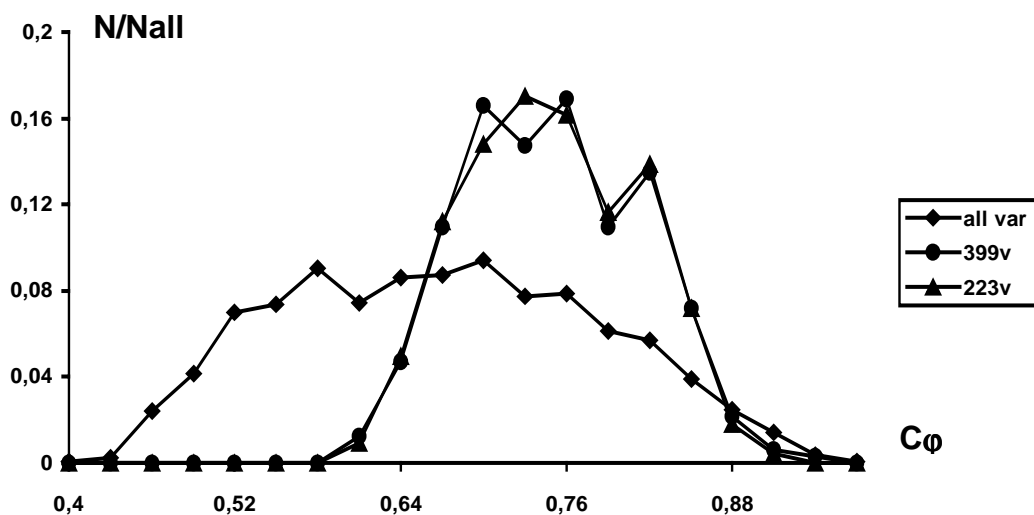
Ансамбль отобранных вариантов с высокой концентрацией хороших вариантов, можно усреднить, рассчитав центроидные фазы и индивидуальные показатели достоверности для каждой фазы по формуле:

$$m(\mathbf{h}) \exp[\varphi^{best}(\mathbf{h})] = \frac{1}{M} \sum_{j=1}^M \exp[i\varphi_j(\mathbf{h})]. \quad (4.1)$$

Полученные фазы с соответствующими показателями достоверности дают лучшее значение фазовой корреляции с правильными фазами, чем решение, полученное усреднением по всем случайно сгенерированным вариантам, и могут служить в качестве начального приближения к решению фазовой проблемы. Результаты сравнения вариантов, усредненных по всем сгенерированным фазовым наборам и только по отобранным наборам, представлены в таблице 4.1. Заметим, что для белков пространственной группы $P2_12_12_1$ наибольший интерес представляет сравнение значений C_φ , рассчитанных по группе рефлексов, не включающей четыре рефлекса с фиксированными фазами.



a)



б)

Рисунок 4.6 Сравнение распределения значений фазовой корреляции C_ϕ для всех случайно сгенерированных вариантов и для отобранных вариантов с максимальными значениями обобщенного правдоподобия.

а) γ -кристаллин Pb; 2000 фазовых наборов разрешения $d=23-100 \text{ \AA}$ были сгенерированы случайно. Распределение значений C_ϕ представлено для трех случаев: все 2000 вариантов, 91 вариант с $GL \geq 0.47$ и 47 вариантов с $GL \geq 0.5$. Обобщенное правдоподобие рассчитывалось по рефлексам в диапазоне разрешения $d=20-100 \text{ \AA}$. Параметры для расчета GL : число псевдоатомов $N=100$, The parameter $B_{glob}=100$.

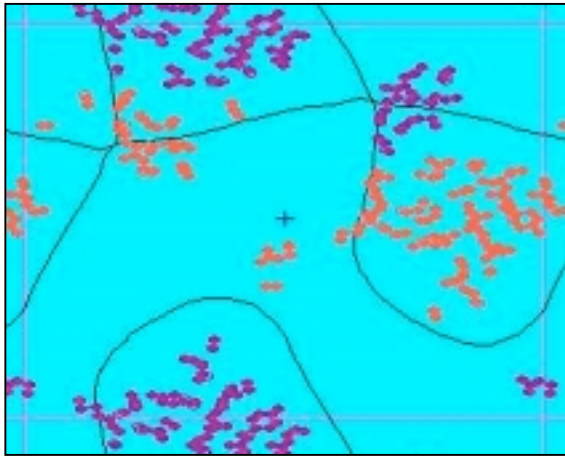
	<i>γ</i> -кристаллин IIIb		фактор элонгации <i>G</i>		комплекс <i>AspRS</i>
	C_{φ} d=23 Å		C_{φ} d=29 Å		C_{φ} d=50 Å
	по всем рефл.	исключая 4 рефл. с фикс. фазами	по всем рефл.	исключая 4 рефл. с фикс. фазами	по всем рефл.
v1	0.71	-0.01	0.60	0.21	0.22
v2	0.83	0.64	0.61	0.47	0.49

Таблица 4.1 Сравнение решений, полученных усреднением всех случайно сгенерированных вариантов и усреднением только отобранных вариантов с максимальными значениями обобщенного правдоподобия.

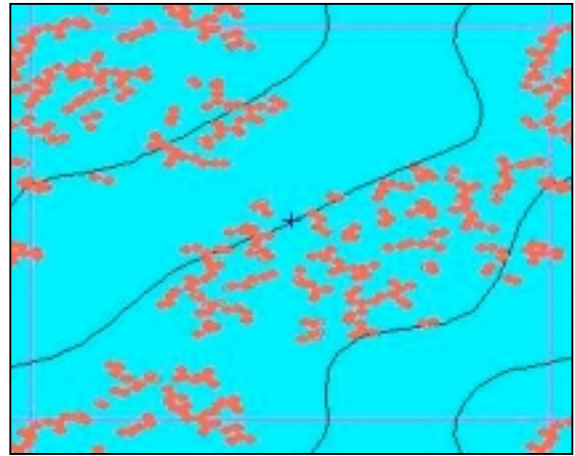
V1 - решение, полученное усреднением 2000 случайно сгенерированных вариантов,

v2- решение, полученное усреднением 100 отобранных вариантов с максимальными значениями обобщенного правдоподобия.

Были рассчитаны синтезы низкого разрешения по экспериментальным модулям и усредненным по отобранным вариантам фазовым наборам. На рис. 4.7 и 4.8 показаны контуры молекулы, как они были определены с помощью с помощью предлагаемой процедуры и атомные модели для *γ*-кристаллина IIIb и фактора элонгации *G*.



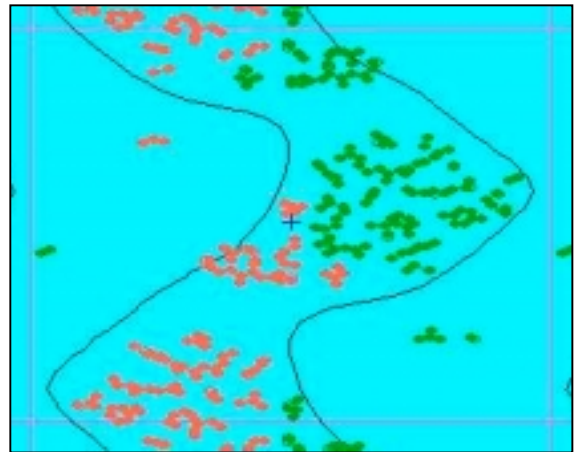
Section Z = 00/40



Section Z = 04/40

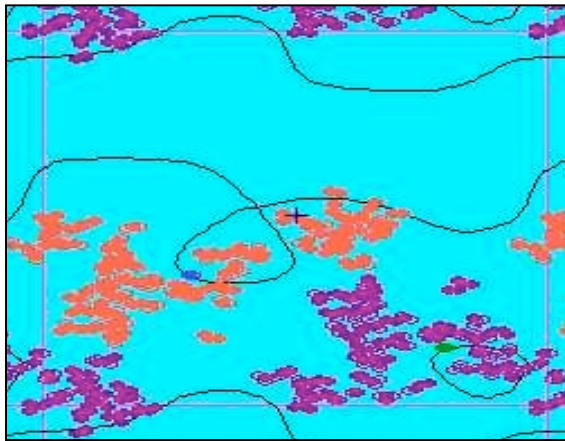


Section Z = 07/40

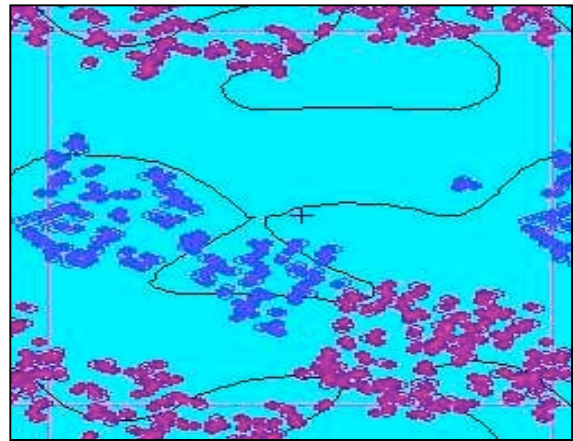


Section Z = 10/40

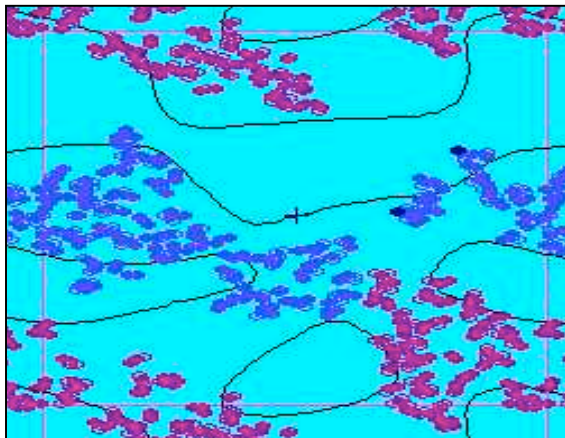
Рисунок 4.7 Синтез, построенный по экспериментальным модулям и фазам, усредненным по отобранным вариантам с максимальными значениями обобщенного правдоподобия, и атомы модели для γ -кристаллин ПШв. 2000 фазовых наборов разрешения $d=23\text{-inf \AA}$ были сгенерированы случайно, из них были отобраны 91 вариант, у которых значение $GL \geq 0.47$ и усреднены. Черный контур выделяет на синтезе область, содержащую 0.45 % точек с максимальными значениями синтеза.



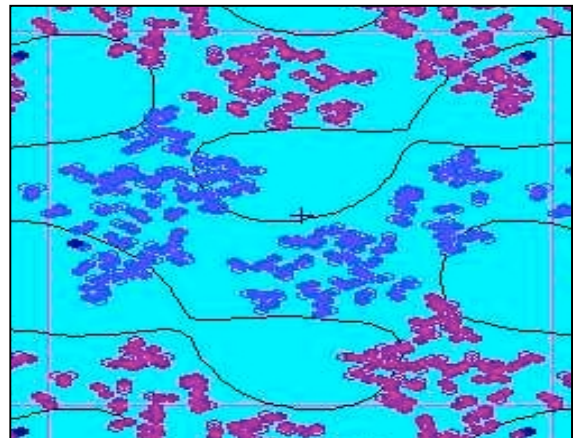
Section Z = 00/40



Section Z = 04/40



Section Z = 07/40



Section Z = 10/40

Рисунок 4.8 Синтез, построенный по экспериментальным модулям и фазам, усредненным по отобранным вариантам с максимальными значениями обобщенного правдоподобия, и атомы модели для фактора элонгации G . 2000 фазовых наборов разрешения $d=29\text{-inf \AA}$ были сгенерированы случайно, из них были отобраны 184 варианта, у которых значение $GL \geq 0$. и усреднены. Черный контур выделяет на синтезе область, содержащую 0.60 % точек с максимальными значениями синтеза.

4.6.3 Сочетание критериев правдоподобия и связности

Согласно исследованиям проведенным Луниным, Луниной и Уржумцевым (Urzumtsev, Lunin & Lunina, 1998), синтез низкого разрешения, соответствующий правильным фазам, имеет столько связных компонент, сколько молекул содержится в элементарной ячейке. Ими был предложен критерий качества фаз низкого разрешения - критерий связности. Используя этот критерий они получили решения для гамма-кристаллина IIIb и фактора элонгации G. Было проведено усреднение решения, полученного отбором и усреднением вариантов с высоким значением обобщенного правдоподобия, и решения, полученного с помощью критерия связности. Как можно видеть из таблиц 4.2 и 4.3, усреднение этих решений дает решение лучшего качества, чем каждое из решений в отдельности.

	C_s <i>по всем рефл.</i>	C_s <i>Исключ. 4 рефл. с фикс. фазами</i>	C_φ <i>по всем рефл.</i>	C_φ <i>Искл. 4 рефл. с фикс. фазами</i>
V1-решение, полученное с помощью критерия правдоподобия	0.82	0.63	0.82	0.64
V2-решение, полученное с помощью критерия связности	0.87	0.76	0.82	0.63
V3 – результат усреднения решений V1 and V2	0.87	0.78	0.92	0.84

Таблица 4.2. Усреднение решений, полученных с помощью критерия правдоподобия и критерия связности для гамма-кристаллина IIIb; $d=24\text{\AA}$. C_s – коэффициент корреляции синтезов, рассчитываемый аналогично C_φ , но с учетом показателей достоверности.

	C_s <i>по всем рефл.</i>	C_s <i>Исключ. 4 рефл. с фикс. фазами</i>	C_ϕ <i>по всем рефл.</i>	C_ϕ <i>Искл. 4 рефл. с фикс. фазами</i>
V1-решение, полученное с помощью критерия правдоподобия	0.66	0.32	0.39	0.02
V2-решение, полученное с помощью критерия связности	0.67	0.41	0.51	0.32
V3 – результат усреднения решений V1 and V2	0.70	0.61	0.52	0.41

Таблица 4.3. Усреднение решений, полученных с помощью критерия правдоподобия и критерия связности для фактора элонгации G; d=29 Å.

4.6.4 Расширение набора фаз

Целью следующей серии тестов было исследовать возможность применимости критерия обобщенного правдоподобия при расширении фазового набора.

Предыдущие тесты (разд. 4.3.1), в которых фазовые наборы формировались путем перебора только сильных рефлексов самого низкого разрешения, показали, что как хорошие, так и плохие варианты могут иметь высокое значение правдоподобия (рис. 4.3 и 4.5). На этом этапе критерий правдоподобия не разделяет хорошие и плохие варианты. Был поставлен вопрос, нельзя ли из вариантов, имеющих высокое значение правдоподобия на первом этапе, выделить хорошие решения при расширении фазового набора, используя процедуру, аналогичную построению “фазового дерева”

(Bricogne & Gilmore, 1990). Были отобраны варианты с высокими значениями правдоподобия, значения фаз для этих вариантов были зафиксированы, и эти варианты образовали “узлы” фазового дерева. Затем были выбраны несколько сильных рефлексов более высокого разрешения, и для этих дополнительных рефлексов проводился полный перебор значений фаз. В результате, для каждого узла были сформированы фазовые наборы, в которых значения фаз, соответствующих “узлам”, были фиксированы, а перебирались значения фаз рефлексов следующей зоны разрешения. Затем, критерий обобщенного правдоподобия рассчитывался для всех вариантов всех узлов полученного фазового дерева. Наблюдалась такая же тенденция, как в и случае случайной генерации фазовых наборов: отбор вариантов с высокими значениями правдоподобия давал в результате множество фазовых наборов с более высокой концентрацией хороших вариантов. Однако, этот эффект наблюдался в узком диапазоне разрешения, только для зон разрешения, содержащих не более 30-ти рефлексов.

Для комплекса аминоксил-тРНК-синтетазы с тРНК удалось в ходе описанной выше процедуры построения “фазового дерева” выделить лучший узел на основании критерия обобщенного правдоподобия. Из всех 4096 вариантов, представленных на рис.4.3, были отобраны только 20 вариантов с самыми высокими значениями правдоподобия. Для каждого варианта фазы первых 12-ти рефлексов, принадлежащих зоне разрешения $68 \text{ \AA} - \text{inf.}$ были зафиксированы, а перебирались фазы 6-ти сильных рефлексов в диапазоне разрешения $48 - 68 \text{ \AA}$. Для всех вариантов 20-ти узлов фазового дерева было рассчитано значение обобщенного правдоподобия. Самые высокие значения правдоподобия были получены для вариантов лучшего узла. Более того, чем выше было значение S_{ϕ} в диапазоне $68 \text{ \AA} - \text{inf.}$, тем более сильная корреляция наблюдалась для вариантов этого узла между S_{ϕ} в диапазоне $48-68 \text{ \AA}$ и значениями обобщенного правдоподобия (Рис.4.9 а)-д)). На следующем этапе, среди всех полученных вариантов были отобраны 5 вариантов с максимальными значениями правдоподобия и усреднены. Полученные значения 18-ти фаз были зафиксированы, а перебирались значения фаз еще

для 5-ти сильных рефлексов разрешения 48 -68 Å . Распределение значений обобщенного правдоподобия и фазовой корреляции для сформированных вариантов представлено на рис.4.10. Опять наблюдается четкая зависимость между качеством фазового набора и значением обобщенного правдоподобия. Усреднение семи вариантов с максимальным значением правдоподобия дает в результате фазовый набор, имеющий фазовую корреляцию $C_\phi = 0.98$ с точными фазами, посчитанную по 23м рефлексам разрешения 68 Å - inf. Синтез рассчитанный с этими фазами и экспериментальными модулями хорошо очерчивает общие контуры комплекса в элементарной ячейке (рис.4.11 и 4.12).

Т.о. в частном случае комплекса аминоксил-тРНК-синтетазы с тРНК критерий правдоподобия в ходе процедуры расширения фазового набора позволил выделить лучшее решение на разрешении 68 Å - inf., а затем расширить это решение до разрешения 48 Å. Однако, применение аналогичной процедуры для γ -кристаллина и фактора элонгации G к такому же успеху не привели. Возможной причиной является близость всех рассматриваемых вариантов в пространственной группе $P2_12_12_1$.

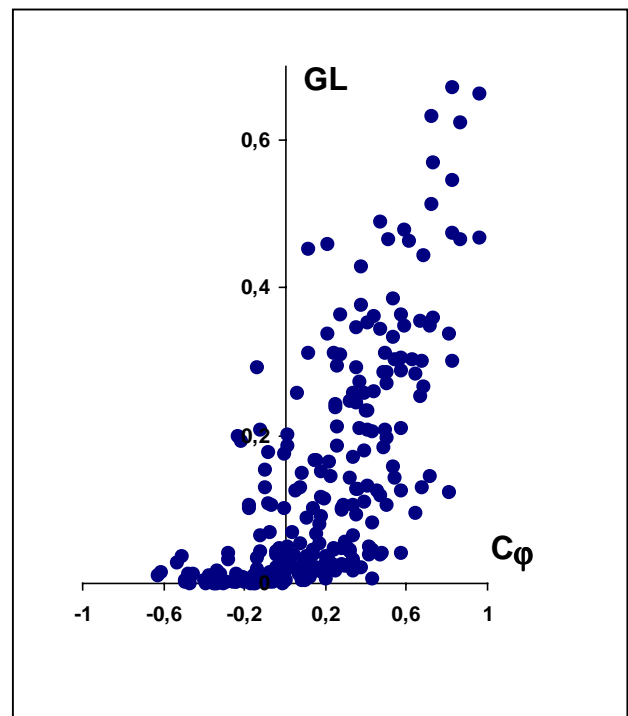
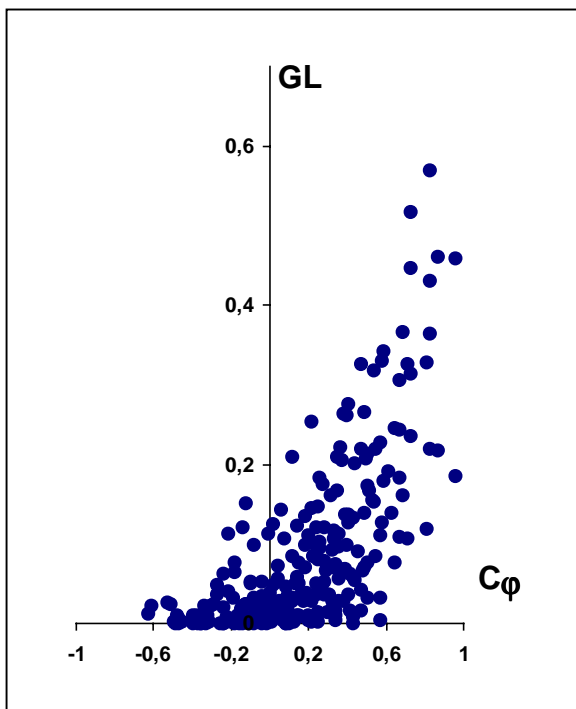
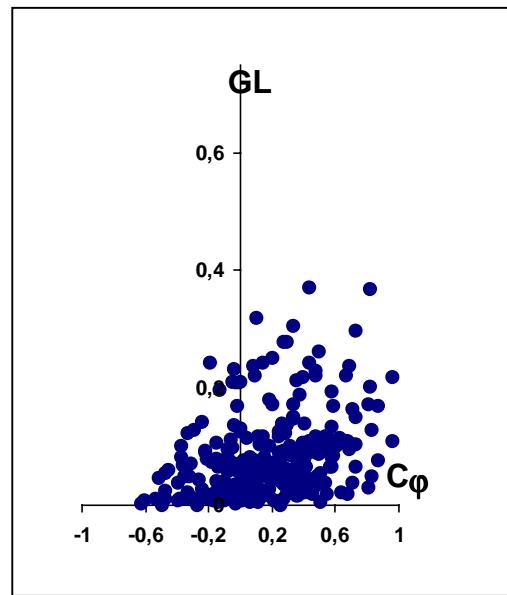
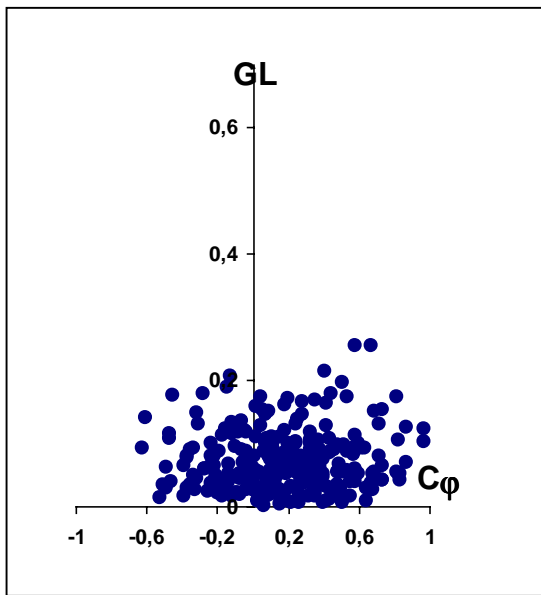


Рисунок 4.9 Расширение фазового набора для 4-х вариантов разрешения $d=68-250\text{\AA}$, имеющих высокие значения обобщенного правдоподобия при первом полном переборе значений фаз. 12 фаз было фиксировано ($d=68-250\text{\AA}$), перебирались значения фаз для 6-ти сильных рефлексов диапазона разрешения $d=48-68\text{\AA}$.

Значение фазовой корреляции фиксированных фаз с правильными фазами:

а) $C_\varphi = 0.49$ б) $C_\varphi = 0.31$ в) $C_\varphi = 0.94$ д) $C_\varphi = 0.9997$

Коэффициент C_φ рассчитывался по рефлексам разрешения $d=48-68\text{\AA}$.

Обобщенное правдоподобие рассчитывалось по рефлексам разрешения $d=40-250\text{\AA}$. Значения параметров для расчета обобщенного правдоподобия: $w=0.64$, $V=8000$, объем области молекулы составлял 0.3 объема элементарной ячейки.

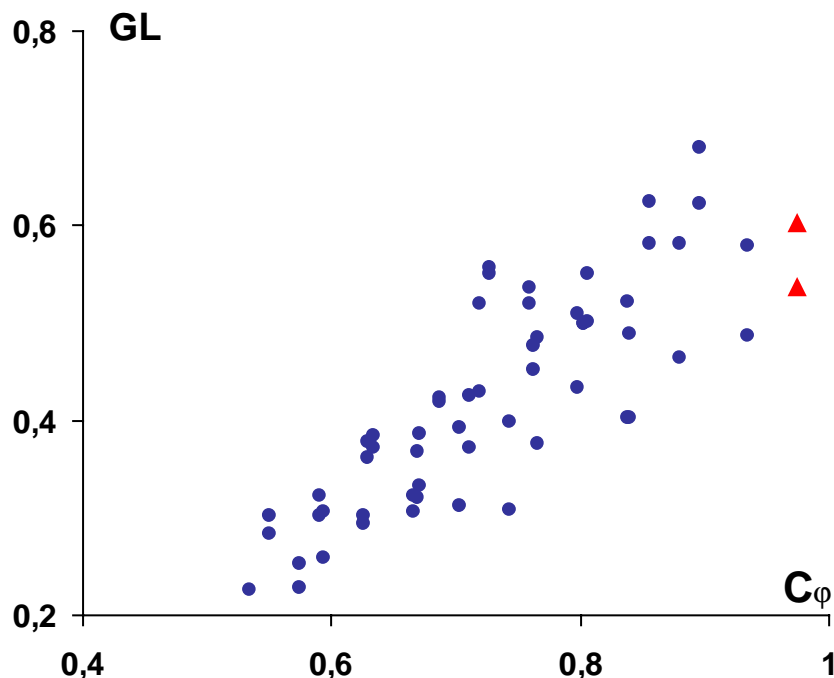


Рисунок 4.10 Второе расширение фазового набора. Комплекс аминоксил-тРНК-синтетазы с тРНК $d=48-250$ Å. Были отобраны после первого расширения фазового набора 5 вариантов с самыми высокими значениями обобщенного правдоподобия. Эти 5 вариантов были усреднены и полученные фазы были зафиксированы. Перебирались фазы для 5-ти сильных рефлексов разрешения $d=48-68$ Å. Коэффициент C_ϕ рассчитывался по рефлексам разрешения $d=48-68$ Å. GL was calculated at $d=40$ Å. The cut-off level was $w=0.86$. Обобщенное правдоподобие вычислялось по всем рефлексам в диапазоне разрешения $d=40-250$ Å. C_ϕ рассчитывался по рефлексам в диапазоне разрешения $d=48-68$ Å. Значения параметров для вычисления обобщенного правдоподобия: $B_{glob}=8000$, объем тестируемой маски области $V=0.3$ объема элементарной ячейки, число псевдоатомов $N=100$, уровень среза $w=0.86$. Треугольники соответствует вариантам, ближайшим к правильному решению.

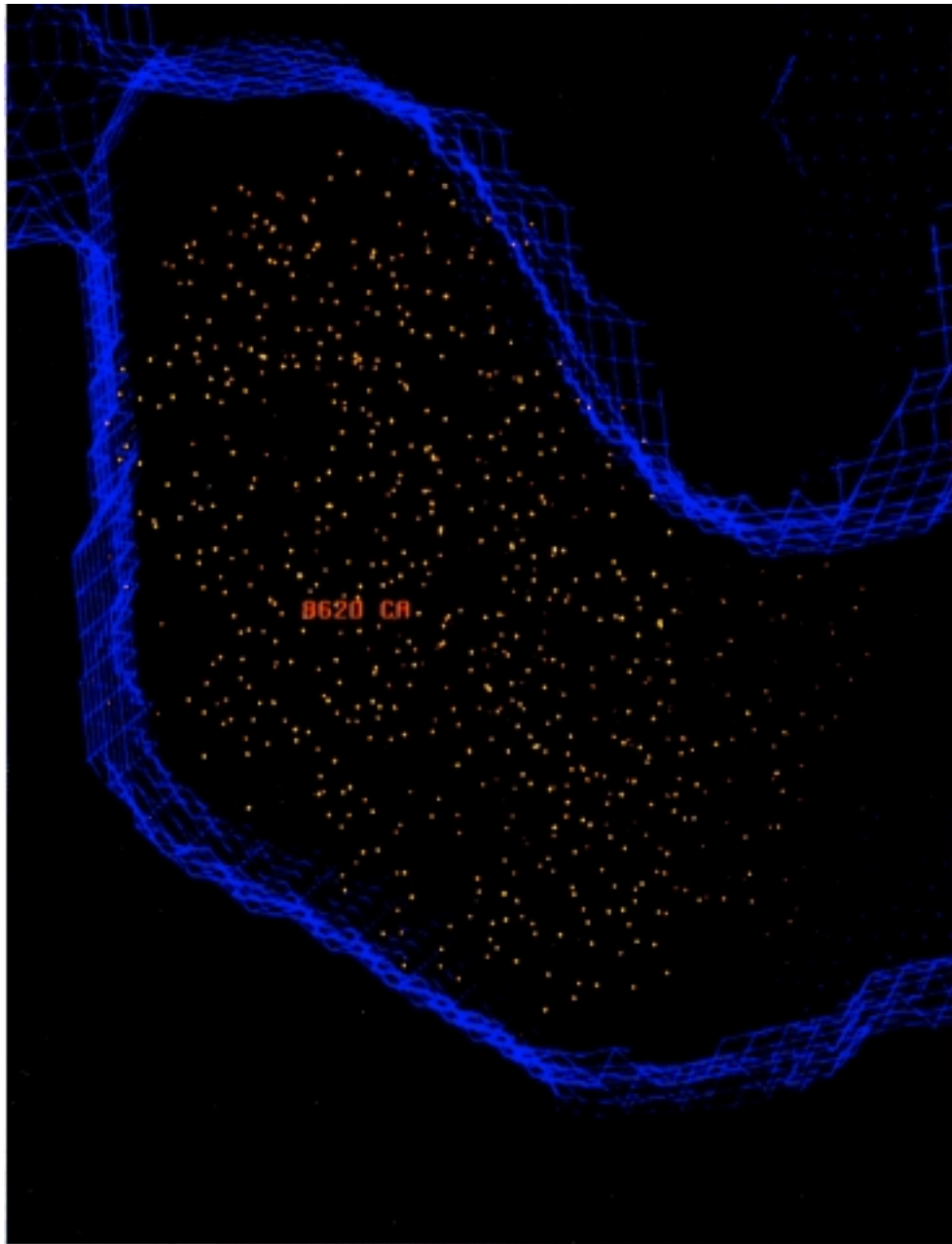
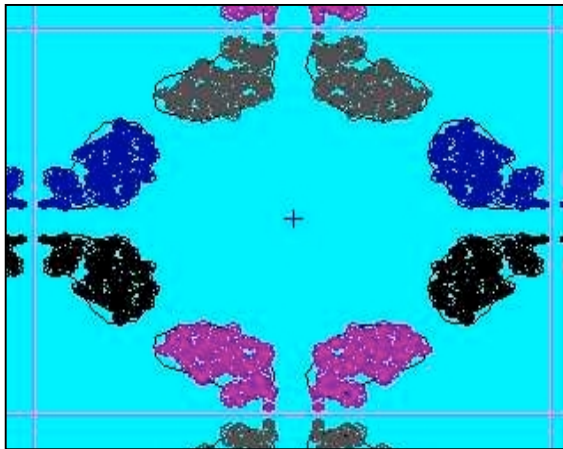
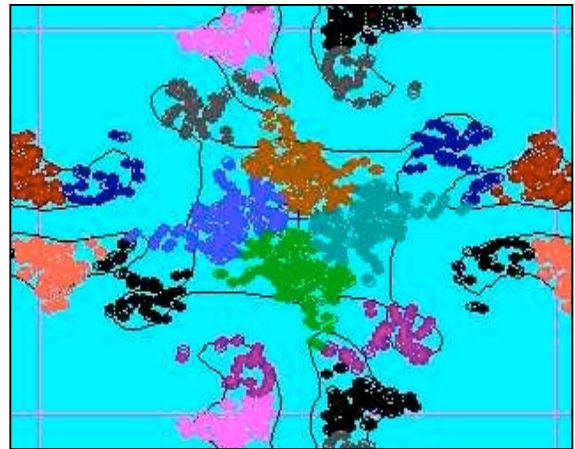


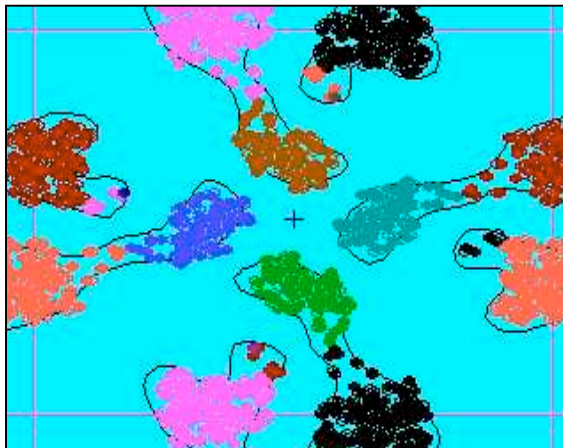
Рисунок 4.11 Синтез, построенный по экспериментальным модулям и фазам, усредненным по отобранным вариантам с максимальными значениями обобщенного правдоподобия при первом расширении фазового набора, и атомы модели для комплекса аминоксил-тРНК-синтетазы с тРНК.



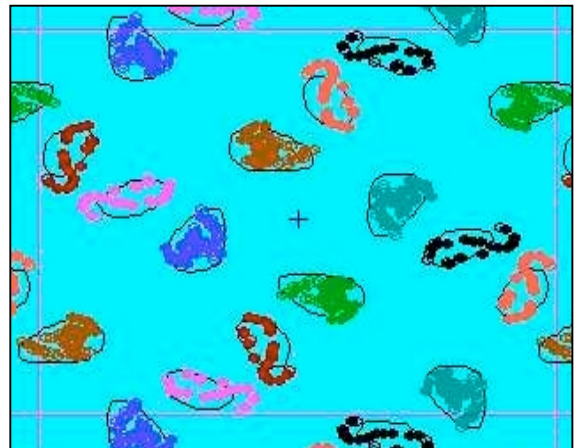
Section Z = 00/40



Section Z = 04/40



Section Z = 07/40



Section Z = 10/40

Рисунок 4.12 Синтез, построенный по экспериментальным модулям и фазам, усредненным по отобранным вариантам с максимальными значениями обобщенного правдоподобия при втором расширении фазового набора, и атомы модели для комплекса аминокил-тРНК-синтетазы с тРНК. Черный контур выделяет на синтезе область, содержащую 0.30% точек с максимальными значениями синтеза.

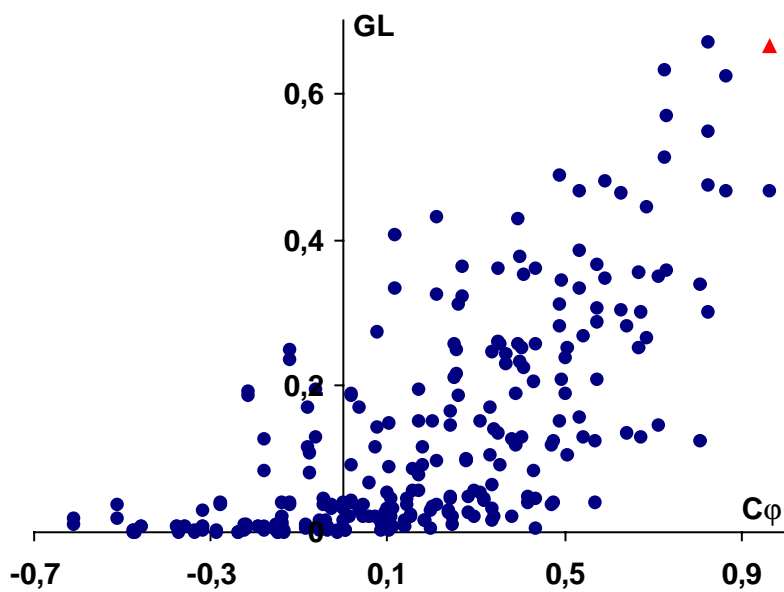
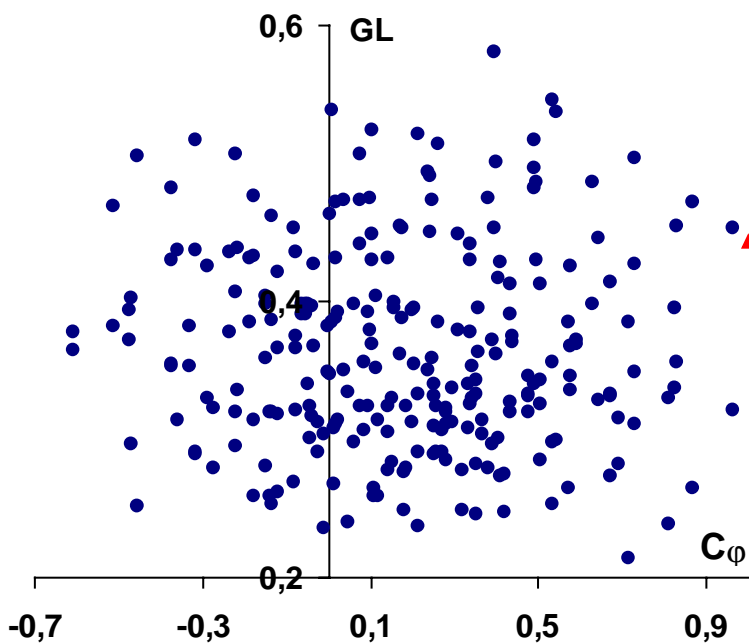
4.6.5 Зависимость результатов от зоны разрешения, по которой рассчитывается правдоподобие

При расчете значения правдоподобия необходимо задать набор рефлексов, по которым рассчитывается корреляция модулей (2.4). Иными словами, необходимо определить группу рефлексов, для которых определяется вероятность того, что рассчитанные модули близки к экспериментальным значениям. Как показывают тесты, зона разрешения, содержащая этот набор рефлексов, является параметром, наиболее сильно влияющим на результаты. На рис.4.13 (а-б) представлены результаты тестов, в которых отличались только зоны разрешения для расчета обобщенного правдоподобия. Корреляция между качеством фазового набора и значением обобщенного правдоподобия наблюдалась, если правдоподобие рассчитывалось по зоне, включающей рефлексы с перебираемыми фазами, а также некоторое количество рефлексов более высокой зоны разрешения. Если же корреляция модулей рассчитывалась только по зоне, не включающей рефлексы с перебираемыми фазами, то корреляция между качеством фазового набора и правдоподобием отсутствовала. Такой вывод был получен на основании результатов тестов со всеми тремя структурами.

4.7 Заключительные замечания

В предлагаемом подходе проблема определения *ab-initio* фаз низкого разрешения была переформулирована как проблема поиска области молекулы в элементарной ячейке. В качестве критерия отбора лучших предполагаемых областей молекулы среди всех рассматриваемых областей использовался критерий обобщенного правдоподобия. Целью данного исследования было определить, позволяет ли критерий обобщенного правдоподобия найти лучший набор фаз.

Во всех проведенных тестах лучшим фазовым наборам соответствовали высокие значения правдоподобия. Однако, однозначной зависимости между значением правдоподобия и качеством фазового



б)

Рисунок 4.13 Зависимость распределения значений фазовой корреляции и обобщенного правдоподобия от зоны разрешения рефлексов, по которым рассчитывается обобщенное правдоподобие. Комплекс аминоксил-тРНК-синтетазы с тРНК. Маски рассчитывались по фазовым наборам для рефлексов разрешения $d=48-250$ А. Обобщенное правдоподобие рассчитывалось по рефлексам зоны разрешения: а) $d=36-48$ А б) $d=40-250$ А.

набора выявлено не было. Наряду с хорошими вариантами высокими значениями правдоподобия обладали и варианты с низким значением коэффициента корреляции с правильными фазами. В случае с данными для комплекса аминоксил-тРНК-синтетазы с тРНК процедура расширения фазового набора позволила выделить лучшее решение среди двадцати вариантов с самым высоким значением правдоподобия на разрешении $d=68 \text{ \AA}$ и расширить полученное решение до разрешения $d=48 \text{ \AA}$. Подчеркнем, однако, что такая процедура сработала только в единичном случае комплекса аминоксил-тРНК-синтетазы с тРНК. Возможными причинами являются высокое качество низкоугловых экспериментальных данных, полученных для этого комплекса, пространственная группа I432, в которой неопределенным является только выбор энантиомера и нет проблем с возможными сдвигами начала координат, компактная упаковка данного комплекса в элементарной ячейке и высокое содержание растворителя.

Тем не менее, случайная генерация большого количества фазовых наборов и отбор вариантов с высоким значением правдоподобия позволили получить для всех тестируемых структур ансамбль вариантов с более высокой концентрацией хороших вариантов, чем среди всех сгенерированных вариантов. Усреднение отобранных вариантов дает вариант лучшего качества, чем усреднение всех первоначально сгенерированных фазовых наборов. Полученное в результате усреднения решение можно рассматривать как первое приближение к решению фазовой проблемы на низком разрешении. Усреднение решения, полученного с помощью критерия обобщенного правдоподобия, с решением, полученным с помощью критерия связности, позволяет улучшить полученное решение. Необходимы дальнейшие исследования по нахождению оптимального способа сочетания этих двух критериев.

ГЛАВА V. ПОЛУЧЕНИЕ АПРИОРНОГО РАСПРЕДЕЛЕНИЯ КООРДИНАТ АТОМОВ АНАЛИТИЧЕСКИМ ПУТЕМ

5.1 Постановка проблемы о поиске априорного распределения, соответствующего максимуму правдоподобия, аналитическим путем

Рассматриваемые в предыдущих главах задачи поиска лучшей сферической области и поиска лучшего фазового набора в конечном итоге сводились к задаче поиска такого априорного распределения координат атомов (из некоторого класса априорных распределений), которому бы соответствовало максимальное значение функции обобщенного правдоподобия, значение которой оценивалось в ходе численного эксперимента по формулам (2.3)-(2.4). Задачу поиска априорного распределения, удовлетворяющего принципу максимума правдоподобия, можно было бы пытаться решать аналитическим путем. Первая серьезная трудность, встающая на этом пути - это получение приближенных формул для совместного распределения структурных факторов (1.24). Для больших отклонений унитарных структурных факторов от ожидаемых значений $\Delta U \sim 1$ приближенное выражение для интеграла (1.24) было получено Бриконом (Bricogne, 1984) с помощью метода перевала. Переходя от обычных структурных факторов к унитарным структурным факторам, полученную им аппроксимацию можно представить в виде:

$$P(\mathbf{U}) = \frac{N^{M/2}}{(2\pi)^M} b^0(\mathbf{U}) \exp(Na^0(\mathbf{U})) \quad (5.1)$$

где M - количество структурных факторов, а функции выражаются через неявные функции, являющиеся решением системы нелинейных уравнений. Последнее обстоятельство существенно затрудняет применение этой формулы. Для случая, когда отклонения унитарных структурных факторов от ожидаемых значений $\Delta U \sim N^\epsilon$ были получены приближенные формулы

для $a^0(\mathbf{U})$ и $b^0(\mathbf{U})$, не содержащие неявных функций (Приложение 1). При таких ограничениях на порядок отклонения унитарных структурных факторов от ожидаемых значений достаточно большие значения структурных факторов не исключаются из рассмотрения.

5.2 Приближение априорного распределения градиентом функции правдоподобия

На низком разрешении карта априорного распределения может быть непосредственно использована для определения предполагаемых областей высокой и низкой концентрации атомов. Учитывая сложность нахождения априорного распределения, соответствующего глобальному максимуму функции правдоподобия, была предпринята попытка решить сначала более простую задачу: найти такое распределение $q(\mathbf{r})$, которому бы соответствовало значение правдоподобия, превышающее значение правдоподобия для равномерного распределения (1.2). Получить такое распределение можно, двигаясь из точки, соответствующей равномерному распределению в пространстве априорных распределений, вдоль градиента функции правдоподобия. Таким образом, искомое распределение ищется в виде:

$$q_\lambda(\mathbf{r}) = q_0(\mathbf{r}) + \lambda \frac{\delta L}{\delta q(\mathbf{r})} \quad (5.2)$$

Где значение λ достаточно мало. Задача определения функции $q(\mathbf{r})$ эквивалентна задаче определения его коэффициентов Фурье. Градиент функции правдоподобия может быть выражен через производные функции правдоподобия по коэффициентам Фурье априорного распределения координат атомов:

$$\frac{\delta L}{\delta q(\mathbf{r})} = \sum_{\mathbf{h}} \frac{\partial L}{\partial G_{\mathbf{h}}} \exp(2\pi i(\mathbf{h}, \mathbf{r})) \quad (5.3)$$

Необходимо подчеркнуть, что градиент функции правдоподобия (5.3) будет отличен от нуля, только если предварительно будут зафиксированы для

данной пространственной группы начало координат и энантиоморф. Из формулы (5.2) легко видеть, что максимумы и минимумы функций $q_\lambda(\mathbf{r})$

и $\frac{\partial L}{\partial G_{\mathbf{h}}}$ совпадают. Следовательно, для того, чтобы выявить в элементарной

ячейке предполагаемые области максимальной и минимальной концентрации атомов, достаточно построить карту градиента функции правдоподобия, рассчитав ее по формуле (5.3).

Используя асимптотическую оценку для совместного распределения структурных факторов (Приложение 1), были получены выражения для производных функции правдоподобия по коэффициентам Фурье априорного распределения $G(\mathbf{h})$ в случае, когда фаза рефлекса \mathbf{h} является абсолютным семиинвариантом (см. Приложение 2). Переписывая в других обозначениях формулы (П2.31) и (П2.32), имеем:

$$\frac{\partial L}{\partial G_{\mathbf{h}}} = \begin{cases} M(\mathbf{h}) & \mathbf{h} \notin S \\ E_{\mathbf{h}}^2 M(\mathbf{h}) + \omega(\mathbf{h}, \mathbf{h})(E_{\mathbf{h}}^4 - 3) & \mathbf{h} \in S \end{cases} \quad (5.4)$$

где

$$M(\mathbf{h}) = \sum_{\mathbf{k} \neq \mathbf{h}} \omega(\mathbf{k}, \mathbf{h})(E_{\mathbf{k}}^2 - 1) \quad (5.5)$$

$$\omega(\mathbf{k}, \mathbf{h}) = \frac{\left\{ \begin{array}{l} \text{number of } (\mu, \nu): \\ \mathbf{R}_{\mu}^T \mathbf{k} + \mathbf{R}_{\nu}^T \mathbf{k} + \mathbf{h} = 0 \end{array} \right\}}{2 \left\{ \begin{array}{l} \text{number of } \mu: \\ \mathbf{R}_{\mu}^T \mathbf{k} = \mathbf{k} \end{array} \right\} \left\{ \begin{array}{l} \text{number of } \nu: \\ \mathbf{R}_{\nu}^T \mathbf{h} = \mathbf{h} \end{array} \right\}} \quad (5.6)$$

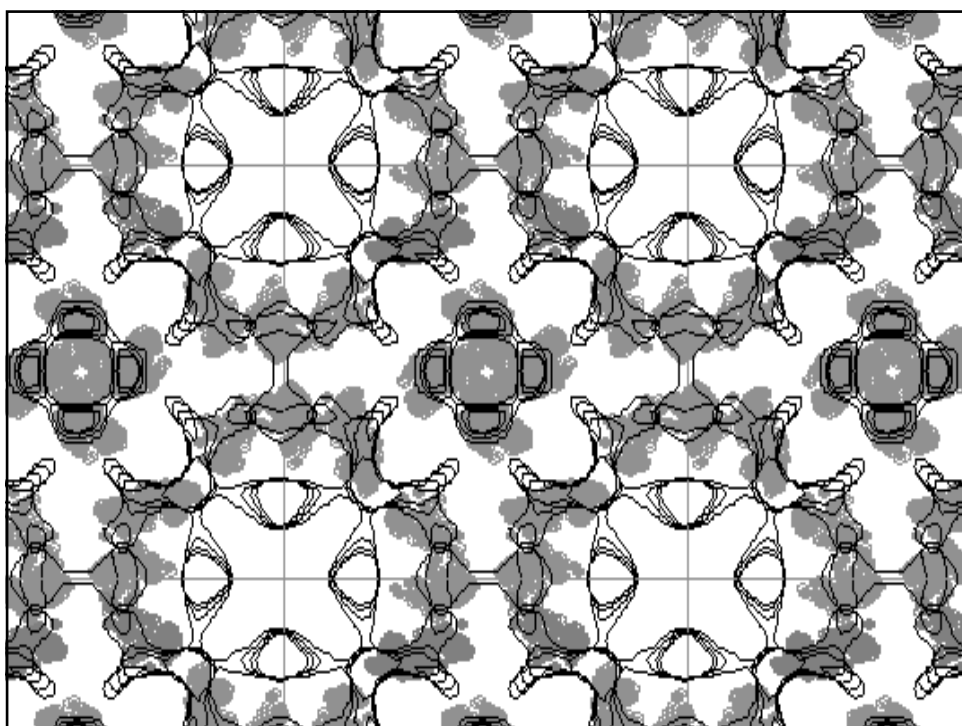
где $E_{\mathbf{h}}$ - нормализованный структурный фактор, а \mathbf{R}_{μ} - матрица симметрии.

5.3 Карты градиента функции правдоподобия для комплекса аминоксил-тРНК-синтетазы с тРНК

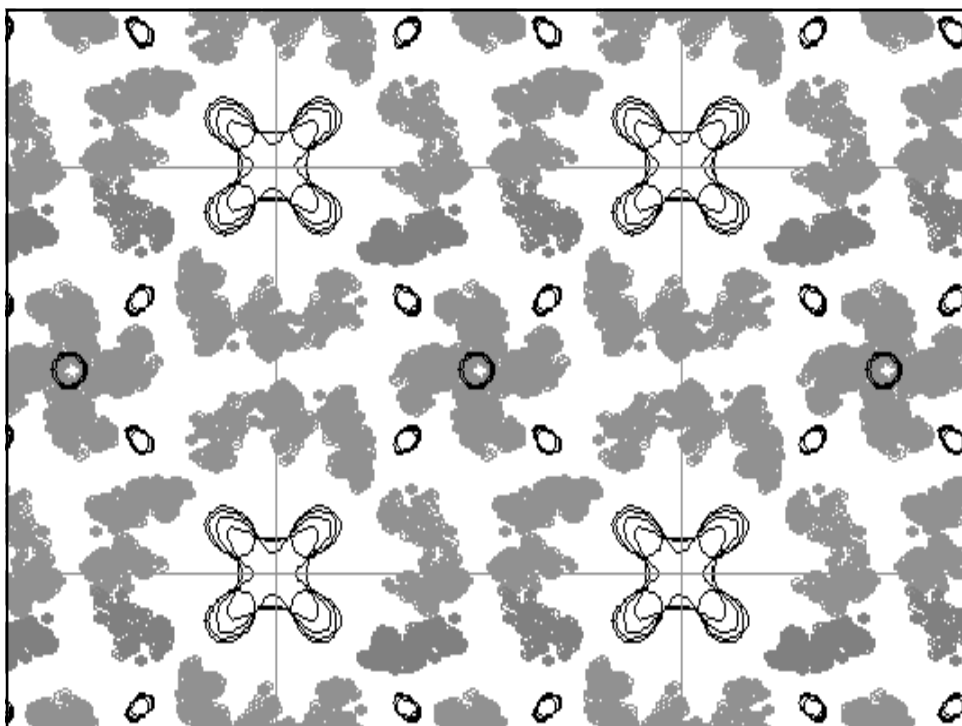
Для построения карт градиента функции правдоподобия были использованы данные нейтронного рассеяния для комплекса комплекса аминоксил-тРНК-синтетазы с тРНК (см. п.2.5.2) В пространственной группе I432 все центросимметричные рефлексы являются абсолютными семиинвариантами. Были выбраны 24 центросимметричных рефлекса, с использованием которых и были построены карты градиента функции правдоподобия. На рис.5.1 (а,б) Приведены карты градиента функции правдоподобия, рассчитанные по формулам (5.3)-(5.6) и положения C_{α} -атомов модели. Из рис.5.1 видно, что рассчитанная функция имеет максимумы в областях, где находятся атомы, а в областях, где атомов почти нет – минимумы. Однако, наблюдаются также дополнительные максимумы градиента функции правдоподобия и в тех частях элементарной ячейки, где концентрация атомов незначительна.

5.4 Заключительные замечания

Использование аналитического подхода к максимизации функции правдоподобия является альтернативой компьютерному Монте-Карловскому моделированию и позволяет достичь некоторого прогресса. Однако, сложность математического аппарата делает цену этого продвижения достаточно высокой. В связи с этим основные усилия были сконцентрированы на развитии компьютерных процедур, реализующих расчет обобщенного правдоподобия и его максимизацию. На сегодняшний день именно этот путь представляется наиболее эффективным.



а)



б)

Рисунок 5.1 Карта градиента функции правдоподобия и атомы модели для комплекса аминокил-тРНК-синтетазы с тРНК. Черный контур выделяет: (а) 25% точек ячейки с максимальными значениями, (б) 5% точек ячейки с минимальными значениями функции градиента правдоподобия.

ПРИЛОЖЕНИЕ 1. ВЫВОД ФОРМУЛЫ ДЛЯ СОВМЕСТНОГО РАСПРЕДЕЛЕНИЯ СТРУКТУРНЫХ ФАКТОРОВ

П1.1 Обозначения

Исследуемая структура рассматривается, в рамках статистического подхода (см. п.2.2), как одна из реализаций серий независимых испытаний, каждое из которых состоит в том, что любой атом структуры случайно и независимо помещается в элементарную ячейку с некоторой плотностью априорного распределения вероятностей $g(\mathbf{x})$. Обозначим $\mathbf{x}_1, \dots, \mathbf{x}_N$ 3-мерные координаты N атомов исследуемой структуры, а $\xi^1(\mathbf{x}), \dots, \xi^M(\mathbf{x})$ – тригонометрические структурные факторы, определенные по формулам (1.20)-(1.22). Сразу отметим, что мы будем рассматривать случай когда тригонометрические структурные факторы действительны. Это с одной стороны упростит выкладки, а с другой стороны, такое предположение справедливо для centrosymmetric структурных факторов группы I432, для которых будут применены выведенные ниже формулы. Структурные факторы для такой модели становятся случайными переменными и выражаются через тригонометрические структурные факторы

$$F^j = \sum_{k=1}^N f^{jk} \xi^j(\mathbf{x}_k), \quad j=1, \dots, M, \quad (\text{П1.1})$$

где предполагается, что $\{f^{jk}\}, j = 1, \dots, M; k = 1, \dots, N$ - известные коэффициенты.

Введем обозначения для смешанных моментов порядка j ($j = \alpha_1 + \dots + \alpha_j$) случайных величин $\xi^1(\mathbf{x}), \dots, \xi^M(\mathbf{x})$:

$$m^{\alpha_1 \dots \alpha_j} = \left\langle \xi^{\alpha_1}(\mathbf{x}) \dots \xi^{\alpha_j}(\mathbf{x}) \right\rangle_{g(\mathbf{x})} = \int_V \xi^{\alpha_1}(\mathbf{x}) \dots \xi^{\alpha_j}(\mathbf{x}) g(\mathbf{x}) d^3 \mathbf{x} \quad (\text{П1.2})$$

Введем также обозначения для их семиинвариантов, первые из которых выражаются через моменты (П1.2) по формулам:

$$\begin{aligned} K^\alpha &= m^\alpha, \\ K^{\alpha\beta} &= m^{\alpha\beta} - m^\alpha m^\beta, \end{aligned} \quad (\text{П1.3})$$

$$\kappa^{\alpha\beta\gamma} = m^{\alpha\beta\gamma} - \left[m^{\alpha} m^{\beta\gamma} + m^{\beta} m^{\alpha\gamma} + m^{\gamma} m^{\alpha\beta} \right] + 2m^{\alpha} m^{\beta} m^{\gamma} ,$$

А последующие могут быть рассчитаны шаг за шагом путем дифференцирования логарифма производящей функции моментов функции:

$$\kappa^{\alpha_1 \dots \alpha_j} = \left. \frac{\partial^j}{\partial t_{\alpha_1} \dots \partial t_{\alpha_j}} \ln M(\mathbf{t}) \right|_{\mathbf{t}=\mathbf{0}} , \quad (\text{П1.4})$$

где функция $M(\mathbf{t})$ разлагается в ряд

$$M(\mathbf{t}) = 1 + \sum_{j=1}^{\infty} \frac{1}{j!} m^{\alpha_1 \dots \alpha_j} t_{\alpha_1} \dots t_{\alpha_j} , \quad (\text{П1.5})$$

Из (П1.5) следует, что:

$$M(\mathbf{0}) = 1 , \quad \left. \frac{\partial^j}{\partial t_{\alpha_1} \dots \partial t_{\alpha_j}} M(\mathbf{t}) \right|_{\mathbf{t}=\mathbf{0}} = m^{\alpha_1 \dots \alpha_j} . \quad (\text{П1.6})$$

Производящая функция моментов связана с характеристической функцией:

$$M(\mathbf{t}) = C(-it) \quad (\text{П1.7})$$

Предполагается, что выполняются соглашения, принятые в тензорной алгебре, а именно, если некоторый индекс появляется в выражении дважды, один раз как нижний индекс, и один – как верхний, то это означает, что по этому индексу производится суммирование. Знак суммирования при этом опускается т.е.

$$a_{\alpha} b^{\alpha} = \sum_{\alpha=1}^M a_{\alpha} b^{\alpha} .$$

Определим также величины, выражающиеся через семиинварианты:

$$\sigma^{\alpha_1 \dots \alpha_j} = \left(\frac{1}{N} \sum_{k=1}^N f^{\alpha_1 k} \dots f^{\alpha_j k} \right) \cdot \kappa^{\alpha_1 \dots \alpha_j} , \quad (\text{П1.8})$$

и связанные через (П1.3) и (П1.2) с плотностью априорного распределения $g(\mathbf{x})$. Введенные коэффициенты симметричны по отношению к перестановке индексов. Предполагается также, что выполняется условие несингулярности

матрицы $\left((\sigma^{\alpha\beta}) \right)_{\alpha, \beta=1}^M$ т.е.:

$$D = \det\left(\left(\sigma^{\alpha\beta}\right)\right) \neq 0 \quad . \quad (\text{П1.9})$$

П1.2 Аппроксимация совместного распределения структурных факторов

Независимость случайных переменных позволяет, используя аппарат характеристических функций, представить совместное распределение величин структурных факторов в виде обратного преобразования Фурье от произведения характеристических функций:

$$\Psi^{-1}\left[C^N\right] = \Psi^{-1}\left[M^N(it)\right] = \Psi^{-1}\left[\exp\{N \ln(M(it))\}\right] \quad (\text{П1.10})$$

Если ввести функцию:

$$S(it) = \ln(M(it)) \quad (\text{П1.11})$$

то совместное распределение для структурных факторов имеет вид:

$$P(\mathbf{F}) = \frac{1}{(2\pi)^M} \int_{\mathbf{R}^M} \exp[NS(it) - it \cdot \mathbf{F}] d^M \mathbf{t}, \quad (\text{П1.12})$$

где $\mathbf{F} = (F^1, \dots, F^M)$ - вектор структурных факторов, а функция $S(\mathbf{t})$, учитывая (П1.11) и (П1.5) представима в виде ряда:

$$S(\mathbf{t}) = \sum_{j=1}^{\infty} \frac{1}{j!} \sigma^{\alpha_1 \dots \alpha_j} t_{\alpha_1} \dots t_{\alpha_j}, \quad (\text{П1.13})$$

связана с величинами $\xi^l(\mathbf{x}), \dots, \xi^M(\mathbf{x})$ и плотностью априорного распределения $g(\mathbf{x})$ координат атомов посредством соотношений (П1.8)-(П1.4)-(П1.3)-(П1.2).

Идея получения оценок при больших значениях N интеграла типа (П1.12) состоит в том, что функция $S(it)$ имеет один глобальный максимум, и основной вклад в интеграл дает малая окрестность точки максимума функции $S(it)$. Такое приближение будет тем точнее, чем больше значение N . В окрестности точки максимума подынтегральная функция раскладывается в ряд и интегрирование (П1.12) заменяется почленным интегрированием полученного ряда, которое можно выполнить

аналитически. Осуществление этой идеи не вызывает трудностей для случая конечных \mathbf{F} .

Более сложный случай возникает, когда ожидаемые значения компонент вектора структурных факторов \mathbf{F} увеличиваются пропорционально \sqrt{N} при возрастании N . В этом случае удобнее перейти к нормализованным структурным факторам. Тогда

$$\mathbf{F} = \sqrt{N} \mathbf{E} \quad (\text{П1.14})$$

где \mathbf{E} - некоторые конечные величины. Учитывая (П1.12) и (П1.14), совместное распределение нормализованных структурных факторов будет иметь вид:

$$P(\mathbf{E}) = \frac{N^{M/2}}{(2\pi)^M} \int_{\mathbf{R}^M} \exp[NS(it) - i\sqrt{N}\mathbf{t} \cdot \mathbf{E}] d^n \mathbf{t}. \quad (\text{П1.15})$$

Присутствие в подинтегральном выражении быстроосциллирующей функции $\exp[i\sqrt{N}\mathbf{t}]$ создает некоторые дополнительные трудности, которые, тем не менее, могут быть преодолены аналитически. Получившаяся в результате аппроксимация совместного распределения имеет вид многомерного Гауссова распределения, умноженного на полиномиальный ряд (Klug, 1958).

Еще более сложной является получение оценки для интеграла (П1.12) в случае, если ожидаемые значения компонент вектора структурных факторов \mathbf{F} увеличиваются пропорционально N при возрастании N . В этом случае удобнее перейти к унитарным структурным факторам \mathbf{U} .

Обозначим:

$$\mathbf{F} = N\mathbf{U} \quad (\text{П1.16})$$

где \mathbf{U} - некоторые конечные величины. Совместное распределение унитарных структурных факторов будет иметь вид:

$$P(\mathbf{U}) = \frac{N^M}{(2\pi)^M} \int_{\mathbf{R}^M} \exp\{N[S(it) - i\mathbf{t} \cdot \mathbf{U}]\} d^M \mathbf{t}. \quad (\text{П1.17})$$

Оценка для интеграла (П1.17) была получена Бриконом “методом перевала” (Bricogne, 1984). Для получения оценки значения интеграла (П1.17) используется возможность продолжить функцию $S(it)$ в многомерное комплексное пространство и изменить контур интегрирования таким образом, чтобы можно было применить метод Лапласа (Федорюк, 1987). Вводя комплексную переменную \mathbf{t} :

$$\mathbf{t} = \mathbf{u} - i\mathbf{v} \quad (\text{П1.18})$$

перепишем совместное распределение унитарных структурных факторов:

$$P(\mathbf{U}) = \frac{N^M}{(2\pi)^M} \int_{\mathbf{R}^M} \exp\{N[S(i\mathbf{u} + \mathbf{v}) - (i\mathbf{u} + \mathbf{v}) \cdot \mathbf{U}]\} d^M \mathbf{u}, \quad (\text{П1.19})$$

для любых $\mathbf{v} \in \mathbf{R}^M$. Свобода в выборе пути интегрирования позволяет погасить нежелательные осцилляции $\exp[iN\mathbf{u} \cdot \mathbf{U}]$ благодаря наличию $S(i\mathbf{u} + \mathbf{v})$ при $\mathbf{v} \neq 0$. На пути с минимальными осцилляциями опять можно применить метод Лапласа для оценки интеграла. Подробное доказательство правомерности такого перехода можно найти в работах Брикона (Bricogne, 1984, 1988).

Если ввести обозначения:

$$H(\mathbf{t}, \mathbf{U}) = S(it) - it \cdot \mathbf{U}, \quad (\text{П1.20})$$

$$S^{\alpha_1 \dots \alpha_j}(\mathbf{t}) = \frac{\partial^j}{\partial t_{\alpha_1} \dots \partial t_{\alpha_j}} S(\mathbf{t}), \quad (\text{П1.21})$$

тогда (П1.19) можно переписать в виде:

$$P(\mathbf{U}) = \frac{N^M}{(2\pi)^M} \int_{\mathbf{R}^M} \exp[NH(\mathbf{u} - i\mathbf{v}, \mathbf{U})] d^M \mathbf{u}, \quad (\text{П1.22})$$

а производные

$$\frac{\partial}{\partial u_\alpha} H(\mathbf{u} - i\mathbf{v}, \mathbf{U}) = i[S^\alpha(i\mathbf{u} + \mathbf{v}) - U^\alpha], \quad (\text{П1.23})$$

$$\frac{\partial}{\partial v_\alpha} H(\mathbf{u} - i\mathbf{v}, \mathbf{U}) = S^\alpha(i\mathbf{u} + \mathbf{v}) - U^\alpha. \quad (\text{П1.24})$$

Вектор \mathbf{v} выбирается таким образом, для того, чтобы производные (П1.21)- (П1.22) были равны нулю в точке $\mathbf{u}=\mathbf{0}$. Вводится вектор-функция $\boldsymbol{\tau}(\mathbf{U}) = (\tau_1(\mathbf{U}), \dots, \tau_M(\mathbf{U}))$, которая определяется неявно как решение системы нелинейных уравнений:

$$S^\alpha(\boldsymbol{\tau}(\mathbf{U})) = U^\alpha, \alpha = 1, \dots, M, \quad (\text{П1.25})$$

Тогда, в точке $\mathbf{u}=\mathbf{0}$, $\mathbf{v}=\boldsymbol{\tau}(\mathbf{U})$:

$$\frac{\partial}{\partial u_\alpha} H(\mathbf{u} - i\mathbf{v}, \mathbf{U}) \Big|_{\mathbf{u}=\mathbf{0}, \mathbf{v}=\boldsymbol{\tau}(\mathbf{U})} = \frac{\partial}{\partial v_\alpha} H(\mathbf{u} - i\mathbf{v}, \mathbf{U}) \Big|_{\mathbf{u}=\mathbf{0}, \mathbf{v}=\boldsymbol{\tau}(\mathbf{U})} = 0 \quad (\text{П1.26})$$

Функция $H(\mathbf{u} - i\boldsymbol{\tau}(\mathbf{U}), \mathbf{U})$ при разложении в ряд Тейлора в точке $\mathbf{u}=\mathbf{0}$ не содержит линейных членов:

$$H(\mathbf{u} - i\boldsymbol{\tau}(\mathbf{U}), \mathbf{U}) = a^0(\mathbf{U}) + \frac{1}{2} a^{\alpha\beta}(\mathbf{U}) u_\alpha u_\beta + \sum_{j=3}^{\infty} \frac{1}{j!} a^{\alpha_1 \dots \alpha_j}(\mathbf{U}) u_{\alpha_1} \dots u_{\alpha_j}, \quad (\text{П1.27})$$

Необходимо отметить, что коэффициенты разложения зависят как от значения \mathbf{U} , так и от неявной функции $\boldsymbol{\tau}(\mathbf{U})$:

$$a^0(\mathbf{U}) = S(\boldsymbol{\tau}(\mathbf{U})) - \mathbf{U} \cdot \boldsymbol{\tau}(\mathbf{U}), \quad (\text{П1.28})$$

$$a^{\alpha_1 \dots \alpha_j}(\mathbf{U}) = \frac{\partial^j}{\partial u_{\alpha_1} \dots \partial u_{\alpha_j}} h(\mathbf{u}, \mathbf{U}) \Big|_{\mathbf{u}=\mathbf{0}} = (i)^j S^{\alpha_1 \dots \alpha_j}(\boldsymbol{\tau}(\mathbf{U})) \quad (\text{П1.29})$$

После замены переменной интегрирования в (П1.22) $\mathbf{t} = \sqrt{N}\mathbf{u}$ и разложения экспоненты в ряд Тейлора:

$$\begin{aligned} & \exp \left[\frac{1}{\sqrt{N}} \sum_{j=3}^{\infty} \frac{1}{N^{(j-3)/2}} \frac{1}{j!} a^{\alpha_1 \dots \alpha_j}(\mathbf{U}) t_{\alpha_1} \dots t_{\alpha_j} \right] = \\ & = 1 + \frac{1}{\sqrt{N}} \frac{1}{6} a^{\alpha\beta\gamma}(\mathbf{U}) t_\alpha t_\beta t_\gamma + \frac{1}{N} \left\{ \frac{1}{24} a^{\alpha\beta\gamma\delta}(\mathbf{U}) t_\alpha t_\beta t_\gamma t_\delta + \frac{1}{2} \left[\frac{1}{6} a^{\alpha\beta\gamma}(\mathbf{U}) t_\alpha t_\beta t_\gamma \right]^2 \right\} + \dots \end{aligned}$$

(П1.30)

оценка для совместного распределения структурных факторов (П1.22) получается в виде:

$$P(\mathbf{U}) = \frac{N^{M/2}}{(2\pi)^M} \left\{ b^0(\mathbf{U}) + \frac{1}{N} b^1(\mathbf{U}) + \frac{1}{N^2} b^2(\mathbf{U}) + \dots \right\} \cdot \exp[Na^0(\mathbf{U})], \quad (\text{П1.31})$$

где $a^0(\mathbf{U})$ определяется по формуле (П1.28), а выражения для коэффициентов $b^0(\mathbf{U}), \dots, b^i(\mathbf{U})$ могут быть получены интегрированием, используя разложение (П1.30)

$$b^0(\mathbf{U}) = \int_{\mathbf{R}^M} \exp\left[\frac{1}{2} a^{\alpha\beta}(\mathbf{U}) t_\alpha t_\beta\right] d^M \mathbf{t} = (2\pi)^{M/2} \left[\det((-a^{\alpha\beta}(\mathbf{U}))) \right]^{-1/2}, \quad (\text{П1.32})$$

$$b^1(\mathbf{U}) =$$

$$\int_{\mathbf{R}^M} \exp\left[\frac{1}{2} a^{\alpha\beta}(\mathbf{U}) t_\alpha t_\beta\right] \left\{ \frac{1}{24} a^{\alpha\beta\gamma\delta}(\mathbf{U}) t_\alpha t_\beta t_\gamma t_\delta + \frac{1}{2} \left[\frac{1}{6} a^{\alpha\beta\gamma}(\mathbf{U}) t_\alpha t_\beta t_\gamma \right]^2 \right\} d^M \mathbf{t}$$

(П1.33)

Преимущество аппроксимации (П1.31) для совместного распределения структурных факторов состоит в том, что функции $b^1(\mathbf{U}), b^2(\mathbf{U}), \dots$ равномерно ограничены для всех возможных величин структурных факторов, и первый член ряда:

$$P(\mathbf{U}) \approx \frac{N^{M/2}}{(2\pi)^M} b^0(\mathbf{U}) \cdot \exp[Na^0(\mathbf{U})], \quad (\text{П1.33})$$

обеспечивает хорошее приближение для всех возможных значений модулей структурных факторов. Однако практическое использование формулы (П1.33)

осложняется тем, что выражения для $a^0(\mathbf{U})$ и $b^0(\mathbf{U})$ содержат неявную функцию, являющуюся решением системы нелинейных уравнений (П1.24). В

следующем разделе выводятся выражения для явного представления $a^0(\mathbf{U})$ и $b^0(\mathbf{U})$.

П1.3 Явное представление $a^0(\mathbf{U})$

Ниже предполагается, что отклонение унитарных структурных факторов от их математических ожиданий порядка $\Delta U \sim N^{-\varepsilon}$. При этом, по сравнению со случаем $\Delta U \sim 1/\sqrt{N}$, достаточно большие значения структурных факторов не исключаются из рассмотрения.

Для упрощения преобразований вводятся новые переменные $\Delta = (\Delta_1, \dots, \Delta_M)$, являющиеся решением системы линейных уравнений:

$$\sigma^{\alpha\beta} \Delta_\beta = U^\alpha - \sigma^\alpha, \quad \alpha = 1, \dots, M, \quad (\text{П1.34})$$

которая имеет решения благодаря условию (П1.10). Значения переменных Δ могут быть выражены через значения унитарных структурных факторов через матрицу $\left((v_{\alpha\beta}) \right)_{\alpha, \beta=1}^M$, обратную матрице $\left((\sigma^{\alpha\beta}) \right)$:

$$\Delta_\alpha = v_{\alpha\beta} (U^\beta - \sigma^\beta) \quad . \quad (\text{П1.35})$$

Выражения для функций $a^0(\mathbf{U})$ and $b^0(\mathbf{U})$ будем искать в виде:

$$a^0(\mathbf{U}) = A(\Delta(\mathbf{U})), \quad b^0(\mathbf{U}) = B(\Delta(\mathbf{U})), \quad (\text{П1.36})$$

где $A(\Delta)$ and $B(\Delta)$ функции, представимые в виде рядов Тейлора:

$$A(\Delta) = A(\mathbf{0}) + \sum_{j=1}^{\infty} \frac{1}{j!} A^{\alpha_1 \dots \alpha_j}(\mathbf{0}) \Delta_{\alpha_1} \dots \Delta_{\alpha_j} \quad , \quad (\text{П1.37})$$

$$B(\Delta) = B(\mathbf{0}) + \sum_{j=1}^{\infty} \frac{1}{j!} B^{\alpha_1 \dots \alpha_j}(\mathbf{0}) \Delta_{\alpha_1} \dots \Delta_{\alpha_j} \quad . \quad (\text{П1.38})$$

Ниже мы продолжаем использовать обозначения для производных:

$$A^{\alpha_1 \dots \alpha_j}(\mathbf{t}) = \frac{\partial^j}{\partial t_{\alpha_1} \dots \partial t_{\alpha_j}} A(\mathbf{t})$$

$$B^{\alpha_1 \dots \alpha_j}(\mathbf{t}) = \frac{\partial^j}{\partial t_{\alpha_1} \dots \partial t_{\alpha_j}} B(\mathbf{t}) \quad (\text{П1.39})$$

Определим функцию:

$$\mathbf{T}(\Delta) = \boldsymbol{\tau}(\mathbf{U}(\Delta)) \quad , \quad (\text{П1.40})$$

где $\boldsymbol{\tau}(\mathbf{U})$ является решением уравнений (4.11), а $\mathbf{U}(\Delta)$ определяется через (П1.34). Тогда $\mathbf{T}(\Delta)$ удовлетворяет следующей системе уравнений:

$$S^\alpha(\mathbf{T}(\Delta)) = \sigma^\alpha + \sigma^{\alpha\beta} \Delta_\beta \quad , \quad \alpha = 1, \dots, M \quad . \quad (\text{П1.41})$$

Вводя обозначения:

$$T_k^{\alpha_1 \dots \alpha_j}(\mathbf{t}) = \frac{\partial^j}{\partial t_{\alpha_1} \dots \partial t_{\alpha_j}} T_k(\mathbf{t}) \quad , \quad \omega_k^{\alpha_1 \dots \alpha_j} = T_k^{\alpha_1 \dots \alpha_j}(\mathbf{0}) \quad , \quad (\text{П1.42})$$

Из (П1.28) и (П1.34) получаем тождество:

$$A(\Delta) = S(\mathbf{T}(\Delta)) - (\sigma^\alpha + \sigma^{\alpha\beta} \Delta_\beta) T_\alpha(\Delta) \quad , \quad (\text{П1.43})$$

и, дифференцируя его по переменным Δ_j и учитывая (П1.41), получаем:

$$A^j(\Delta) = \left[S^\alpha(\mathbf{T}(\Delta)) - (\sigma^\alpha + \sigma^{\alpha\beta} \Delta_\beta) \right] T_\alpha^j(\Delta) - \sigma^{\alpha j} T_\alpha(\Delta) = -\sigma^{\alpha j} T_\alpha(\Delta) \quad , \quad (\text{П1.44})$$

Из последней формулы следуют соотношения:

$$A^{\alpha_1 \dots \alpha_j}(\Delta) = -\sigma^{\alpha\alpha_1} T_\alpha^{\alpha_2 \dots \alpha_j} \quad , \quad (\text{П1.45})$$

$$A^{\alpha_1 \dots \alpha_j}(\mathbf{0}) = -\sigma^{\alpha\alpha_1} \omega_\alpha^{\alpha_2 \dots \alpha_j} \quad , \quad (\text{П1.46})$$

Т.о. проблема определения неизвестных коэффициентов в разложении (П1.37) сводится к проблеме вычисления коэффициентов $\omega_k^{\alpha_1 \dots \alpha_j}$ разложения функций $T_k(\Delta)$ в ряд Тейлора:

$$T_k(\Delta) = T_k(\mathbf{0}) + \sum_{j=1}^{\infty} \frac{1}{j!} \omega_k^{\alpha_1 \dots \alpha_j} \Delta_{\alpha_j} \quad . \quad (\text{П1.47})$$

Из соотношения (П1.13) следует:

$$S^\alpha(\mathbf{0}) = \sigma^\alpha \quad , \quad S^{\alpha\beta}(\mathbf{0}) = \sigma^{\alpha\beta} \quad . \quad (\text{П1.48})$$

Вместе с условием (П1.9) это означает, что система уравнений (П1.41) имеет для малых Δ единственное решение $\mathbf{T}(\Delta)$, определенное дополнительным условием

$$\mathbf{T}(\mathbf{0}) = \mathbf{0} \quad . \quad (\text{П1.49})$$

Все производные функции $\mathbf{T}(\Delta)$ в точке $\Delta=0$ могут быть получены пошаговым дифференцированием соотношений:

$$S^{\alpha}(\mathbf{T}(\Delta)) = \sigma^{\alpha} + \sigma^{\alpha\beta} \Delta_{\beta} \quad , \quad \alpha = 1, \dots, M \quad . \quad (\text{П1.50})$$

Дифференцируя, получаем для всех α и j :

$$S^{\alpha\gamma}(\mathbf{T}(\Delta)) \cdot \mathbf{T}_{\gamma}^j(\Delta) = \sigma^{\alpha j} \quad . \quad (\text{П1.51})$$

Из последней формулы следует, что при $\Delta=0$:

$$-\sigma^{\alpha\gamma} \omega_{\gamma}^j = -\sigma^{\alpha j} \quad . \quad (\text{П1.52})$$

Учитывая условие (П1.9) получаем:

$$\omega_{\gamma}^j = \begin{cases} 1 & \text{for } j = \gamma \\ 0 & \text{for } j \neq \gamma \end{cases} \quad , \quad (\text{П1.53})$$

Дальнейшее дифференцирование (П1.51) по переменным Δ_k приводит к соотношениям:

$$S^{\alpha\gamma}(\mathbf{T}(\Delta)) \cdot \mathbf{T}_{\gamma}^{jk}(\Delta) + S^{\alpha\beta\gamma}(\mathbf{T}(\Delta)) \cdot \mathbf{T}_{\beta}^k(\Delta) \cdot \mathbf{T}_{\gamma}^j(\Delta) = 0 \quad , \quad (\text{П1.54})$$

при $\Delta=0$ получаем:

$$-\sigma^{\alpha\gamma} \omega_{\gamma}^{jk} = \sigma^{\alpha\beta\gamma} \omega_{\beta}^k \omega_{\gamma}^j = \sigma^{\alpha jk} \quad . \quad (\text{П1.55})$$

Процесс дифференцирования может быть продолжен и любые значения $\omega_{\alpha}^{j_1 \dots j_m}$ порядка m могут быть получены как функции величин ω и σ порядков $1, \dots, m-1$. Продолжим дифференцирование:

$$S^{\alpha\gamma}(\mathbf{T}(\Delta)) \cdot \mathbf{T}_{\gamma}^{jkl}(\Delta) + S^{\alpha\beta\gamma}(\mathbf{T}(\Delta)) \cdot \mathbf{T}_{\beta}^{jk}(\Delta) \cdot \mathbf{T}_{\gamma}^l(\Delta) + S^{\alpha\beta\gamma}(\mathbf{T}(\Delta)) \cdot \mathbf{T}_{\beta}^{jl}(\Delta) \cdot \mathbf{T}_{\gamma}^k(\Delta) \\ + S^{\alpha\beta\gamma}(\mathbf{T}(\Delta)) \cdot \mathbf{T}_{\beta}^{lk}(\Delta) \cdot \mathbf{T}_{\gamma}^j(\Delta) + S^{\alpha\beta\gamma\delta}(\mathbf{T}(\Delta)) \cdot \mathbf{T}_{\beta}^j(\Delta) \cdot \mathbf{T}_{\gamma}^k(\Delta) \cdot \mathbf{T}_{\delta}^l(\Delta) = 0 \quad ,$$

(П1.56)

Из последней формулы аналогично тому, как это было сделано выше, получаем:

$$\begin{aligned} -\sigma^{\alpha\gamma} \omega_{\gamma}^{jkl} &= \sigma^{\alpha\beta j} \omega_{\beta}^{kl} + \sigma^{\alpha\beta k} \omega_{\beta}^{jl} + \sigma^{\alpha\beta l} \omega_{\beta}^{jk} + \sigma^{\alpha jkl} = \\ &= \sigma^{\alpha jkl} - \nu_{\beta\gamma} \left(\sigma^{\alpha j\beta} \sigma^{kl\gamma} + \sigma^{\alpha k\beta} \sigma^{jl\gamma} + \sigma^{\alpha l\beta} \sigma^{jk\gamma} \right). \end{aligned} \quad (\text{П1.57})$$

Объединяя (П1.43), (П1.44), (П1.49), (П1.55) и (П1.57) получаем приближенное выражение для $A(\Delta)$ в виде:

$$\begin{aligned} A(\Delta) &= -\frac{1}{2} \sigma^{\alpha\beta} \Delta_{\alpha} \Delta_{\beta} + \frac{1}{6} \sigma^{\alpha\beta\gamma} \Delta_{\alpha} \Delta_{\beta} \Delta_{\gamma} \\ &+ \frac{1}{24} \left[\sigma^{\alpha\beta\gamma\delta} - \nu_{pq} \left(\sigma^{\alpha\beta p} \sigma^{\gamma\delta q} + \sigma^{\alpha\beta q} \sigma^{\gamma\delta p} + \sigma^{\alpha\beta p} \sigma^{\gamma\delta q} \right) \right] \Delta_{\alpha} \Delta_{\beta} \Delta_{\gamma} \Delta_{\delta} + \dots \end{aligned} \quad (\text{П1.58})$$

П1.4 Явное представление $b^0(\mathbf{U})$

Предэкспоненциальная функция $b^0(\mathbf{U})$ (П1.32) зависит от переменной неявно через решение системы уравнений (П1.24). Получим приближенное представление для этой функции в явной форме. Введем в рассмотрение матрицу $\left(\left(\nu_{\alpha\beta} \right) \right)$, обратную к матрице $\left(\left(\sigma^{\alpha\beta} \right) \right)$ и матрицу $\left(\left(\mu_{\beta}^{\alpha} \right) \right)$, являющуюся квадратным корнем из матрицы $\left(\left(\nu_{\alpha\beta} \right) \right)$. Тогда выполняется:

$$\sigma^{\alpha\beta} \mu_{\alpha}^j \mu_{\beta}^k = \begin{cases} 1 & \text{for } j = k, \\ 0 & \text{for } j \neq k \end{cases} \quad (\text{П1.59})$$

Подставляя (П1.29) в выражение (П1.32)

$$B(\Delta) = \int_{\mathbf{R}^M} \exp \left[-\frac{1}{2} S^{\alpha\beta} (\mathbf{T}(\Delta)) t_{\alpha} t_{\beta} \right] d^M \mathbf{t} \quad (\text{П1.60})$$

и, затем, дифференцируя, получаем:

$$B^j(\Delta) = -\frac{1}{2} \int_{\mathbf{R}^M} S^{\alpha\beta\gamma}(\mathbf{T}(\Delta)) \cdot T_\gamma^j(\Delta) \cdot t_\alpha t_\beta \exp\left[-\frac{1}{2} S^{\alpha\beta}(\mathbf{T}(\Delta)) t_\alpha t_\beta\right] d^M \mathbf{t},$$

(П1.61)

$$B^j(\mathbf{0}) = -\frac{1}{2} \sigma^{\gamma\delta j} \int_{\mathbf{R}^M} t_\gamma t_\delta \exp\left[-\frac{1}{2} \sigma^{\alpha\beta} t_\alpha t_\beta\right] d^M \mathbf{t} \quad . \quad (\text{П1.62})$$

После введения новых переменных интегрирования $t_\alpha = \mu_\alpha^\beta x_\beta$, интеграл (П1.62) преобразуется в к виду:

$$B^j(\mathbf{0}) = -\frac{1}{2\sqrt{D}} \sigma^{\alpha\beta j} \mu_\alpha^p \mu_\beta^q \int_{\mathbf{R}^M} x_p x_q \exp\left[-\frac{1}{2} \sum_{k=1}^M x_k^2\right] d^M \mathbf{x},$$

(П1.63)

где D определяется по формуле (П1.9). Учитывая, что значение интеграла (П1.63) равно нулю при $p \neq q$, получаем:

$$B^j(\mathbf{0}) = -\frac{(2\pi)^{M/2}}{2\sqrt{D}} \sigma^{\alpha\beta j} \mu_\alpha^p \mu_\beta^p = -\frac{(2\pi)^{M/2}}{2\sqrt{D}} v_{\alpha\beta} \sigma^{\alpha\beta j} \quad (\text{П1.64})$$

Из (П1.48) получаем:

$$B(\mathbf{0}) = \frac{(2\pi)^{M/2}}{2\sqrt{D}}, \quad (\text{П1.65})$$

Выписывая первые два члена разложения функции $B(\Delta)$ в ряд Тейлора, получаем выражение:

$$B(\Delta) = \frac{(2\pi)^{M/2}}{2\sqrt{D}} \cdot \left\{ 1 - \frac{1}{2} v_{\alpha\beta} \sigma^{\alpha\beta\gamma} \Delta_\gamma + \dots \right\}, \quad (\text{П1.66})$$

в котором следующие члены ряда могут быть получены аналогично тому, как это было сделано для $B^j(\mathbf{0})$.

ПРИЛОЖЕНИЕ 2. ПОЛУЧЕНИЕ АНАЛИТИЧЕСКИХ ВЫРАЖЕНИЙ ДЛЯ ПРОИЗВОДНЫХ ФУНКЦИИ ПРАВДОПОДОБИЯ ПО КОЭФФИЦИЕНТАМ ФУРЬЕ АПРИОРНОГО РАСПРЕДЕЛЕНИЯ КООРДИНАТ АТОМОВ

Предполагается, что каждый атом исследуемой структуры помещается в элементарную ячейку случайно и независимо с одним и тем же априорным распределением координат атомов $g(\mathbf{x})$:

$$g(\mathbf{x}) = \frac{1}{V} \sum G(\mathbf{h}) e^{-2\pi i(\mathbf{h}, \mathbf{x})} \quad (\text{П2.1})$$

где

$$G(\mathbf{o}) = 1 \quad (\text{П2.2})$$

и коэффициенты Фурье $G(\mathbf{h})$ обладают симметрией пространственной группы исследуемой структуры. В частном случае равномерного априорного распределения все коэффициенты $G(\mathbf{h})$ за исключением $G(\mathbf{0})$ равны нулю. Будем предполагать, что априорное распределение $g(\mathbf{x})$ не слишком сильно отличается от равномерного. Мы будем считать, что коэффициенты $G(\mathbf{h})$ малы, и при получении приближенных выражений для функции правдоподобия будем оставлять только линейные по $G(\mathbf{h})$ члены.

Упростим еще более рассматриваемую задачу. Будем рассматривать только centrosymmetric рефлексы $\mathbf{h} \in J$. В нашем случае такая задача имеет смысл так как полученные формулы будут применены к комплексу аминоацил-тРНК-синтетазы с тРНК (пространственная группа I432). На очень низком разрешении большая часть рефлексов в этой группе является centrosymmetric. Будем также учитывать, что в данной пространственной группе все преобразования симметрии содержат только матрицы вращения R_μ и не содержат векторов сдвига $t_\mu = 0$.

Получим приближенное выражение для совместного распределения набора centrosymmetric рефлексов $\mathbf{h} \in J$. Будем также предполагать, что все коэффициенты $G(\mathbf{h}) = 0$ если $\mathbf{h} \notin J$.

Приближенное выражения для совместного распределения структурных факторов было получено в Приложении 1. Оно имеет вид:

$$P(\{U\}) \propto b(\Delta) e^{Na_0(\Delta)} (1 + O(1/N)) \quad (\text{П2.3})$$

где выражение для $a_0(\Delta)$ дается формулой (П1.58), $b(\Delta)$ определяется по формуле (П1.66), а величины $\{\Delta_\alpha\}$ связаны с исходными переменными системой линейных уравнений (П1.34). Распишем выражение (П2.3) через коэффициенты априорного распределения $G(\mathbf{h})$.

Подставляя (П2.1) в выражение (П1.2) получаем смешанный момент случайных величин $\xi^\alpha(\mathbf{x}), \xi^\beta(\mathbf{x})$ для пространственной группы I432 в виде:

$$m^{\alpha\beta} = \frac{1}{n} \sum_{v=1}^n G(\mathbf{h}_\alpha + \mathbf{R}_v^T \mathbf{h}_\beta) \quad (\text{П2.4})$$

где n - число симметрий, а \mathbf{R}_μ - матрицы поворота. Если обозначить $\varepsilon(\mathbf{h})$ - число преобразований симметрии таких, что

$$\mathbf{R}_v^T(\mathbf{h}) = \mathbf{h} \quad (\text{П2.5})$$

то выражение (П2.4) можно переписать в виде:

$$m^{\alpha\beta} = \frac{\sqrt{\varepsilon_\alpha \varepsilon_\beta}}{n} [\delta_{\alpha\beta} + \chi_{\alpha\beta}] \quad (\text{П2.6})$$

где

$$\delta_{\alpha\beta} = \begin{cases} 1 & \alpha = \beta \\ 0 & \alpha \neq \beta \end{cases} \quad (\text{П2.7})$$

$$\chi_{\alpha\beta} = \frac{1}{\sqrt{\varepsilon_\alpha \varepsilon_\beta}} \sum_v G(\mathbf{h}_\alpha + \mathbf{R}_v^T \mathbf{h}_\beta) - \delta_{\alpha\beta} \quad (\text{П2.8})$$

Учитывая, что все коэффициенты $G(\mathbf{h})=0$ для $\mathbf{h} \notin J$, получаем:

$$\sum_v G(\mathbf{h}_\alpha + \mathbf{R}_v^T \mathbf{h}_\beta) = \sum_{j \in J} G_j \Omega \quad (\text{П2.9})$$

где

Ω - число таких преобразований симметрии V , для которых возможно представление с некоторым μ :

$$\mathbf{h}_\alpha + \mathbf{R}_\nu^T \mathbf{h}_\beta = \mathbf{R}_\mu^T \mathbf{h}_j \quad (\text{П2.10})$$

либо

$$\mathbf{h}_\alpha + \mathbf{R}_\nu^T \mathbf{h}_\beta = -\mathbf{R}_\mu^T \mathbf{h}_j \quad (\text{П2.11})$$

Заметим, что если при некотором μ выполняется равенство (П2.10), то так как рефлекс \mathbf{h}_j - центросимметричный, существует такое преобразование симметрии \mathbf{R}_λ^T , что $\mathbf{h}_j = -\mathbf{R}_\lambda^T \mathbf{h}_j$. После подстановки этого представления в правую часть (П2.10) и учитывая также, что $\mathbf{R}_\mu^T \mathbf{R}_\lambda^T = \mathbf{R}_\gamma^T$, где последнее преобразование симметрии принадлежит данной группе симметрии, получаем, что равенство (П2.11) можно убрать из определения для Ω и переформулировать его следующим образом:

Ω - число таких преобразований симметрии ν , для которых при некотором μ выполняется равенство:

$$\mathbf{h}_\alpha + \mathbf{R}_\nu^T \mathbf{h}_\beta + \mathbf{R}_\mu^T \mathbf{h}_j = 0 \quad (\text{П2.12})$$

Легко показать, что если для данного преобразования симметрии ν существует хотя бы одно преобразование \mathbf{R}_μ^T такое, что справедливо соотношение (П2.12), то таких преобразований симметрии существует ровно ε_j штук. Поэтому,

$$\Omega = 1/\varepsilon_j \{ \text{число таких пар } (\mu, \nu) \text{ таких, что выполняется (П2.12)} \}.$$

Условие (П2.12) эквивалентно условию:

$$\left(\mathbf{R}_\mu^T\right)^{-1} \mathbf{h}_\alpha + \left(\mathbf{R}_\mu^T\right)^{-1} \mathbf{R}_\nu^T \mathbf{h}_\beta + \mathbf{h}_j = 0 \quad (\text{П2.13})$$

Если ввести обозначение:

$$\omega(\alpha, \beta, \gamma) = \{ \text{число таких пар } (\mu, \nu) \text{ таких, что} \\ \mathbf{R}_\mu^T \mathbf{h}_\alpha + \mathbf{R}_\nu^T \mathbf{h}_\beta + \mathbf{h}_\gamma = 0 \} \quad (\text{П2.14})$$

тогда выражение (П2.8) можно окончательно переписать в виде:

$$\chi_{\alpha\beta} = \frac{1}{\sqrt{\varepsilon_\alpha \varepsilon_\beta}} \sum_{j \in J} \frac{1}{\varepsilon_j} \omega(\alpha, \beta, j) G_j \quad (\text{П2.15})$$

Тогда получаем следующее выражение для кумулянтов второго порядка:

$$\sigma^{\alpha\beta} = m^{\alpha\beta} - m^\alpha m^\beta = \frac{\sqrt{\varepsilon_\alpha \varepsilon_\beta}}{n} [\delta_{\alpha\beta} + \chi_{\alpha\beta}] + 0(G^2) \quad (\text{П2.16})$$

Поскольку предполагается, что априорное распределение координат атомов мало отличается от равномерного, для элементов матрицы ((V)), обратной к матрице ((σ)), будет справедливо соотношение:

$$v_{\alpha\beta} = \frac{n}{\sqrt{\varepsilon_\alpha \varepsilon_\beta}} [\delta_{\alpha\beta} - \chi_{\alpha\beta}] + 0(G^2) \quad (\text{П2.17})$$

Подставляя (П2.17) и (П1.35) в первые два слагаемых формулы (П1.58), а затем преобразовывая, получаем выражение для первого слагаемого:

$$\begin{aligned} \sigma^{\alpha\beta} \Delta_\alpha \Delta_\beta &= \sum_{j,k} \frac{n}{\sqrt{\varepsilon_j \varepsilon_k}} (\delta_{jk} - \chi_{jk} + 0(G^2)) (U_j - G_j)(U_k - G_k) = \\ &= n \sum_j \frac{1}{\varepsilon_j} U_j^2 - 2n \sum_j \frac{1}{\varepsilon_j} U_j G_j - n \sum_{\alpha\beta} \frac{1}{\varepsilon_\alpha \varepsilon_\beta} \sum_j \frac{1}{\varepsilon_j} \omega(\alpha, \beta, j) U_\alpha U_\beta G_j + 0(G^2) \end{aligned} \quad (\text{П2.18})$$

и для второго слагаемого формулы (П1.58):

$$\begin{aligned} \sigma^{\alpha\beta\gamma} \Delta_\alpha \Delta_\beta \Delta_\gamma &= \\ &= \sum_{\alpha,\beta,\gamma} \sum_{j,k,l} v_{\alpha j} v_{\beta k} v_{\gamma l} (U_j - G_j)(U_k - G_k)(U_l - G_l) = \\ &= n \sum_{j,k,l} \frac{\omega(j,k,l)}{\varepsilon_j \varepsilon_k \varepsilon_l} U_j U_k U_l - 3n \sum_{j,k,l} \frac{\omega(j,k,l)}{\varepsilon_j \varepsilon_k \varepsilon_l} U_k U_l G_j + 0(G)0(U^3) + 0(G^2) \end{aligned} \quad (\text{П2.19})$$

Суммируя (П2.18) и (П2.19), получаем выражение для $a_0(\{U\}, \{G\})$:

$$a_0 = n \left\{ -\frac{1}{2} \sum_\alpha \frac{1}{\varepsilon_\alpha} U_\alpha^2 + \sum_j \frac{1}{\varepsilon_j} U_j G_j + \frac{1}{2} \sum_{\alpha\beta} \sum_{j \in S} \frac{\omega(\alpha, \beta, j)}{\varepsilon_\alpha \varepsilon_\beta \varepsilon_j} U_\alpha U_\beta G_j + \right.$$

$$+ \sum_{\alpha, \beta, \gamma} \frac{\omega(\alpha, \beta, \gamma)}{\varepsilon_\alpha \varepsilon_\beta \varepsilon_\gamma} U_\alpha U_\beta U_\gamma + 0(G)0(U^3) + 0(G^2) \} \quad (П2.20)$$

Согласно введенным предположениям о порядке величин унитарных структурных факторов и коэффициентов Фурье априорного распределения, все слагаемые выражения (П2.20) начиная со второго можно считать малыми и после подстановки выражения (П2.20) в показатель экспоненты разложить ее в ряд Тейлора. В результате получим:

$$\begin{aligned} \exp\{Na_0\} = & \exp\left\{-\frac{Nn}{2} \sum_{\alpha} \frac{1}{\varepsilon_\alpha} U_\alpha\right\} \left(1 + \right. \\ & + Nn \sum_j \frac{1}{\varepsilon_j} U_j G_j + \frac{Nn}{2} \sum_{\alpha\beta} \sum_{j \in S} \frac{\omega(\alpha, \beta, j)}{\varepsilon_\alpha \varepsilon_\beta \varepsilon_j} U_\alpha U_\beta G_j + \\ & \left. + \frac{(Nn)^2}{6} \left(\sum_j \frac{1}{\varepsilon_j} U_j G_j \right) \sum_{\alpha, \beta, \gamma} \frac{\omega(\alpha, \beta, \gamma)}{\varepsilon_\alpha \varepsilon_\beta \varepsilon_\gamma} U_\alpha U_\beta U_\gamma + 0(G)0(U^3) + 0(G^2) \right) \end{aligned} \quad (П2.21)$$

Аналогично расписываем предэкспоненциальный множитель $b(\Delta)$ через значения унитарных структурных факторов и коэффициенты Фурье априорного распределения. Подставив (П1.35), (П2.16) и (П2.17) в (П1.66) и проведя тождественные преобразования, получаем:

$$\begin{aligned} B(U) = & \frac{(2\pi)^{M/2}}{2\sqrt{D}} \left\{ 1 - \frac{1}{2} \sum_{\alpha, j} \frac{\omega(\alpha, \alpha, j)}{\varepsilon_\alpha \varepsilon_j} U_j + \frac{1}{2} \sum_{\alpha, j} \frac{\omega(\alpha, \alpha, j)}{\varepsilon_\alpha \varepsilon_j} G_j + \right. \\ & \left. + 0(U^2) + 0(G)0(U) + 0(G^2) \right\} \end{aligned} \quad (П2.22)$$

Выражение для совместного распределения структурных факторов получается перемножением (П2.21) и (П2.22). С точностью до множителя, не зависящего от значений унитарных структурных факторов, оно будет иметь вид:

$$\begin{aligned}
P(U) \propto \exp \left\{ -\frac{Nn}{2} \sum_{\alpha} \frac{1}{\varepsilon_{\alpha}} U_{\alpha} \right\} & \left(1 + \right. \\
+ Nn \sum_j \frac{1}{\varepsilon_j} U_j G_j + \frac{Nn}{2} \sum_{\alpha\beta} \sum_{j \in S} \frac{\omega(\alpha, \beta, j)}{\varepsilon_{\alpha} \varepsilon_{\beta} \varepsilon_j} U_{\alpha} U_{\beta} G_j + & \\
+ \frac{(Nn)^2}{6} \sum_{\alpha, \beta, \gamma, j} \frac{\omega(\alpha, \beta, \gamma)}{\varepsilon_{\alpha} \varepsilon_{\beta} \varepsilon_{\gamma} \varepsilon_j} U_{\alpha} U_{\beta} U_{\gamma} U_j G_j - & \\
- \frac{Nn}{2} \sum_{\alpha, \beta, j} \frac{\omega(\alpha, \beta, j)}{\varepsilon_{\alpha} \varepsilon_{\beta} \varepsilon_j} U_{\beta} U_j G_j + \frac{1}{2} \sum_{\alpha, j} \frac{\omega(\alpha, \alpha, j)}{\varepsilon_{\alpha} \varepsilon_j} G_j + \dots & \quad (П2.23)
\end{aligned}$$

Для упрощения выкладок перейдем к нормализованным структурным факторам:

$$E_{\alpha} = Nn \frac{U_{\alpha}^2}{\varepsilon_{\alpha}} \quad (П2.24)$$

Выражение для функции правдоподобия имеет вид:

$$L = \frac{P(|E_1|, \dots, |E_M|)}{\int P(|E_1|, \dots, |E_M|) d|E_1| \dots d|E_M|} \quad (П2.25)$$

Интегрируя по неизвестным фазам выражение (П2.23), получаем с точностью до множителя распределение модулей нормализованных структурных факторов. Заметим, что так как все рассматриваемые рефлексы центросимметричны, то интегрирование в данном случае заменяется суммированием. Итак, приближенное выражение для числителя (П2.25) с точностью до знака имеет вид:

$$\begin{aligned}
P(|E_1|, \dots, |E_M|) \propto \exp \left\{ -\frac{1}{2} \sum_{\alpha} E_{\alpha}^2 \right\} & \left(1 + \frac{1}{2} \sum_{j \in S} \sum_{\alpha} \Xi(\alpha, j) E_{\alpha}^2 G_j + \right. \\
\frac{1}{2} \sum_{j \in S} \left(\sum_{\alpha} \Xi(\alpha, j) [(E_{\alpha}^2 - 1) E_j^2 + 1] \right) & G_j + \dots \quad (П2.26)
\end{aligned}$$

где введено обозначение:

$$\Xi(\alpha, j) = \frac{\omega(\alpha, \alpha, j)}{\varepsilon_{\alpha} \varepsilon_j} \quad (П2.27)$$

Чтобы получить выражение для знаменателя (П2.25), надо проинтегрировать (П2.26) по модулям нормализованных структурных факторов. Проведя необходимые выкладки, получаем выражение для знаменателя:

$$1 + \frac{1}{2} \sum_{j \in S} \sum_{\alpha} \Xi(\alpha, j) G_j + \frac{1}{2} \sum_j \sum_{\alpha \neq j} \Xi(\alpha, j) G_j + \frac{1}{2} \gamma_4 \sum_j \Xi(j, j) G_j, \quad (\text{П2.28})$$

где введено обозначение:

$$\gamma_4 = \frac{\int e^{-t^2/2} t^4 dt}{\int e^{-t^2/2} dt} \quad (\text{П2.29})$$

Учитывая малость коэффициентов Фурье G_j и разлагая в ряд Тейлора выражения (П2.28) в минус первой степени, окончательно получаем выражение для функции правдоподобия:

$$L = \exp \left\{ -\frac{1}{2} \sum_{\alpha} E_{\alpha}^2 \right\} \left(1 + \frac{1}{2} \sum_{j \in S} \sum_{\alpha} \Xi(\alpha, j) (E_{\alpha}^2 - 1) G_j + \right. \\ \left. \frac{1}{2} \sum_{j \in S} \sum_{\alpha \neq j} \Xi(\alpha, j) (E_{\alpha}^2 - 1) E_j^2 G_j + \frac{1}{2} \sum_{j \in S} \Xi(j, j) (E_{\alpha}^4 - \gamma_4) G_j \right) \quad (\text{П2.30})$$

Отсюда получаем оценку для производной функции правдоподобия по коэффициентам G_j с точностью до постоянного для всех производных множителя.

$$\frac{\partial L}{\partial G_j} = \frac{1}{2} \sum_{\alpha \neq j} \frac{\omega(\alpha, \alpha, j)}{\varepsilon_{\alpha} \varepsilon_j} (E_{\alpha}^2 - 1) \quad \text{если } j \notin S$$

(П2.31)

$$\frac{\partial L}{\partial G_j} = \frac{1}{2} E_j^2 \sum_{\alpha \neq j} \frac{\omega(\alpha, \alpha, j)}{\varepsilon_{\alpha} \varepsilon_j} (E_{\alpha}^2 - 1) + \frac{\omega(j, j, j)}{\varepsilon_j^2} (E_j^4 - \gamma_4) \quad \text{если } j \in S$$

(П2.32)

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Предложен подход к выбору области, занимаемой макромолекулой в элементарной ячейке кристалла из класса альтернативных областей, основанный на статистическом моделировании, принципе максимального правдоподобия и компьютерной процедуре Монте-Карловского типа, позволяющей оценивать величину правдоподобия. Применимость подхода

продемонстрирована в тестовых расчетах на белках с известной атомной структурой.

2. Предложен метод трансформирования фазовой проблемы рентгеноструктурного анализа в проблему выбора области, занимаемой макромолекулой, из класса альтернативных областей.

3. Предложен подход к *ab-initio* к решению фазовой проблемы при низком разрешении, основанный на фильтрации случайно сгенерированных фазовых наборов при помощи критерия отбора, являющегося аналогом статистического правдоподобия, и усреднении отобранных вариантов. С помощью предложенного подхода для тестируемых структур получены значения фаз, которые можно использовать в качестве первого приближения к решению фазовой проблемы.

4. Получены явные асимптотические формулы для совместного распределения вероятностей набора структурных факторов, пригодные для использования при работе с сильными рефlekсами. Найденные формулы применены для расчета градиента функции правдоподобия и приближения априорного распределения градиентом функции правдоподобия.

В заключение автор выражает глубокую благодарность своему научному руководителю д.ф.-м.н. Лунину В.Ю. за ясную постановку задач, постоянное руководство и внимание.

Автор благодарит к.ф.-м.н. Лунину Н.Л. за помощь при подготовке программ.

Автор выражает особую благодарность Сковороде Т.П. за участие в обсуждениях и постоянные ценные советы.

ЛИТЕРАТУРА

1. Боровков А.А. Математическая статистика. Издательство Института математики. Новосибирск, 1997.
2. ван дер Варден Б.Л.. Математическая статистика. Пер. с нем. Л.Н. Большева. Издательство Иностранной литературы. Москва, 1960.
3. Лунин В.Ю. Использование метода максимального правдоподобия для оценки ошибок при определении фаз в кристаллографии белка. Препринт НЦБИ АН СССР. Пущино, 1982.
4. Ф.Мостеллер, Р.Рурке, Дж.Томас. Вероятность. Пер. с англ. В.В. Фирсова. Москва. Изд. "Мир", 1969.
5. Федорюк М.В. Асимптотика: Интегралы и ряды. М. Наука, 1987.
6. Чиргадзе Ю.Н., Невская Н.А., Фоменкова Н.П., Никонов С.В., Сергеев Ю.В., Бражников Е.В., Гарбер М.Б., Лунин В.Ю., Уржумцев Ф.Г., Верносова Е.А. Пространственная структура гамма-кристаллина IIIb из хрусталика глаза телят при разрешении 2,5 Å. (1986) Доклады АН СССР, т.290, в.2, 492-495.
7. Ширяев А.Н. Вероятность. М. Наука, 1989.
8. AEvarsson, A., Braznihnikov, E., Garber, M., Zhelnotsova, J., Chirgadze, Yu., al-Karadaghi, S., Svensson, L.A. and Liljas, A. Three-Dimensional Structure of the Ribosomal Translocase: Elongation Factor G from *Thermus Thermophilus*. (1994), *Embo Journal*, 13, 3669-3677.
9. Andersson, K.M. Positioning a protein molecule in the asymmetric unit using very low order reflections and packing restraints. (1999). *J. Appl. Cryst.* , **32**, 530-535.
10. Andersson, K.M., Hovmöller, S. Phasing Proteins at Low Resolution (1996). *Acta Cryst. D***52**, 1174-1180.
11. Andersson, K.M., Hovmöller, S. (1997). Seventeenth European Crystallographic Meeting. Lisboa-Portugal, 24/28 August 1997. Programme book of abstracts, 54.
12. Blow, D.M. & Crick, F., H., C. The Treatment of Errors in the Isomorphous Replacement Method (1959). *Acta Cryst.* **12**, 1195-1202.

13. Bricogne G. Maximum Entropy and the Foundations of Direct Methods. (1984). *Acta Cryst.* **A40**, 410-445.
14. Bricogne G. A Bayesian Statistical Theory of the Phase Problem. I. A Multichannel Maximum-Entropy Formalism for Constructing Generalized Joint Probability Distributions of Structure Factors. (1988). *Acta Cryst.* **A44**, 517-545.
15. Bricogne G. Efficient Sampling Methods for Combinations of Signs, Phases, Hyperphases, and Molecular Orientations. (1997). *Methods in Enzymology*, volume 276. *Macromolecular Crystallography*, part A, edited by Charles W. Carter, Jr., 424-448.
16. Bricogne G. & Gilmore C.,J. A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. I. Theory, Algorithms and Strategy. (1990). *Acta Cryst.* **A46**, 284-297.
17. Brünger A.T. Free R value: A Novel Statistical Quantity for Assessing the Quality of Crystal Structures. (1992) *Nature(London)*, v.355, 472-475.
18. Cochran, W. Relations between the phases of the structure factors. (1955). *Acta Cryst.* **8**, 473-478.
19. Derrick, J.P., Wingley, D.B. The Third Igg-Binding Domain from Streptococcal Protein G: An Analysis by X-ray Crystallography of the Structure Alone and in a Complex with Fab (1994). *J.Mol.Biol.*, **243**, 906.
20. De Titta, G., Weeks, C., Thuman, P., Miller, R. & Hauptman, H. Structure solution by minimal function phase refinement and Fourier filtering. I. Theoretical basis. (1994). *Acta Cryst.* **A50**, 203-210.
21. Doublet, S., Xiang, S., Gilmore C.,J. & Bricogne G. Overcoming Non-Isomorphism by Phase Permutation and Likelihood Scoring: Solution of the TrpRS Crystal Structure. (1994). *Acta Cryst.* **A50**, 164-182.
22. Germain, G. & Woolfson, M.M. On the application of phase relationships to complex structures. (1968) *Acta Cryst* **B24**, 91-96.
23. Germain, G., Main, P. & Woolfson, M.M. On the application of phase relationships to complex structure. II. Getting a good start. (1970) *Acta Cryst* **B26**, 274-285.

24. Giacovazzo, C. A General Approach to Phase Relationships: The Method of Representations (1977). *Acta Cryst.* **A33**, 934-944.
25. Giacovazzo, C. The method of representations of structure seminvariants. II. New theoretical and practical aspects (1980). *Acta Cryst.* **A36**, 362-372.
26. Gilmore, C.,J., Bricogne, G. & Bannister, C. A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. II. Application to small molecules. (1990). *Acta Cryst.* **A46**, 297-308.
27. Gilmore C.,J. Maximum Entropy and Bayesian Statistics in Crystallography: a Review of Practical Applications. (1996). *Acta Cryst.* **A52**, 561-589.
28. Gilmore C.,J. & Bricogne G. MICE Computer Program (1997). *Methods in Enzymology*, volume 276. *Macromolecular Crystallography*, part A, edited by Charles W. Carter, Jr. ,65-78.
29. Gilmore C.J., Dong W. & Bricogne G. A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. VI. The use of error-correcting codes as a source of phase permutation and their application to the phase problem in power, electron and macromolecular crystallography. (1999). *Acta Cryst.* **A55**, 70-83.
30. Hauptman, H. On the identity and estimation of those cosine invariants, $\cos(\varphi_l + \varphi_m + \varphi_n + \varphi_p)$, which are probably negative (1974a). *Acta Cryst.* **A30**, 472-476.
31. Hauptman, H. On the theory and estimation of the cosine invariants $\cos(\varphi_l + \varphi_m + \varphi_n + \varphi_p)$ (1974b). *Acta Cryst.* **A30**, 822-829.
32. Hauptman, H. A joint probability distribution of seven structure factors (1975a). *Acta Cryst.* **A31**, 671-679.
33. Hauptman, H. A new method in the probabilistic theory of the structure invariants (1975b). *Acta Cryst.* **A31**, 680-687.
34. Hauptman, H. Probabilistic theory of the structure invariants: extension to the unequal atom case with application to neutron diffraction (1976). *Acta Cryst.* **A32**, 877-882.
35. Hauptman, H. A Sequence of Nested Neighborhoods of the Structure Invariant $\varphi_h + \varphi_k + \varphi_l + \varphi_m$ (1977). *Acta Cryst.* **A33**, 553-555.

36. Hauptman, H. A minimal principle in the phase problem. (1991). Crystallographic Computing 5 from Chemistry to Biology; Proceedings of the International School of Crystallographic Computing, Bischenberg, France (1990); D. Moras, A.D. Podjarny & J.C. Thierry (eds.). IUCr Oxford University Press. 324-332.
37. Hauptman, H. & Karle, J. The probability distribution of the magnitude of a structure factor: II. The non-centrosymmetric crystal. (1953). *Acta Cryst.* **6**, 136.
38. Hauptman, H. & Karle, J. Seminvariants for centrosymmetric space groups with conventional centered cells. (1959). *Acta Cryst.* **12**, 93-97.
39. Harris, G.W. Fast Ab-initio Calculation of Solvent Envelopes for Protein Structures. (1995). *Acta Cryst.*, **D51**, 695-702.
40. Hašek, J. & Schenk, H. On the Comparison of Different Sets of Structure-Factor Phases (1992). *Acta Cryst.* **A48**, 693-695.
41. Hendrickson, W., A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation (1991). *Science* **254**, 51-58.
42. Hienerman, J.J.L. The Use of Structural Information in the Phase Probability of a Triple Product. (1977). *Acta Cryst.* **A33**, 100-106.
43. Jaynes, E.T. Informational Theory and Statistical Mechanics (1957). *Physical Review*, v. 106, num. 4, 620-630.
44. Karle, J. & Karle, I.L. An application of the symbolic addition method to the structure of L-arginine dihydrate. (1964). *Acta Cryst.* **17**, 835-842.
45. Karle, J. & Karle, I.L. Symbolic addition procedure for phase determination for centrosymmetric and noncentrosymmetric crystals. (1966). *Acta Cryst.* **21**, 849-859.
46. Karle, J. & Hauptman, H. A theory of phase determination for the four types of non-centrosymmetric space groups $1P222$, $2P22$, $3P_12$, $3P_22$. (1956). *Acta Cryst.* **9**, 635-651.
47. Karle, J. & Hauptman, H. Seminvariants for non-centrosymmetric space groups with conventional centered cells. (1961). *Acta Cryst.* **14**, 217-223.
48. Klug, A. Joint Probability Distribution of Structure Factors and the Phase Problem. (1958). *Acta Cryst.* **11**, 515-543.

49. Kraut, J. The Shape of the Chymotrypsinogen Molecule from X-Ray Diffraction (1958). *Biochim. Biophys. Acta*, **30**, 265-270.
50. Lunin, V.Y. The likelihood based choice of priors in statistical approaches to the phase problem. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp.451-454. Kluwer Academic Publishers.
51. Lunin, V.Y., Lunina, N.L. The Map Correlation Coefficient for Optimally Superposed Maps. (1996) *Acta Cryst.*, **A52**, 365-368.
52. Lunin, V.Y. & Skovoroda, T.P. R-Free Likelihood Based Estimates of Errors for Phases Calculated from Atomic Models (1995) *Acta Cryst.*, **A51**, 880-887.
53. Lunin, V.Y. & Urzhumtsev, A.G. Improvement of Protein Phases by Coarse Model Modification. (1984) *Acta Cryst.*, **A40**, 269-277.
54. Lunin, V.Y. & Urzhumtsev, A.G. Maximal Likelihood refinement. It works, but why? (Seminar notes). (1999) *CCP4 Newsletter on Protein Crystallography*, **N. 37**, 14-28.
55. Lunin, V.Y., Urzhumtsev, A.G. & Skovoroda, T.P. Direct Low-resolution Phasing from Electron-Density Histograms in Protein Crystallography. (1990). *Acta Cryst.* **A46**, 540-544.
56. Lunin, V.Y. & Woolfson, M.M. Mean Phase Error and the Map Correlation Coefficient. (1993) *Acta Cryst.*, **D49**, 530-533.
57. Main, P. Towards an ab-initio method for the determination of macromolecular crystal structures. (1999). *Collected Abstracts, XVIIIth IUCr Congress & General Assembly, 4-13 August 1999, Glasgow, Scotland*, M12.BB.001, 183.
58. Miller, R. & Weeks, C.M. Shake-and-bake: applications and advances. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp.389-400. Kluwer Academic Publishers.
59. Moras D., Lorber B., Romby P., Ebel J.-P., Giegé R., Lewitt-Bentley A., Roth M. Yeast tRNA^{Asp}-Aspartyl-tRNA Synthetase: The Crystalline Complex. (1983). *J. Biomol. Structure & Dynamics* **1**, 209-223.
60. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method (1997) *Acta Cryst.*, **D53**, 240-255.

61. Pannu & R.J. Read Improved Structure Refinement Through Maximum Likelihood (1996) *Acta Cryst.*, **A52**, 659-668.
62. Perutz, M.E. Isomorphous Replacement and Phase Determination in Non-centrosymmetric Space Groups. (1956) *Acta Cryst.*, **9**, 867-873.
63. Podjarny, A.D., Rees, B., Thierry, J.-C., Cavarely, J., Jesior, J.C., Roth, M., Lewitt-Bentley, A., Kahn, R., Lorber, B., Ebel, J.-P., Giege, R., Moras, D. Yeast tRNA^{Asp}-Aspartyl-tRNA Synthetase Complex: Low Resolution Crystal Structure (1987). *Journal of Biomolecular Structures & Dynamics* **5**, 187-198.
64. Podjarny, A.D., Rees, B. & Urzhumtsev, A.G. Density Modification in X-ray Crystallography. (1996). In *Methods in Molecular Biology*, **56**, 205-226.
65. Podjarny, A.D., Schevitz, R.W. & Sigler, P. Phasing low resolution macromolecular structure factors by matricial direct methods. (1981) *Acta Cryst.*, **A37**, 662-668.
66. Podjarny, A.D. & Urzhumtsev, A.G. Low-resolution Phasing. (1997). *Methods in Enzymology*, volume 276. *Macromolecular Crystallography, part A*, edited by Charles W. Carter, Jr., 641-658.
67. Podjarny, A.D., Urzhumtsev, A.G. & Lunin, V.Y. Model based very low resolution phasing. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp.421-431. Kluwer Academic Publishers.
68. Read, R.J. Improved Fourier coefficients for maps using phases from partial structures with errors (1986) *Acta Cryst.*, **A42**, 140-149.
69. Read, R.J. Structure-factor probabilities for related structures (1990) *Acta Cryst.*, **A46**, 900-912.
70. Read, R.J. Model Phases: Probabilities and Bias (1997). In *Methods in Enzymology*, Academic Press, San Diego., Charles W., Carter, Jr., R.M. Sweet, eds., 277, part B, 110-128.
71. Rossmann, M.G. (1972). *The Molecular Replacement Method.*, Gordon & Breach; New York, London, Paris.
72. Shannon, C., E. & Weaver, W. (1949) *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, USA.

73. Sevcik, J., Dodson, E. & Dodson, G.G. Determination and restrained least-squares refinement of the structures of ribonuclease Sa and its complex with 3'-guanylic acid at 1.8 Å resolution. (1991) *Acta Cryst.*, **B47**, 240-253.
74. Urzhumtsev, A.G., Lunin, V.Y. & Lunina, N.L. Seminvariant density decomposition and connectivity analysis in the low resolution macromolecular phasing. (1998). *ECM-18 Abstracts, XVIIIth European Cryst. Meeting, 16-20 August 1998, Prague, Republic Czech, E5-P7, Bulletin of the Czech and Slovak Crystallographic Association*, **5B**, 482-483.
75. Urzhumtsev, A.G., Skovoroda, T.P. & Lunin, V.Y. A procedure compatible with X-PLOR for the calculation of electron-density maps weighted using an R-free-likelihood-based approach. (1996) *J. Appl. Cryst.*, **29**, 741-744.
76. Urzhumtsev, A.G., Podjarny, A.D. & Navaza, J. On the solution on the molecular replacement problem at very low resolution. (1994). *J. CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **30**, 29-36.
77. Urzhumtsev, A.G., Podjarny, A.D. On the Solution of the Molecular-Replacement Problem at Very Low resolution: Application to Large Complexes. (1995). *Acta Cryst.* **A51**, 888-895.
78. Urzhumtsev, A.G., Vernoslova, E.A. & Podjarny, A.D. Approaches to very low resolution phasing of the ribosome 50S particle from *Thermus thermophilus* by Few-Atoms-Models and Molecular Replacement methods. (1996). *Acta Cryst.* **D52**, 1092-1097.
79. Volkman, N., Hottenträger, S., Hansen, H.A.S., Zaytsev-Bashan, A., Sharon, R., Yonath, A. & Wittmann, H.G. Characterization and Preliminary Crystallographic Studies on Large Ribosomal Subunits from *Thermus thermophilus*. (1990). *J. Mol. Biol.* **216**, 239-241.
80. Weeks, C., De Titta, G., Hauptman, H., Thuman, P. & Miller, R. Structure solution by minimal function phase refinement and Fourier filtering. II. Implementation and applications. (1994). *Acta Cryst.* **A50**, 210-220.
81. White, P. & Woolfson, M.M. The Application of Phase Relationships to Complex Structures. VII. Magic Integers. (1975). *Acta Cryst.*, **A29**, 133-137.

82. Wilson, A.,J.,C. The Probability Distribution of X-ray Intensities. (1949). Acta Cryst., **2**, 318-321.
83. Woolfson, M.M. Structure Determination by the Method of Permutation Syntheses. (1954). Acta Cryst., **7**, 65-67.
84. Woolfson, M.M. Direct Methods - from Birth to Maturity. (1987). Acta Cryst., **A43**, 593-612.
85. Xiang, S., Carter, C.W., Jr., Bricogne, G. & Gilmore, C.J. Entropy Maximization Constrained by Solvent Flatness: a New Method for Macromolecular Phase Extension and Map Improvement (1993). Acta Cryst., **D49**, 193-212.