

# МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ В КРИСТАЛЛОГРАФИЧЕСКОМ МОЛЕКУЛЯРНОМ ЗАМЕЩЕНИИ НА ПРИМЕРЕ РЕШЕНИЯ СТРУКТУРЫ КРИСТАЛЛА РНК-БЕЛКОВОГО КОМПЛЕКСА

Волчков С. А.

(Россия, Пущино)

*На примере решения кристаллической структуры РНК-белкового комплекса показано успешное применение метода максимального правдоподобия в задаче кристаллографического молекулярного замещения, решение которой оказалось невозможным при использовании традиционных методов.*

**Введение.** В основе рентгеновской кристаллографии лежит применение принципов дифракции рентгеновских лучей на структуре кристалла. Для определения пространственной структуры кристалла необходимо найти функцию, описывающую распределение электронной плотности в кристалле исследуемого объекта. Это распределение обладает тройной периодичностью и может быть представлено в виде трехмерного ряда Фурье:

$$\rho(x, y, z) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \mathbf{F}(hkl) \exp(-2\pi i(hx + ky + lz)), \quad (1)$$

где  $\rho$  — электронная плотность,  $V$  — объем элементарной ячейки,  $h, k, l$  — целочисленные индексы, определяющие положение дифракционных максимумов в пространстве (индексы Миллера). Комплексный коэффициент  $\mathbf{F}(hkl)$  называется структурным фактором и может быть представлен в виде  $\mathbf{F}(hkl) = F(hkl) \exp(i\varphi_{hkl})$ , где  $F(hkl)$  — модуль структурного фактора, а  $\varphi_{hkl}$  — фаза.

Регистрируемая дифракционная картина позволяет измерить интенсивности дифракционных максимумов и вычислить значения модулей  $F(hkl)$  структурных факторов. Вся информация о фазах теряется. Определение значений фаз структурных факторов является центральной проблемой кристаллографии.

Один из методов решения фазовой проблемы — молекулярное замещение (МЗ). Основная идея МЗ — использование известной структуры гомологичной и структурно-эквивалентной молекулы, которая должна быть сориентирована в элементарной ячейке кристалла так, чтобы наилучшим образом аппроксимировать положение искомой молекулы. Ориентированная таким образом молекула позволяет рассчитать фазы отражений, которые комбинируют с экспериментально определенными амплитудами и используют для построения карт электронной плотности и построения модели искомой молекулы.

МЗ представляет собой задачу шестимерного поиска, требующего колоссального машинного времени. Один из подходов к решению — разделение ее на две последовательные трехмерные задачи: поиск ортогонального преобразования, обеспечивающего совпадение ориентаций модели и искомой молекулы, и последующий поиск трансляционного вектора, задающего положение ориентированной модели в элементарной ячейке кристалла.

Большинство традиционных методов, применяющихся в задачах МЗ, основываются на расчете функции Паттерсона [1–2]. Эта функция является Фурье-трансформантой величины  $F(hkl)^2$  и определяет систему межатомных векторов, а не индивидуальных положений атомов. Ввиду того что в выражении для функции Паттерсона коэффициентами перед экспонентой являются  $F(hkl)^2$ , а не  $F(hkl)$ , как в (1), карту Паттерсона можно рассчитать, не имея никакой информации о фазах. Совмещение карт Паттерсона модельной и искомой структур в элементарной ячейке искомой структуры позволяет решить задачу МЗ. Однако отсутствие хороших стартовых моделей или наличие нескольких копий макромолекулы в независимой части элементарной ячейки препятст-

вует успешному применению традиционных методов для решения целого ряда структур.

В последнее время во многих областях макромолекулярной кристаллографии стали успешно внедряться методы, основанные на анализе статистического правдоподобия. Применительно к МЗ такой подход был реализован в программе «Phaser» [3-4], которая и использовалась в данной работе.

**Цель работы.** В данной работе требовалось на примере сложной задачи показать успешность применения метода максимального правдоподобия в кристаллографическом МЗ.

**Исследуемый объект.** В работе исследовалась кристаллическая структура комплекса рибосомного белка L1 из бактерии *Thermus thermophilus* (TthL1) с 38-нуклеотидным фрагментом мРНК из *Methanococcus vannielii*. Помимо своей функции в рибосоме, белок L1 является репрессором трансляции своей мРНК. При избытке белка L1 в клетке он связывается со своей собственной мРНК, препятствуя её посадке на рибосому. Получение структуры высокого разрешения белка в комплексе с мРНК является необходимым этапом изучения механизма регуляции.

**Метод максимального правдоподобия в молекулярном замещении.** Принцип метода максимального правдоподобия прост: лучшая модель наилучшим способом соответствует наблюдаемым экспериментальным данным. Соответствие определяется правдоподобием — вероятностью получения имеющихся экспериментальных данных при использовании тестируемой модели.

Для генерации целевой функции правдоподобия предполагается, что отражения независимы [3–4]. Эта аппроксимация позволяет находить полное правдоподобие произведением правдоподобий, рассчитанных для отдельных отражений. Правдоподобие функции вращения для нецентросимметричного отражения  $r$  определяется выражением [3]:

$$P - RF_r = \frac{2F_o}{\sigma_\Delta^2} \exp\left(-\frac{F_o^2 + D^2 F_{\text{big}}^2}{\sigma_\Delta^2}\right) I_0\left(\frac{2F_o D F_{\text{big}}}{\sigma_\Delta^2}\right). \quad (2)$$

Правдоподобие функции трансляции [4]:

$$P - TF_r = \frac{2F_0}{\sigma_{\Delta}^2} \exp\left(-\frac{F_0^2 + D^2 F_c}{\sigma_{\Delta}^2}\right) I_0\left(\frac{2F_0 D F_c}{\sigma_{\Delta}^2}\right). \quad (3)$$

Здесь  $I_0$  — модифицированная функция Бесселя нулевого порядка,  $\sigma$  и  $D$  в совокупности учитывают полноту и точность модели,  $F_0$  — экспериментальное значение амплитуды для отражения  $r$ ,  $F_c$  и  $F_{\text{big}}$  — рассчитанные по модели значения амплитуды. Расчет амплитуд по модели, в случае вращения, выполняется отдельно для каждой копии модели, составляющей разные асимметричные части ячейки. Из этого набора амплитуд для каждого отражения выбирается максимальное рассчитанное значение амплитуды  $F_{\text{big}}$ .

Правдоподобие для модели по формулам (2–3) рассчитывается на сетке углов, а потом на сетке позиций. Углы ориентации и вектор трансляции, обеспечивающие такое расположение модели в ячейке, при котором правдоподобие имеет максимальное значение, выбираются решением задачи МЗ.

Метод максимального правдоподобия на основе выражений (2–3) был реализован в программе «Phaser» [3–4]. Авторы программы, за счет ряда аппроксимаций, добились существенного уменьшения вычислительных затрат. Тестовые расчеты показали, что и полная функция правдоподобия, и модифицированные функции быстрого поиска обладают более высокой чувствительностью к верным решениям, чем ранее использовавшиеся в МЗ функции.

**Построение модели.** Для построения поисковой модели были использованы ранее определенные структуры изолированного белка TthL1 и комплекса архейного белка MjaL1 с фрагментом мРНК. Белок L1 является двухдоменным. На основе структурных и биохимических данных предполагалось, что в комплексе TthL1-мРНК ориентация первого домена белка относительно мРНК должна быть такой же, как и в комплексе MjaL1-мРНК. Фрагменты мРНК в комплексах отличаются только на один нуклеотид. Поэтому в качестве стартовой поисковой модели исполь-

зовались домен I структуры изолированного белка TthL1 и мРНК из структуры комплекса MjaL1-мРНК. Их относительная ориентация была выполнена путем наложения первого домена TthL1 на первый домен MjaL1 в структуре комплекса MjaL1-мРНК.

**Позиционирование модели.** Первоначально задача решалась при помощи широко известных программ «AmoRe», «Molrep» и «CNS», алгоритм поиска которых основан на расчете функции Паттерсона, но решение не было найдено. Сложность задачи состояла в том, что независимая часть элементарной ячейки содержит четыре копии комплекса, при этом в стартовой поисковой модели отсутствует второй домен белка.

Вероятностный же подход, реализованный в программе «Phaser»[3-4], позволил однозначно выделить правильный набор решений. В независимой части элементарной ячейки были последовательно расположены четыре копии стартовой модели. При поиске рассчитывалось не само правдоподобие, а логарифм правдоподобия. Также результаты анализировались при помощи  $Z$ -score — числа стандартных отклонений логарифма правдоподобия от среднего значения всех рассчитанных логарифмов правдоподобия:  $Z - score_i = \frac{\log P_i - \langle \log P_n \rangle}{\sigma}$ .

$$\text{бля: } Z - score_i = \frac{\log P_i - \langle \log P_n \rangle}{\sigma} .$$

Таблица 1. Значение числа $Z$ -score на различных этапах поиска								
Этап поиска	RF1	TF1	RF2	TF2	RF3	TF3	RF4	TF4
$Z$ -score для верного решения	5.6	6.3	5.0	14.1	6.0	13.8	4.3	16.9
$Z$ -score для лучшего неверного решения	5.4	4.2	4.8	12.9	4.7	9.32	3.9	6.1

В табл. 1. приведены значения этого числа для верного решения и лучшего неверного решения на различных этапах поиска.

ка. В ней RF1 и TF1 обозначают соответственно этапы поиска ориентации и последующей трансляции для первого мономера. Расчет правдоподобия для каждого следующего мономера велся с учетом уже расположенных в ячейке копий модели. Из таблицы 1. видно, что на каждом этапе поиска верное решение возглавляет список потенциальных ориентаций или трансляций. На конечном этапе были найдены углы вращения и трансляционные векторы для четырех копий второго домена белка (результаты не представлены).

**Заключение.** Методом «Молекулярного замещения» определена структура кристалла РНК-белкового комплекса. Данную задачу не удавалось ранее решить с использованием традиционных подходов из-за наличия в независимой части ячейки четырех копий комплекса и неудовлетворительного качества стартовой модели. В работе решение было найдено с помощью программы «Phaser», эффективно реализующей принцип максимального правдоподобия. На каждом этапе поиска верное решение возглавляло список потенциальных решений.

#### СПИСОК ЛИТЕРАТУРЫ

6. Бландел Т., Джонсон Т. Кристаллография белка. — Москва: Мир, 1979.
7. Brunger A.T. Patterson correlation searches and refinement. // *Methods in enzymology*. Acad. Press, 1997. — Vol. 276. — P. 558–580.
8. Storoni, L.C., McCoy, A.J., and Read, R.J. Likelihood-enhanced fast rotation functions // *Acta Crystallographica*. — 2004. — Vol. D60. — P. 432–438.
9. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. Likelihood-enhanced fast translation functions // *Acta Crystallographica*. — 2005. — Vol. D61. — P. 458–464.

**APPLICATION OF MAXIMUM LIKELIHOOD APPROACH IN  
X-RAY MACROMOLECULAR REPLACEMENT TO  
DETERMINE RNA-PROTEIN STRUCTURE**

**Volchkov S. A.**

(Russia, Pushchino)

*We successfully applied maximum likelihood approach of X-ray molecular replacement method to determine the crystal structure of the RNA-protein complex. The traditional approaches were failed.*