

蛋白质晶体结构测定通用程序之二

——概率法计算相角

宋时英 李军

(中国科学院生物物理研究所)

晶体结构的测定方法实质上就是测定相角的方法。目前,测定生物大分子晶体结构的主要方法是多对同晶置换法。早在五十年代Harker^[1]就提出利用几何作图法来测定相角,但是由于实验误差(例如:同晶度不好,重原子参数测定不够准确等)所引起的相角多解现象,给作图法解相角带来了困难。为了确定相角的唯一解,Blow和Crick^[2]提出了同晶置换概率法。后来,Narath^[3]和Matthews^[6]发展了这一方法,引进了反常散射信息,综合成为同晶置换与反常散射联合概率法。现在这种方法在蛋白质晶体结构测定中已得到广泛应用。本文将介绍用013-BCY语言编制的概率法计算蛋白质母体相角的通用程序。

一、方法概要

1. 计算同晶概率密度分布

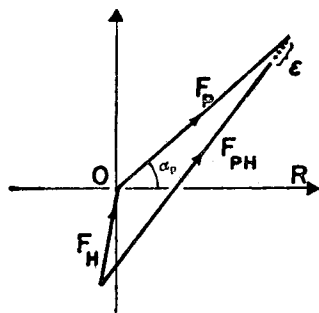


图 1

如图 1 所示,在理想的情况下,向量三角形是闭合的,即: $F_{PH} = F_P + F_H$ 。但是由于存在实验误差,使得三角形往往不能完全闭合。为了处理方便,假定全部误差来自观察值 (F_{PH}) 的不准确。设闭合误差为 ϵ , 则:

$$\epsilon(\phi) = (F_{PH})_o - (F_{PH})_c$$

式中:

$$(F_{PH})_o = \left[\frac{1}{2} (F_{PH}^2(+)) + F_{PH}^2(-) - (F_H'')^2 \right]^{\frac{1}{2}}$$

$$(F_{PH})_c = [F_P^2 + F_H^2 + 2F_P F_H \cos(\phi - \alpha_H)]^{\frac{1}{2}}$$

F_P 为蛋白质母体结构因子的模

F_{PH} 为衍生物结构因子的模。 $(F_{PH})_o$ 为其计算值, $F_{PH}^{(+)}$ 与 $F_{PH}^{(-)}$ 分别为倒易空间正、负半球结构因子的模。

F_H 为重原子的结构因子的模(不包括 F_H'')。

F_H'' 为重原子反常散射虚部分量的向量和的模。

α_H 为重原子的相角。

因为 $\epsilon(\phi)$ 为一服从高斯分布的连续随机变量,故其概率密度分布可以写成:

$$P(\phi) = N_1 \exp(-\epsilon^2(\phi)/2E^2)$$

根据概率可乘性定理,对于 I 个相互独立的重原子衍生物,其同晶联合概率密度分布:

$$P_{iso}(\phi) = N_1 \exp \left[- \sum_{i=1}^I (\epsilon_i^2(\phi)/2E_i^2) \right]$$

式中: 归一化因子

$$N_1 = 1 / \int_0^{2\pi} \exp \left[- \sum_{i=1}^I (\epsilon_i^2(\phi)/2E_i^2) \right] d\phi$$

标准偏差

$$E_i^2 = \sum_{j=1}^n [(F_{PH})_o - (F_{PH})_c]^2 / n$$

n 为衍射点数。

在中心对称情况下,蛋白质母体相角 α_p 只可能是 0° 或 180° , 所以 $\epsilon(\phi)$ 是一个离散型的

随机变量,这时计算就可简化:

α_p 为 0° 的概率分布为:

$$P_+ = N_3 [\exp(-\varepsilon_1^2/2E^2) + \exp(-\varepsilon_2^2/2E^2)]$$

α_p 为 180° 的概率分布为:

$$P_- = N_3 [\exp(-\varepsilon_3^2/2E^2) + \exp(-\varepsilon_4^2/2E^2)]$$

这里, $P_+ + P_- = 1$

$$N_3 = 1/\exp[-(\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \varepsilon_4^2)/2E^2]$$

$$\varepsilon_1 = (\mathbf{F}_H)_c - (F_{pH} - F_p)$$

$$\varepsilon_2 = (\mathbf{F}_H)_c + (F_{pH} + F_p)$$

$$\varepsilon_3 = (\mathbf{F}_H)_c - (F_{pH} + F_p)$$

$$\varepsilon_4 = (\mathbf{F}_H)_c + (F_{pH} - F_p)$$

若为 I 个衍生物,则:

$$P_+ = N_3 \left\{ \exp \left[- \sum_{i=1}^I (\varepsilon_{1i}^2/2E_i^2) \right] + \exp \left[- \sum_{i=1}^I (\varepsilon_{2i}^2/2E_i^2) \right] \right\}$$

$$P_- = N_3 \left\{ \exp \left[- \sum_{i=1}^I (\varepsilon_{3i}^2/2E_i^2) \right] + \exp \left[- \sum_{i=1}^I (\varepsilon_{4i}^2/2E_i^2) \right] \right\}$$

这里, N_3 为联合概率的归一化因子。当 $(P_+ - P_-) > 0$, α_p 取零度; 当 $(P_+ - P_-) < 0$, 则 α_p 取 180° 。 m 因子为 $|P_+ - P_-|$, 当 $P_+ = P_-$ 时 $m = 0$, 此点不参加富氏加和。

2. 计算反常散射概率密度分布

因为有反常散射效应, 致使同一种重原子衍生物的 $F_{pH}^{(+)}$ 与 $F_{pH}^{(-)}$ 不相等, 如图 2 所示。

假设 $\varepsilon_+ = (F_{pH}^{(+)} - F_{pH}^{(-)})_c$ 。

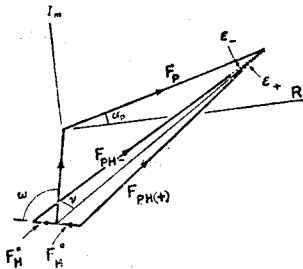


图 2

$$\varepsilon_- = (F_{pH}^{(-)})_c - (F_{pH}^{(+)})_c$$

则总闭合误差为:

$$\begin{aligned} \varepsilon' &= \varepsilon_+ - \varepsilon_- = -[(F_{pH}^{(+)})_c \\ &\quad - (F_{pH}^{(-)})_c] + [(F_{pH}^{(+)})_c \\ &\quad - (F_{pH}^{(-)})_c] \\ &= -(\Delta m)_c + (\Delta m)_c \end{aligned}$$

ε' 为 ϕ 角的函数, 从而可导出:

$$\varepsilon'(\phi) = -(\Delta m)_c + 2F_H'' \cos(\nu + \omega)$$

$\varepsilon'(\phi)$ 服从高斯分布, 它的概率密度分布为:

$$P'(\phi) = N_2 \exp[-(\varepsilon'^2(\phi)/2E'^2)]$$

I 个重原子衍生物的反常散射联合概率密度分布为:

$$P_{\text{ano}}(\phi) = N_2 \exp \left[- \sum_{i=1}^I (\varepsilon_i'^2(\phi)/2E_i'^2) \right]$$

式中:

$$\begin{aligned} \cos \nu &= [(F_H)_c^2 + F_p(A_H \cos \phi \\ &\quad + B_H \sin \phi)] / [(F_H)_c(F_{pH})_c] \end{aligned}$$

$$\sin \nu = [F_p(B_H \cos \phi - A_H \sin \phi)] / [(F_H)_c(F_{pH})_c]$$

$$\cos \omega = (A_H A_H'' + B_H B_H'') / [(F_H'')_c(F_H)_c]$$

$$\sin \omega = (A_H B_H'' - A_H'' B_H) / [(F_H'')_c(F_H)_c]$$

ν 为 F_{pH} 与 F_H 的夹角, ω 为 F_H 与 F_H'' 的夹角。

A_H, B_H 为 F_H 的实、虚部。

A_H'', B_H'' 为 F_H'' 的实、虚部。

将这种方法与同晶置换法联合使用, 有助于消除多解, 提高测定相角的精确度。

3. 计算“同晶”与“反常”联合概率密度分布

$$\begin{aligned} P(\phi) &= \prod_{i=1}^I P_i(\phi) \prod_{i=1}^I P_i'(\phi) \\ &= P_{\text{iso}}(\phi) P_{\text{ano}}(\phi) \end{aligned}$$

从而可以求出最佳结构因子 F_b 的实、虚部:

$$A_b = F_p N \int_0^{2\pi} \cos \phi P(\phi) d\phi$$

$$B_b = F_p N \int_0^{2\pi} \sin \phi P(\phi) d\phi$$

$$F_b = A_b + iB_b$$

最佳相角 $\phi_b = \text{tg}^{-1}(B_b/A_b)$

式中 N 为联合概率密度分布的归一化因子。

从联合概率密度分布曲线中可以获得最可几相角 ϕ_M 和最佳相角 ϕ_b 。独立的重原子衍生物越多,用概率法消除误差效果越好。 ϕ_b 就越接近于 α_p 的真值。用 F_b 进行富里叶合成,则得到均方差最小的电子云密度分布。

另外,为了衡量所算的相角的可信度,程序中还计算了以下三个因子:

$$(1) m = \frac{\sum_{j=1}^n [(A_b^2 + B_b^2)/F_b^2]_j^2}{n}$$

$$(2) \overline{|\phi_b - \phi_M|} = \frac{\sum_{j=1}^n |\phi_b - \phi_M|_j}{n}$$

$$(3) \overline{|\phi_b - \alpha_H|} = \frac{\sum_{j=1}^n |\phi_b - \alpha_H|_j}{n}$$

m 因子与 E 、 E' 取值有关,只有当标准偏差取值合理时才能客观地反映相角可信度,它越接近于 1 越好。 $\overline{|\phi_b - \phi_M|}$ 反映概率分布曲线的质量,它越小越好。 $\overline{|\phi_b - \alpha_H|}$ 可以用来衡量相对 K 的取值是否恰当,它越接近 $\pi/2$ 越好。

程序中第一轮의 E 和 E' 是估计给出,以后各轮则采用上轮计算的结果。

二、程序结构

1. 输入参数。
2. 衍射强度数据分类、排队,统一。
3. 判断是否全部衍生物参加运算。
4. 计算 $\sin^2\theta/\lambda^2$ 。
5. 计算温度因子(各向同性或各向异性)
6. 计算 f_{00}
7. 计算 $A_H, B_H, A''_H, B''_H, F_H, F''_H, \alpha_H$ 。
8. 判断是否中心对称衍射点:如是,则转 9,否则转 10。
9. 按中心对称公式计算 $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$, 转 16。
10. 按非中心对称公式计算 $(F_{pH})_o, (F_{pH})_c$,

以及 $\varepsilon_i^2(\phi)/2E_i^2$ 。

11. 判断是否计算反常散射概率密度分布:如是,则转 12,否则令 $P'(\phi) = 1$, 转 14。
12. 计算 $(\Delta m)_o, (\Delta m)_c, \varepsilon_i^{*2}(\phi)/2E_i^{*2}$ 。
13. 计算 $N_2, P_{ano}(\phi)$ 。
14. 计算 $N_1, P_{iso}(\phi)$ 。
15. 计算 $N, P(\phi), \phi_M, A_b, B_b, \phi_b, m_o$
16. 用中心对称公式计算 $P_+, P_-, \phi_M, \phi_b, m$ 等,并输出计算结果。
17. 控制台开关控制是否要逐点输出 $h, K, l, P_{iso}(\phi), P_{ano}(\phi), P'(\phi), \phi_M, \phi_b, m$ 等。
18. 重新计算 E, E' , 并输出。
19. 计算 $m, \overline{|\phi_b - \phi_M|}, \overline{|\phi_b - \alpha_H|}$, 并输出。
20. 输出所有衍射点的 h, K, l, ϕ_M, ϕ_b 和 m_o
21. 停机。
22. 重新按“继续”键,转去做电子云密度程序或去进行交替修正母体相角。

三、程序功能

本程序可以在任意空间群内计算 $0^\circ - 360^\circ$ 范围内间隔为 5° 的蛋白质母体相角的概率密度分布,并求出最可几相角、最佳相角以及 m 因子等。

四、讨论

本程序包括一个通用的计算 F_c 的程序,并与一个通用电子云密度计算程序相连,同时留有进行交替修正的接口。凡重原子衍生物不超过五个,独立区衍射点数在 5000 以内均可使用该程序计算。此程序编成后曾用两个不同空间群的模型数据验算,并在去五肽胰岛素晶体结构测定中得到实际应用^[7]。

在程序的编制过程中林政炯、戴金壁同志曾提出宝贵意见,王家槐、李家瑶同志曾协助部分验证工作,在此一并致谢。

参考文献

- [1] Harker, D.: *Acta Cryst.*, 9, 1, 1956.

[2] Blow, D. M. and Crick, F. H. C.: *Acta Cryst.*, 12, 794, 1959.
 [3] Narth, A. C. T.: *Acta Cryst.*, 18, 212, 1965.
 [4] Kartha, G.: *Acta Cryst.*, 14, 680, 1961.
 [5] Cullis, A. F. and et al.: *Proc. Roy. Soc.*, A265,

15, 1961.
 [6] Motthews, B. W.: *Acta Cryst.*, 20, 230, 1966.
 [7] 李家瑶等: 待发表。

[本文于 1979 年 12 月 25 日收到]

体表电位频谱分析技术

刘承德 陈鸿钧 朱庆富 陶宝山

(安徽中医学院经络研究所)

在生物医学中,广泛地应用心电图、脑电图、肌电图、胃电图以及皮肤电反射及神经动作电位的记录等,检测和分析各种器官的机能状态。根据检测的结果可以作出体表电位随时间变化的曲线,进行瞬态反应的时程和幅值的分析。但从理论上讲,对各种器官功能作动态分析,应当从该系统的瞬态反应及其频率特性两个方面着手,只有这样才能得到较全面的知识。从瞬态反应分析可看到该系统活动的时序规律,而从频率特性分析则可看出该系统各组成成份的活动规律。实际上,一个线性系统的瞬态反应中包含了它的频率特性的全部信息。这种瞬态反应和频率特性的对应关系,在数学上可阐述为设在时间轴上出现的连续信号为 $x(t)$,它的频谱为 $X(f)$,则两者的关系如下:

$$X(f) = \int_{-\infty}^{\infty} x(t) \exp(-i2\pi ft) dt \quad (1)$$

(1) 式即富里叶变换公式。它能将信号在时间域内的研究转变为在频率域内的研究,即能对信号作频谱分析。

过去,由于富里叶变换的计算量太大,有关器官活动规律的资料,只能从瞬态反应一个方面去分析,而把大量的频率域的有用信息都丢失了。现在有了成熟的“快速富里叶变换”(FFT)使计算量大大减少,从而可以借助通用数字电子计算机作频谱分析。本文介绍在体表电位的研究中,运用 FFT 时,如何解决离散数值抽样,进行 Basic 语言程序设计等问题。

离散数值抽样

各种体表电位的记录,都是连续的时间函数,记为 $x(t)$ 。我们是用数字计算机进行 FFT 运算,故首先要对连续信号 $x(t)$ 进行离散抽样,即按一定的时间间隔 Δ 进行取值,得到离散序列 $x(n\Delta)$, (其中 $\Delta = \dots -2, -1, 0, 1, 2, \dots$)。根据抽样定理及我们的经验,认为在离散抽样中必须注意以下几个重要的参数关系。

首先要注意 Δ 的选择,使其保证由 $x(n\Delta)$ 可以恢

复原始信号 $x(t)$ 。由抽样定理 (1) 可知,必须使: $\Delta \leq 1/2f_h$; 式中 f_h 为频谱中的最高可能频率。

其次要注意到频率的分辨率 F (或称频率分量间的增量;以 Hz 为单位)和最小记录长度 t_p (以秒为单位)的关系,即根据 $t_p = 1/F$ 的关系来选择。最后要在保持其他参量不变的情况下,增加被分析谱的高频范围或提高频率分辨率的唯一办法是增加记录长度内的点数 N 。如果 f_h 和 F 都已给定, N 必须满足 $N \geq 2f_h/F$ 。

实践证明,上述几个重要参数的关系是实现 FFT 算法必须满足的最低条件。

考虑一个 N 点的 FFT,由抽样得到时间序列 $x(K)$, ($K = 0, 1, 2, \dots, N-1$)。对应于公式 (1),根据离散富氏变换 (DFT) 公式:

$$X(f) = \sum_{K=0}^{N-1} x(K) \exp(-i2\pi Kf/N) \quad (2)$$

式 (2) 中 $f = 0, 1, 2, \dots, N-1$ 。 N 个点的频率序列 $X(f)$ 称为时间序列 $x(K)$ 的离散频谱,是一个复数。时间序列 $x(K)$ 的振幅谱定义为:

$$A(k) = \{Re^2[X(f)] + Im^2[X(f)]\}^{1/2} \quad (3)$$

时间序列 $x(K)$ 的相位谱定义为:

$$P(k) = \text{arc tg} \{Im[X(f)]/Re[X(f)]\} \quad (4)$$

FFT 的 Basic 子程序

在有关 FFT 的书籍中^[2],都是用 Fortran 或 Algol 语言编写的,这就必须使用大型通用电子计算机。为了适应只配有 Basic 语言的 Dsj-130 小型通用机,我们自行编制了 FFT 的 Basic 子程序,已上机通过,并应用于两种体表电位的频谱分析。这个 FFT 的 Basic 子程序为:

```
200 REM THIS IS FFT AND IFFT SUBROUTINE
201 IF Y = 0 GOTO 205
202 FOR I = 0 TO N4
203 LET S[1] = -S[1]
```