

АКАДЕМИЯ НАУК СССР  
НАУЧНЫЙ ЦЕНТР БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ  
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР

ПРЕПРИНТ

А.Г.УРЖУМЦЕВ

**ИСПОЛЬЗОВАНИЕ ЛОКАЛЬНОГО  
УСРЕДНЕНИЯ ПРИ АНАЛИЗЕ  
ИЗОБРАЖЕНИЙ МАКРОМОЛЕКУЛ  
НА КАРТАХ РАСПРЕДЕЛЕНИЯ  
ЭЛЕКТРОННОЙ ПЛОТНОСТИ**

ПУЩИНО • 1985

УДК 548.73

Приводится описание способа улучшения изображений макромолекул на картах распределения электронной плотности с помощью процедуры локального усреднения и примеры применения такого подхода к данным по ряду белков.

Работа будет направлена в журнал "Acta Crystallographica".

### 1. Задачи анализа изображения макромолекул при рентгеноструктурном исследовании

Рентгеноструктурный анализ является важнейшим методом исследования пространственной структуры молекул. Конечным результатом здесь является пространственная структура, представленная в виде набора координат центров всех (для макромолекул – неводородных) атомов молекулы и параметров рассеяния этими атомами рентгеновских лучей.

Экспериментальные данные метода не позволяют определять параметры атомов для макромолекул, и промежуточным результатом при рентгеноструктурных исследованиях является определение функции  $\rho(\vec{r})$  распределения электронной плотности в кристаллах исследуемого вещества. Функция  $\rho(\vec{r})$  в силу кристалличности объекта является периодической по каждой из координатных осей и может быть представлена в виде ряда Фурье

$$\rho(\vec{r}) = V^{-1} \sum_{\vec{s}} F(\vec{s}) \exp\{i\varphi(\vec{s})\} \exp\{-2\pi i(\vec{s}, \vec{r})\},$$

где  $V$  – объем элементарной ячейки,  $\vec{s}$  – целочисленные точки обратного пространства,  $F(\vec{s}) \exp\{i\varphi(\vec{s})\}$  – комплексные коэффициенты ряда Фурье.

Модули  $F(\vec{s})$  коэффициентов Фурье с той или иной степенью точности могут быть определены из рентгеновского эксперимента.

При этом набор данных  $\{F(\vec{s}), \vec{s} \in S\}$ , естественно, получается конечным, и величина предельного разрешения  $d_{MIN} = \min_{\vec{s} \in S} \{1/|\vec{s}|\}$

является важной характеристикой полученного набора. Значения фаз  $\varphi(\vec{s})$  из эксперимента напрямую получить не удастся, и для определения (приближенных) значений фаз применяются методы изоморфного замещения, аномального рассеяния, молекулярного замещения /1/, использующие различную дополнительную информацию. Полученные таким образом значения фаз часто содержат существенные ошибки.

В результате обрыва ряда Фурье и наличия фазовых ошибок ис-

следовател ь вместо настоящей функции  $\rho(\vec{r})$  имеет дело с некоторым ее приближением  $\rho_c(\vec{r})$ , подчас воспроизводящим  $\rho(\vec{r})$  довольно грубо.

Функция  $\rho_c(\vec{r})$  обычно рассчитывается на трехмерной сетке с соответствующим данному разрешению  $d_{MIN}$  шагом. Для визуального анализа  $\rho_c(\vec{r})$  представляется, как правило, набором карт "двумерных" распределений электронной плотности в сечениях  $\vec{r} = c$  - наборами линий одинакового уровня для функций  $\rho_c(x, y, c)$ .

Анализ функции  $\rho_c(\vec{r})$  на различных этапах исследования может заключаться в решении следующих задач:

- 1) определение формы, границ, "центра" молекулы;
- 2) определение основных структурных элементов молекулы, определение хода полипептидной цепи;
- 3) построение и анализ атомной модели.

## 2. Использование усреднения для анализа границ молекулы

Анализ границ молекулы проводится обычно при невысоком разрешении -  $d_{MIN} \sim 3-10 \text{ \AA}$ . При этом определение границ молекулы интересно как само по себе, так и важно при улучшении качества карт для построения атомных моделей /2/. К сожалению, в ряде исследований по имеющимся картам  $\rho_c(\vec{r})$  эти границы достоверно провести не удастся из-за низкого качества карт, обусловленного, например, трудностями при получении изоморфных производных. Подчеркнем, что в тех ситуациях, когда расположение молекулы все-таки удастся определить, границы молекулы проводятся исследователем вручную, что является трудоемким и вносит дополнительные ошибки.

Нами был предложен следующий способ автоматического определения границ молекулы и ее "центра".

Основная идея состоит в следующем.

Вспомним, что анализ положения молекулы заключается в проведении линий заданного уровня  $\rho_j$ ,  $j=1, \dots, J$ , то есть в выборе точек  $\vec{r}_k$  со значениями функции  $\rho_c(\vec{r}_k) \geq \rho_1$  и определении закономерностей их расположения. В идеальной ситуации при правильном выборе  $\rho_1$  такие точки  $\vec{r}_k$  будут сгруппированы "внутри" изображения молекулы. Таким образом, если взять шар радиуса  $R$ , где  $R$  - характерный размер молекулы (например, радиус инерции), то количество точек  $\vec{r}_k$  с  $\rho(\vec{r}_k) \geq \rho_1$ , лежащих внутри шара, будет постепенно снижаться при продвижении центра шара от "центра" молекулы к ее границам и будет минимальным, когда центр шара окажется в середине промежутка между изображениями соседних молекул.

Предлагаемый нами способ выделения границ молекулы состоит в следующем:

1) "удаляются" (полагаются равными  $g_0 \leq g_1$ ) значения  $g_c(\vec{z})$  в точках  $\vec{z}_k$  таких, что  $g_c(\vec{z}_k) < g_1$ :

$$g_{NUL}(\vec{z}) = \begin{cases} g_c(\vec{z}), & g_c(\vec{z}) \geq g_1 \\ g_0, & g_c(\vec{z}) < g_1 \end{cases};$$

2) для каждой точки  $\vec{z}_j$  определяется сумма значений "за-нуленной" функции  $g_{NUL}(\vec{z}_j)$ :

$$g_{SUM}(\vec{z}_j) = \sum_{e: |\vec{z}_j - \vec{z}_e| \leq R} g_{NUL}(\vec{z}_e);$$

3) проводится анализ положения молекулы для функции  $g_{SUM}(\vec{z})$ ; при достаточном качестве исходной функции  $g_c(\vec{z})$  область молекулы может быть определена как набор точек  $\vec{z}_k$  таких, что  $g_{SUM}(\vec{z}_k) \geq g_1^{SUM}$ , где  $g_1^{SUM}$  - некоторый критический уровень.

При больших значениях радиуса  $R$  этап суммирования представляется достаточно трудоемким. Однако для быстроты вычислений не требуется при переходе от точки к точке заново вычислять всю сумму значений плотности внутри шара, а нужно только пересчитывать изменения суммы за счет передвижения соответствующего "мениска" (рис. 1).

Важным в вычислительном плане является следующий факт, на который было указано В.Ю.Луниным (НИВЦ АН СССР). Поскольку вместо  $g_{SUM}(\vec{z})$  мы можем анализировать

$$g_{AV}(\vec{z}) = V_R^{-1} g_{SUM}(\vec{z}),$$

где  $V_R = 4\pi R^3/3$  - объем шара радиуса  $R$ , то вместо суммирования по узлам сетки внутри шара  $S_R$  мы можем проводить усреднение по этому шару. Записывая

$$g_{AV}(\vec{z}) = g_{NUL}(\vec{z}) * a_R(\vec{z}), \quad \text{где}$$

$$a_R(\vec{z}) = \left(\frac{4}{3}\pi R^3\right)^{-1} \cdot \begin{cases} 1, & |\vec{z}| \leq R \\ 0, & |\vec{z}| > R \end{cases}$$

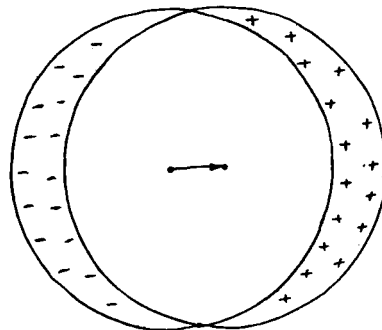


Рис. 1. Схема вычисления функции  $g_{SUM}(\vec{z})$  при последовательном обходе узлов сетки

и используя свойство преобразования Фурье переводить свертку функций в произведение их коэффициентов Фурье

$$\vec{F}_{AV}(\vec{s}) = \vec{F}_{NUL}(\vec{s}) \cdot A(|\vec{s}|), \text{ где}$$

$A(s) = 3(\sin 2\pi sR - 2\pi sR \cdot \cos 2\pi sR) / (2\pi sR)^3$  - интерференционная функция, получаем алгоритм выделения молекулы в следующем виде:

1) рассчитывается

$$\varrho_{NUL}(\vec{z}) = \begin{cases} \varrho_c(\vec{z}), & \varrho_c(\vec{z}) \geq \varrho_1 \\ \varrho_0, & \varrho_c(\vec{z}) < \varrho_1 \end{cases};$$

2) вычисляются  $\vec{F}_{NUL}(\vec{s})$  - коэффициенты Фурье функции  $\varrho_{NUL}(\vec{z})$ ;

3) вычисляются  $\vec{F}_{AV}(\vec{s}) = \vec{F}_{NUL}(\vec{s}) \cdot A(|\vec{s}|)$ ;

4) по  $\vec{F}_{AV}(\vec{s})$  рассчитывается искомая функция  $\varrho_{AV}(\vec{z})$ ;

5) область молекулы определяется как набор точек  $\vec{z}_k$  таких, что  $\varrho_{AV}(\vec{z}_k) \geq \varrho_1^{AV}$ , где  $\varrho_1^{AV}$  - некоторый заданный уровень.

Использование алгоритма быстрого преобразования Фурье /3-5/ позволяет определять  $\varrho_{AV}(\vec{z})$  по  $\varrho_c(\vec{z})$  достаточно быстро и, кроме того, независимо от шага сетки, на которой рассчитана функция  $\varrho_c(\vec{z})$ , то есть без внесения дополнительной погрешности.

### 3. Результаты применения усреднения

Проверка предложенного способа выделения области молекулы была проведена на данных по белкам термитазе, лектину и кристаллину ШБ, исследуемых соответственно в ИК АН СССР, ИМГ АН СССР и ИБ АН СССР. Результаты применения методики к данным по термитазе с любезного разрешения Э.Г. Арутюняна и А.В. Теплякова приведены ниже.

Исходная функция  $\varrho_c(\vec{z})$  (разрешением до 4 Å) с фазами, определенными по методу изоморфного замещения, была рассчитана на сетке с шагом  $\sim 2$  Å по каждой из координатных осей. Карты распределения электронной плотности в термитазе представлены на рис. 2. Определение расположения молекулы по таким картам является затруднительным.

Непосредственное усреднение  $\varrho_c(\vec{z})$  по шару радиусом 20 Å без удаления точек с  $\varrho < \varrho_1$  картины существенно не изменило (рис. 3). Однако в ситуации, когда все значения  $\varrho_c(\vec{z})$  меньшие 0.2 е/Å<sup>3</sup> были заменены на 0.2 е/Å<sup>3</sup> при последующем усреднении с тем же радиусом 20 Å, граница и центр молекулы стали видны даже неспециалисту (рис. 4).

При выборе линий уровня на всех картах использовался прием, предложенный нам А.А. Вагиным (ИК АН СССР). Прием заклю-

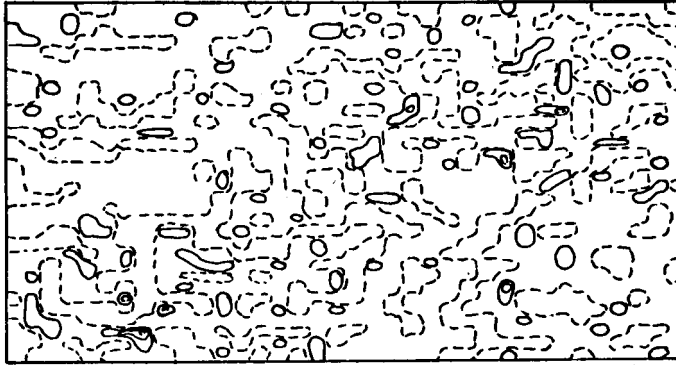


Рис. 2. Изображение функции  $\rho_{4.0}(\bar{z})$  распределения электронной плотности в термитазе,  $\bar{z} = 10/48$ . Данные метода изоморфного замещения, разрешение  $4\text{Å}$

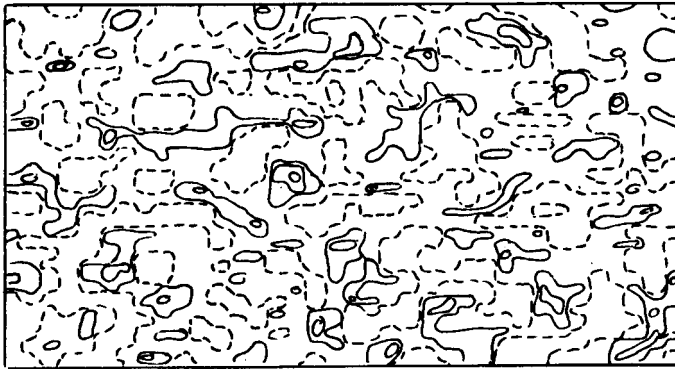


Рис. 3. Изображение функции  $\rho_{sum}(\bar{z})$  для термитазы,  $\bar{z} = 10/48$ . Усреднение  $\rho_{4.0}(\bar{z})$  без предварительного "удаления" точек с  $\rho_{4.0}(\bar{z}) < \rho_1$ , радиус  $R = 20\text{Å}$

чается в следующем: внутри внешней изолинии должен находиться заданный процент общего числа узлов сетки (на рис. 2-4 оставались  $2/3$  общего числа точек). Соответствующее значение  $\rho_1^{AV}$  может быть определено из гистограммы распределения значений  $\rho_{AV}(\bar{z})$ .

Для определения при разрешении  $3.0\text{Å}$  границ молекул лектина, исследуемого в ИМГ АН СССР /6/ и  $\gamma$ -кристаллина ШБ, изучаемого в ИБ АН СССР /7/, использовалось усреднение по шару ра-

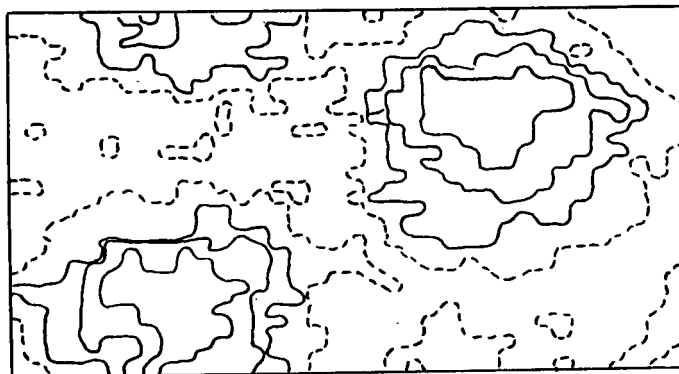


Рис. 4. Изображение функции  $\rho_{sum}(\vec{z})$  для термитазы,  $Z = 10/48$ . Усреднение  $\rho_{4.0}(\vec{z})$  с предварительным "удалением" точек с  $\rho_{4.0}(\vec{z}) < 0,2 \text{ e}/\text{Å}^3$ , радиус  $R = 20 \text{ Å}$

диусом  $16 \text{ Å}$ . Величина  $\rho_1$ , по-прежнему, оставалась равной  $0,2 \text{ e}/\text{Å}^3$ . В этих примерах границы, форма и "центр" молекул вырисовывались так же четко, как и на примере с термитазой.

#### 4. Использование усреднения для анализа более тонких деталей пространственной структуры молекулы

Предложенный выше прием оказывается эффективным и при анализе карт для выявления более тонких деталей пространственной организации молекулы. При этом, естественно, радиус  $R$  шара должен быть существенно уменьшен.

При анализе хода полипептидной цепи, по-видимому, эффективным является усреднение  $\rho_c(\vec{z})$  по шару с радиусом  $3-5 \text{ Å}$ . Нами был проведен следующий эксперимент.

С помощью метода грубой атомной модели /8/ набор фаз структурных факторов для термитазы был расширен до  $3,5 \text{ Å}$ . Однако проведение хода цепи было затруднительным (рис. 5). Усреднение  $\rho_c(\vec{z})$  описанным выше способом с радиусом усреднения  $3 \text{ Å}$  позволило отчетливо увидеть ход цепи значительной части молекулы (рис. 6).

При анализе карт более высокого разрешения при построении атомной модели усреднение с малым радиусом порядка  $1,5-2,5 \text{ Å}$ , по-видимому, может оказаться полезным для удаления "шума" на соответствующих картах.



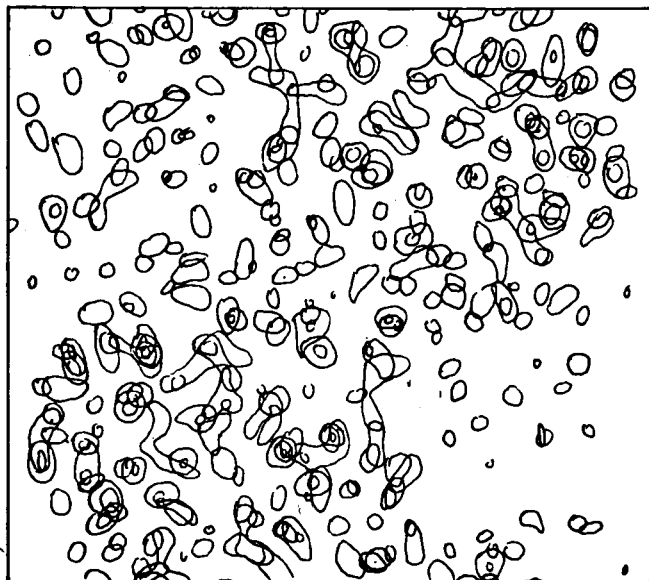


Рис. 5. Изображение функции  $\rho_{3.5}(\vec{r})$  распределения электронной плотности в термитазе,  $\vec{z} = 12\frac{1}{2}16/48$ . Данные метода использования грубой модели, разрешение  $3.5 \text{ \AA}$

## 5. Резюме

1. Предложенный алгоритм выделения молекулы путем усреднения  $\rho_c(\vec{r})$  позволяет при большом радиусе шара усреднения ( $R \sim 15-25 \text{ \AA}$ ) автоматически выделять границы молекулы на имеющихся картах распределения электронной плотности.

2. Использование усреднения с радиусом  $3-5 \text{ \AA}$  позволяет облегчить анализ хода полипептидной цепи.

3. Выполнение усреднения может проводиться "напрямую", путем суммирования  $\rho_{NVL}(\vec{r})$  по узлам сетки; однако более точным, особенно при малом радиусе  $R$ , является расчет  $\rho_{AV}(\vec{r})$  путем перехода к коэффициентам Фурье.

4. Удаление точек со значениями функции  $\rho_c(\vec{r})$ , меньшими заданного уровня  $\rho_1$ , является важным для получения требуемого результата.

5. Значение уровня обводки для анализа границ молекулы на картах  $\rho_{AV}(\vec{r})$  может быть выбрано исходя из гистограммы распределения  $\rho_{AV}(\vec{r})$  путем выбрасывания приблизительно трети общего числа узлов сетки. Значение уровня срезки  $\rho_1$  выбирается из анализа карт  $\rho_c(\vec{r})$ , например,  $\rho_1 \approx F_{000}/V$ , либо также из соответствующей гистограммы.

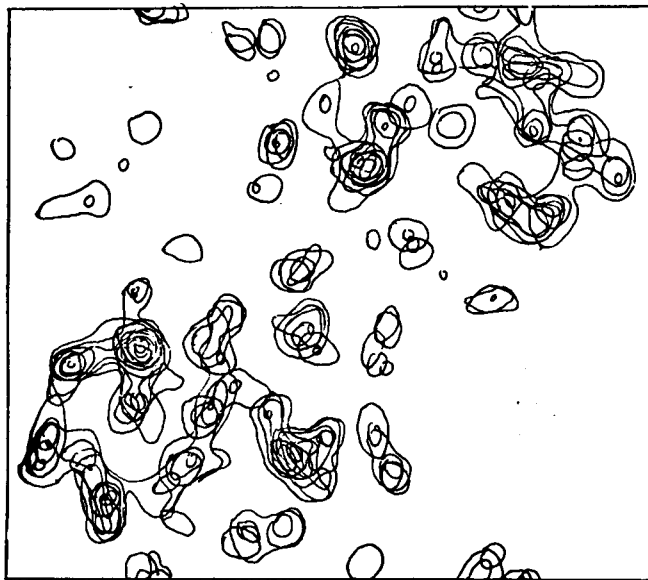


Рис. 6. Изображение функции  $\rho_{\text{sum}}(\vec{z})$  для термитазы,  $\vec{z} = 12\vec{i} + 16\vec{j} / 48$ . Усреднение  $\rho_{3.5}(\vec{z})$ , радиус  $R = 3.0 \text{ \AA}$

На настоящий момент открытыми остаются вопросы об улучшении карт путем усреднения по шару малого диаметра и о критическом качестве исходных карт  $\rho_c(\vec{z})$ , позволяющем решать обсуждаемые выше задачи.

Данная работа была в значительной мере инициирована участием автора в исследовании пространственной структуры термитазы. Автор благодарит сотрудников ИК АН СССР Э.Г.Арутюняна и А.В.Теплякова за плодотворную совместную работу по этому белку и за предоставленные данные по термитазе. Автор благодарен также зав. лаб. ИБ АН СССР Ю.Н.Чиргадзе и сотрудникам ИМГ АН СССР В.З.Плетневу и Ю.Д.Лобсанову за предоставленные данные по  $\gamma$ -кристаллину и лектину. Автор глубоко признателен сотруднику НИВЦ АН СССР В.Ю.Лунину за постоянное внимание, плодотворные обсуждения и ценные советы по данной работе.

#### ЛИТЕРАТУРА

1. Бландел Т., Джонсон Л. Кристаллография белка. М., Мир, 1979.
2. Уржумцев А.Г. Использование смешанной атомной модели для

улучшения наборов фаз структурных факторов при рентгеноструктурном исследовании макромолекул. - Дисс. на соиск. уч. степ. канд. физ.-мат. наук. М., 1985.

3. Cooley I.W., Tukey I.W. - Math. Comput., 1965, v. 19, p. 297-301.

4. Ten Eyck L.F. - Acta Crystallogr., 1973, v. A29, p. 183-191.

5. Лунин В.Ю. Комплекс программ "Быстрое преобразование Фурье". Материалы по мат. обесп. ЭВМ, ФОРТРАН, вып. 3. Пушкино, ОНТИ НЦБИ АН СССР, 1979.

6. Riskulov R.R., Dobrokhotova Z.D., Kuzev S.V., Lobsanov Yu.D., Lubnin M.Yu., Mokulskaya T.D., Myshko G.E., Proskudina L.T., Rogacheva M.M., Saprykina L.F., Khrenov A.A., Mokylyskii M.A. - FEBS Lett., 1984, v. 165, p. 97-101.

7. Chirgadze Yu.N., Sergeev Yu.V., Fomenkova N.P., Oreshin V.D. - FEBS Lett., 1981, v. 131, p. 81-84.

8. Lunin V.Yu., Urzhumtsev A.G. - Acta Crystallogr., 1984, v. A40, p. 269-277.

25.12.85 г. Т21511. Тир. 200. Зак. 6677Р. Уч.-изд.л. 0,5.  
Усл. печ. л. 0,68. Изд. № 396. Бесплатно. ОНТИ НЦБИ АН СССР  
в Пушкине