

АКАДЕМИЯ НАУК СССР

КРИСТАЛЛОГРАФИЯ

(ОТДЕЛЬНЫЙ ОТТИСК)

2

МОСКВА · 1983

УДК 548.737:547.96

ЛУНИН В. Ю., УРЖУМЦЕВ А. Г.

ПОВЫШЕНИЕ РАЗРЕШЕНИЯ КАРТ ЭЛЕКТРОННОЙ ПЛОТНОСТИ БЕЛКОВ ПУТЕМ УТОЧНЕНИЯ МОДЕЛЬНОЙ СТРУКТУРЫ

I. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

Приводится описание численных экспериментов по расширению и уточнению набора фаз, определяемых при рентгеноструктурном исследовании белка. Улучшение набора фаз производится путем построения и уточнения пробной модели структуры.

При расшифровке кристаллической структуры белка довольно распространенной является ситуация, когда метод изоморфного замещения (дополненный иногда данными по аномальному рассеянию) позволяет определить фазы структурных факторов лишь при разрешении порядка 3–3,5 Å, а надежность таких фаз порой недостаточна. Это приводит к тому, что карты электронной плотности, построенные по такому набору фаз, сложны в интерпретации и затрудняют (или даже делают невозможным) построение атомной модели.

В то же время экспериментальный набор модулей структурных факторов, как правило, бывает шире, чем полученный набор фаз, и может содержать, например, данные до разрешения 2,5–2 Å. В связи с этим возникает желание как-то использовать дополнительную информацию о модулях структурных факторов с тем, чтобы улучшить набор фаз — уточнить значение известных фаз и определить фазы, значения которых были неизвестны. Подчеркнем, что речь идет о ситуации, когда мы не в состоянии проинтерпретировать синтез и построить атомную модель структуры, так как при наличии такой модели задача улучшения набора фаз решается в процессе уточнения модели тем или иным способом.

В 1978 г. Агарвал и Айзекс в [1] высказали идею, как, стартуя с набора фаз среднего разрешения (например, 3 Å) и набора модулей структурных факторов более высокого разрешения (например, 2 Å), определить фазы для тех структурных факторов, для которых они не были определены.

Предложенный метод состоял в следующем: а) по функции распределения электронной плотности, рассчитанной из стартового набора фаз, строится некоторая фиктивная модель структуры; эта модель строится автоматически специальной программой, не претендует на стереохимическое соответствие истинной структуре и должна лишь правильно отражать распределение электронной плотности в кристалле; б) параметры атомов этой фиктивной модели уточняются программой кристаллографического уточнения в обратном пространстве без учета стереохимических ограничений; в) рассчитывается новая функция распределения электронной плотности (более высокого разрешения) с использованием экспериментальных значений модулей структурных факторов и рассчитанных по уточненной модели фаз.

В данной работе приводятся результаты двух численных экспериментов по применению этой методики. Первый эксперимент относится к задаче расширения набора фаз, т. е. к определению фаз, ранее отсутствовавших. Второй — к задаче уточнения имеющегося набора фаз.

Расширение набора фаз

Для теста был взят белок актинидин (выбор объекта случаен), кристаллизующийся в гр. $P2_12_12_1$ с параметрами ячейки $a=78,2$ Å, $b=81,8$ Å, $c=33,05$ Å. В независимой части находится одна молекула, содержащая

218 аминокислотных остатков. По координатам атомов, взятым из банка белковых молекул, рассчитаны модули $F(\mathbf{s})$ и фазы $\phi(\mathbf{s})$ структурных факторов до разрешения 2 Å (здесь $\mathbf{s}=(h, k, l)$ — целочисленный вектор обратного пространства, $|\mathbf{s}|=2 \sin \theta/\lambda$). При этом всем атомам приписан изотропный температурный фактор $B_t=20$ Å. Будем далее называть этот набор координат моделью M_0 .

В первом эксперименте набор модулей $F(\mathbf{s})$ по зоне до 2 Å (15009 рефлексов) имитирует набор экспериментально измеренных модулей структурных факторов нативного белка, а набор фаз $\phi(\mathbf{s})$ по зоне до 3 Å (4641 рефлекс) имитирует стартовый набор фаз, полученных по какой-либо методике. Наша задача состояла в том, чтобы, используя только эти данные, определить значения фаз для зоны $2 \text{ Å} \leq d = 1/|\mathbf{s}| = \lambda/2 \sin \theta < 3 \text{ Å}$. Фазы из этой зоны, рассчитанные по модели M_0 , в процессе эксперимента использовать не разрешалось, и они служили лишь для наблюдения за ходом работы.

Первый этап работы заключался в построении синтеза электронной плотности с разрешением 3 Å по стартовому набору фаз и последующем автоматическом построении атомной модели белка (мы будем ее обозначать M_1) по этому синтезу.

При построении модели M_1 считалось, что изображение на синтезе электронной плотности любого атома с температурным фактором B_t можно аппроксимировать гауссовой кривой

$$\rho_{B_t}(\mathbf{r}) = C \left(\frac{4\pi}{B+B_t} \right)^{\frac{3}{2}} \exp \left(-\frac{4\pi^2 R^2(\mathbf{r})}{B+B_t} \right),$$

где $R(\mathbf{r})$ — расстояние в Å от начала координат до точки с координатами $\mathbf{r}=(x, y, z)$, а C и B — некоторые константы, зависящие от разрешения, с которым построен синтез (см. [2]). В нашем случае значения этих констант были определены как $C=12,1$, $B=71,4$ Å. Процесс построения модели шел следующим образом. Сначала на синтезе выбирался максимум плотности, при этом его координаты считались координатами центра некоторого атома. Температурный фактор B_t для этого атома подбирался так, чтобы значение $\rho_{B_t}(0)$ совпадало со значением плотности в этом максимуме. Если выполнение этого условия возможно лишь при $B_t < 0$, то для такого атома принудительно устанавливалось $B_t=0$. (Отметим, что последнее условие приводит к тому, что мощный пик на синтезе, отвечающий группе атомов, представляется в модели M_1 несколькими атомами с близкими координатами центров.) Затем из синтеза вычиталась плотность $\rho_B(r)$ определенного таким образом атома. После этого происходил возврат к началу процесса до тех пор, пока максимальное значение на синтезе не становилось меньше заданного уровня.

В результате работы программы по указанному алгоритму создана модель M_1 из 2168 атомов. Отметим сразу, что, исходя из результатов численных экспериментов, мы считаем полезным в начале работы включить в модель больше атомов, чем их есть на самом деле в структуре. Это связано с двумя обстоятельствами: во-первых, при нашем способе построения модели часть атомов окажется не на своем месте, что будет обнаружено в процессе уточнения, и такие атомы в дальнейшем мы удалим из модели; во-вторых, поскольку мы ограничили величину максимума электронной плотности отдельного атома значением $C(4\pi/B)^{\frac{3}{2}}$, то возможно, что крупные атомы (например, атомы серы) будут формироваться несколькими модельными атомами с близкими координатами центров.

Для контроля по модели M_1 рассчитан набор фаз. Результаты сравнения его с контрольным набором фаз (рассчитанным по модели M_0) приведены ниже. Отметим также, что сравнение фаз показывает рост средней ошибки с уменьшением величины модуля структурного фактора.

Следующим этапом работы явилось уточнение модели M_1 . Для уточнения использовалась программа уточнения в обратном пространстве [3].

Таблица 1

Расширение набора фаз. План уточнения модельной структуры

Уточняемая модель	d_{\min}	Количество рефлексов	R^*	Количество циклов по B	$B_{\text{стрем}}$	Количество циклов по X	Полученная модель	Количество атомов	R^*
M_1	3,0	4641	29,5	2	—	3	M_2	2168	12,9
M_2	3,0	4641	12,9	2	80	3	M_3	2142	9,4
M_3	2,7	6294	15,5	2	70	3	M_4	2043	12,0
M_4	2,5	7874	16,1	2	70	3	M_5	2033	13,2
M_5	2,2	11353	21,0	2	70	3	M_6	2026	16,4
M_6	2,0	15009	21,4	2	70	3	M_7	2022	18,2
M_{10}	3,0	4641	17,1	2	—	3	M_{11}	2933	7,8
M_{11}	2,5	7874	14,7	2	70	3	M_{12}	2872	10,1
M_{12}	2,2	11353	16,1	2	50	3	M_{13}	2560	13,6
M_{13}	2,0	15009	18,0	2	40	3	M_{14}	2295	16,6
M_{14}	2,0	15009	16,6	1	35	2	M_{15}	2080	16,8
M_{17}	2,0	15009	22,6	2	—	3	M_{18}	2205	16,5
M_{18}	2,0	15009	16,5	2	50	3	M_{19}	1782	16,9
M_{19}	2,0	15009	16,9	2	45	3	M_{20}	1681	15,4

$$* R = \frac{\sum |F - F^c|}{\sum F} \cdot 100,$$

Таблица 2

Расширение набора фаз. Зависимость средней фазовой ошибки (в град) для нецентросимметричных рефлексов от $s^2 = (2\sin \theta/\lambda)^2$

Модель	Границы интервалов по s^2									Σ
	0,000— 0,028	0,028— 0,055	0,055— 0,083	0,083— 0,111	0,111— 0,139	0,139— 0,167	0,167— 0,194	0,194— 0,222	0,222— 0,250	
M_1	22	19	22	30	60	66	70	72	76	57
M_2	17	15	18	27	55	64	69	70	74	54
M_3	17	14	17	28	55	62	69	70	73	53
M_4	19	14	18	27	50	60	64	68	71	51
M_5	18	13	17	26	47	56	62	64	68	49
M_6	18	13	17	24	42	51	57	58	63	45
M_7	17	13	17	23	39	48	53	54	58	42
M_{10}	17	15	18	25	42	52	57	59	63	45
M_{11}	15	13	16	24	40	50	56	59	64	44
M_{12}	15	13	16	23	37	47	52	54	60	41
M_{13}	16	13	17	22	33	44	47	49	55	38
M_{14}	17	14	17	22	32	41	43	45	51	36
M_{15}	16	15	18	22	31	39	42	43	50	35
M_{17}	21	17	18	24	34	42	44	46	51	37
M_{18}	18	12	14	19	29	36	38	42	47	33
M_{19}	18	13	15	20	28	35	36	40	46	32
M_{20}	18	13	15	20	27	33	35	39	44	31
Количество рефлексов	388	816	1105	1308	1462	1706	1850	1991	2087	12713

минимизирующую критерий

$$R = \sum_s (F(s) - F^c(s))^2,$$

как функцию параметров $\mathbf{r} = (x, y, z)$ и B_t для каждого атома (здесь F^c — модули рассчитанных по модели структурных факторов, F — экспериментальные значения).

Уточнение состояло из шести серий с постепенным увеличением от серии к серии числа рефлексов, включаемых в уточнение. В каждой серии вначале проводилось несколько циклов уточнения B_t , а затем несколько циклов уточнения координат атомов. По ходу работы периодически удалялись из модели атомы со значениями $B_t > B_{\text{срезки}}$ (такие атомы считались включенными в модель неправильно). Сведения о ходе процесса уточнения приведены в табл. 1, а в табл. 2 — результаты сравнения фаз, рассчитанных по модели на разных этапах уточнения, с контрольными. Отметим, что улучшение фаз при уточнении происходило в основном лишь для тех структурных факторов, модули которых были включены в уточнение.

Дальнейшего улучшения набора фаз удалось достигнуть, применив следующую методику. Вначале мы резко сократили число атомов в модели M_7 , оставив всего 1140 атомов с $B_t < 30$. По полученной модели (модель M_8) рассчитаны фазы ϕ^* для зоны до 2,5 Å. Далее проведено три цикла уточнения координат модели M_8 , и по полученной после такого уточнения модели M_9 по зоне до 2,5 Å рассчитаны модули F^c и ϕ^c . После этого построен разностный синтез разрешения 2,5 Å с коэффициентами $F^d \exp(i\phi^d)$, определенными равенством

$$F^d \exp(i\phi^d) = \begin{cases} F \exp(i\phi) - F^c \exp(i\phi^c), & d \geq 3 \text{ Å} \\ F \exp(i\phi^*) - F^c \exp(i\phi^c), & 3 \text{ Å} > d \geq 2,5 \text{ Å} \end{cases}$$

На разностном синтезе по приведенному выше алгоритму автоматически определено 1793 атома, которые добавлены к модели M_8 (модель M_{10}). Далее проведено уточнение этой модели по плану, приведенному в табл. 1. Результаты сравнения фаз с контрольными, приведенные в табл. 2, показывают эффективность разностных синтезов в исследуемой методике.

Ввиду достигнутого эффекта от применения разностного синтеза этот прием повторен на разрешении 2 Å. Из модели M_{15} удалены атомы с $B_t > 20 \text{ Å}^2$ так, что осталось 1013 атомов — модель M_{16} . После этого построен разностный синтез разрешением 2 Å, по которому к модели M_{16} добавлено 1192 атома — модель M_{17} . Уточнение модели M_{17} привело к дальнейшему уменьшению средней фазовой ошибки, что отражено в табл. 1, 2.

Уточнение набора фаз

Во втором численном эксперименте имитировалась ситуация, когда стартовый набор фаз (при разрешении 3 Å) имеет ошибки, что приводит к искажениям синтеза, с которого начинается работа. Задача состояла в том, чтобы, оставаясь в рамках того же разрешения (не привлекая для работы ни модули, ни фазы структурных факторов с $d < 3 \text{ Å}$), уменьшить ошибку в фазах стартового набора.

Стартовый набор фаз создан внесением случайных ошибок в точные значения контрольных фаз, рассчитанных по модели M_0 . Ошибки вносились таким образом, чтобы средняя ошибка в фазах росла с ростом $s = -2 \sin \theta / \lambda$ и убывала с ростом силы рефлекса. Более точно для данного структурного фактора ошибка фазы δ генерировалась как значение случайной величины с плотностью распределения вероятностей

$$P(\delta) \sim \exp(t(|s|)F^2 \cos(\delta)),$$

где $t(|s|)$ — параметр, регулирующий средний уровень ошибки, $t(|s|)$ выбрано так, что среднее значение ошибки по всем нецентросимметричным рефлексам зоны $d \geq 3 \text{ Å}$ составило 37°. Данные о распределении средней ошибки по зонам по s^2 приведены в табл. 3.

Дальнейшая работа по улучшению набора фаз велась в основном так же, как в предыдущем эксперименте, за исключением того, что все уточнение велось по зоне $d \geq 3 \text{ Å}$. После трех серий уточнения температурных факторов и координат следовали разностные синтезы. План работы приведен в табл. 4.

Результаты, приведенные в табл. 3, показывают, что, несмотря на наличие ошибок в стартовом наборе фаз, улучшение фаз по данной методике

Таблица 3

Уточнение набора фаз. Зависимость средней фазовой ошибки (в град) для нецентросимметричных рефлексов от $s^2 = (2\sin \theta/\lambda)^2$

Модель	Границы интервалов по s^2				Σ
	0,000—0,028	0,028— —0,055	0,055—0,083	0,083—0,111	
Исходное возмущение фаз	24	28	36	47	37
M_1	27	26	31	43	34
M_2	24	24	28	39	31
M_3	25	24	27	39	31
M_5	24	23	28	41	31
M_6	19	16	20	29	22
M_7	16	14	19	29	21
M_9	21	21	26	37	28
M_{10}	19	16	19	31	23
M_{11}	17	15	20	31	22
Количество рефлексов	388	816	1105	1308	3617

Таблица 4

Уточнение набора фаз. План уточнения модельной структуры

Уточняе- мая модель	d_{min}	Количес- тво реф- лексов	$R *$	Количе- ство цик- лов по B	$B_{спреки}$	Количес- ство цик- лов по X	Получен- ная модель	Количес- ство ато- мов	$R *$
M_1	3,0	4641	32,4	2	—	3	M_2	2025	14,9
M_2	3,0	4641	14,9	2	50	3	M_3	1859	11,4
M_5	3,0	4641	31,2	2	—	3	M_6	2469	12,6
M_6	3,0	4641	12,6	2	50	3	M_7	2469	8,2
M_9	3,0	4641	27,3	2	—	3	M_{10}	2040	11,4
M_{10}	3,0	4641	11,4	2	50	3	M_{11}	2040	9,0

$$* R = \frac{\sum |F - F^C|}{\sum F} \cdot 100.$$

сводит ошибки до такого же уровня, какой был достигнут в процессе уточнения модели, построенной по идеальному синтезу (сравни данные по модели M_{11} в табл. 3 с данными по модели M_3 в табл. 2). Дальнейшее снижение ошибки в фазах может быть достигнуто (как это и произошло в первом эксперименте) за счет подключения к уточнению модулей структурных факторов более высокого разрешения.

Выводы

Полученные в тестах результаты — средняя ошибка предсказанных в первом эксперименте фаз 31° и снижение средней ошибки стартового набора с 37 до 22° — показывают высокую надежность фаз, рассчитанных даже по очень грубой модели структуры. Использование таких фаз в дополнение к стартовой фазовой информации, полученной иными методами, позволяет получить карты электронной плотности, содержащие большее число деталей, нежели стартовые. Это может заметно облегчить интерпретацию карт и построение модели.

В работе по уточнению модели использована программа [3], любезно предоставленная ее авторами Ю. Н. Чиргадзе, которая затем была исправ-

лена и адаптирована к ЭВМ серии ЕС в НИВЦ АН СССР А. Г. Уржумцевым. Все расчеты велись на ЭВМ ЕС-1040.

Авторы благодарны Г. Н. Борисюк за консультации по вопросам теории вероятностей и статистики, а также сотрудникам Института белка АН СССР Ю. Н. Чиргадзе и Ю. В. Сергееву за ценные обсуждения практических вопросов методики. Особо авторы хотят выразить признательность операторам и программистам группы сопровождения задач НИВЦ АН СССР и ее руководителю В. А. Королеву за помощь в организации и проведении трудоемких расчетов.

Литература

1. Agarwal R. C., Isaacs N. W. Proc. Nat. Acad. Sci. U.S.A., 1977, **74**, 2835.
2. Лунин В. Ю., Уржумцев А. Г. Повышение разрешения карт электронной плотности белков путем уточнения модельной структуры. I. Описание метода. Препринт. ОНТИ НЦБИ АН СССР, Пущино, 1981.
3. Agarwal R. C. Acta crystallogr., A, 1978, **34**, 791.

Научно-исследовательский
вычислительный центр АН СССР

Поступила в редакцию
15.X.1981