

双波长双相角结构不变量的比邻原理*

孙宏林 张 纲 郭东耀

(吉林大学理论化学研究所)

1988年5月20日收到

本文得到双波长双相角结构不变量一级比邻 (Neighborhood) 的一个严格解。对比计算及加随机误差计算表明,文献[5]的初级近似方法是简明可靠的。Hauptman 做出的双相角结构不变量只与化学组成有关的结论在双波长情况实际上仍然有效。提出一个快速计算方案。

一、引 言

双波长或多波长方法是解决生物大分子晶体结构的理想技术之一,它可以从一个晶体获得两套或多套实验数据。自从比邻原理问世以来^[1,2],同传统蛋白质结构的技术相结合的比邻问题一直是受到重视的^[3,4]。双波长方法也是人们感兴趣的课题之一。1986年郭东耀和 Hauptman 在研究联合概率定积分定理时,曾给出该问题的一个近似解^[5,6],计算的初步结果看来是较好的。由于这个问题的重要性,在理论上尚需严格论证,在实践上寻找合适的计算技术也是有益的。本文就是在这种学术背景下进行的。

二、关于联合概率的近似方法和结果

具有反常散射的双波长双相角结构不变量有六个

$$\begin{aligned} Q_1 &= \varphi_H + \varphi_{\bar{H}}, Q_2 = \phi_H + \phi_{\bar{H}}, Q_3 = \varphi_H - \phi_H, \\ Q_4 &= \varphi_{\bar{H}} + \phi_{\bar{H}}, Q_5 = \varphi_H + \phi_{\bar{H}}, Q_6 = \varphi_H - \phi_{\bar{H}}, \end{aligned} \quad (1)$$

式中四个是独立的。它们是由构成一级比邻的四个归一化结构因子的模

$$|E_H|, |E_{\bar{H}}|, |G_H|, |G_{\bar{H}}|$$

所对应的相角 $\varphi_H, \varphi_{\bar{H}}, \phi_H, \phi_{\bar{H}}$ 所组成的。 E 与 G 用来区别两种不同的波长。归一化结构因子的模与相角的概率参数分别用

$$R, \bar{R}, S, \bar{S}; \phi, \bar{\phi}, \psi, \bar{\psi}$$

来表示。 P 1空间群的这四个归一化结构因子联合概率密度为

$$\begin{aligned} P(R, \bar{R}, S, \bar{S}; \phi, \bar{\phi}, \psi, \bar{\psi}) &= \frac{R\bar{R}S\bar{S}}{(2\pi)^3} \int_{\rho, \bar{\rho}, \sigma, \bar{\sigma}=0}^{\infty} \int_{\theta, \bar{\theta}, x, \bar{x}=0}^{2\pi} \\ &\cdot \rho \bar{\rho} \sigma \bar{\sigma} \exp \{ -i [R \rho \cos(\theta - \phi) + \bar{R} \bar{\rho} \cos(\bar{\theta} - \bar{\phi}) + S \sigma \cos(x - \psi)] \} \end{aligned}$$

* 国家自然科学基金资助的课题。

$$+ \bar{S} \sigma \cos(\bar{x} - \phi)] \} \prod_{j=1}^N q_j d\rho d\bar{\rho} d\sigma d\bar{\sigma} d\theta d\bar{\theta} dx d\bar{x}, \quad (2)$$

式中

$$\begin{aligned} q_j = & \left\langle \exp \left\{ i \left[\frac{|f_j| \rho}{\alpha_{20}^{1/2}} \cos(\delta_j + 2\pi H r_j - \theta) \right. \right. \right. \\ & + \frac{|f_j| \bar{\rho}}{\alpha_{20}^{1/2}} \cos(\delta_j - 2\pi H r_j - \bar{\theta}) \\ & + \frac{|g_j| \sigma}{\alpha_{02}^{1/2}} \cos(\Delta_j + 2\pi H r_j - x) \\ & \left. \left. \left. + \frac{|g_j| \bar{\sigma}}{\alpha_{02}^{1/2}} \cos(\Delta_j - 2\pi H r_j - \bar{x}) \right] \right\} \right\rangle_{r_j}. \end{aligned} \quad (3)$$

在上式中 f_j 及 g_j 代表两个不同波长的原子散射因子, δ_j 及 Δ_j 为它们的相角.

$$\alpha_{mn} = \sum_{j=1}^N |f_j|^m |g_j|^n. \quad (4)$$

把 q_j 展开成级数, 而且令

$$\begin{aligned} y = & -\frac{1}{2} \left\{ \frac{|f_j|^2}{\alpha_{20}} \rho \bar{\rho} \cos(2\delta_j - \theta - \bar{\theta}) \right. \\ & + \frac{|g_j|^2}{\alpha_{02}} \sigma \bar{\sigma} \cos(2\Delta_j - x - \bar{x}) \\ & + \frac{|f_j| |g_j|}{(\alpha_{20} \alpha_{02})^{1/2}} [\rho \sigma \cos(\delta_j - \Delta_j - \theta + x) + \rho \bar{\sigma} \cos(\delta_j + \Delta_j - \theta - \bar{x}) \\ & \left. + \rho \sigma \cos(\delta_j + \Delta_j - \bar{\theta} - x) + \rho \bar{\sigma} \cos(\delta_j - \Delta_j - \bar{\theta} + \bar{x}) \right] \}, \end{aligned} \quad (5)$$

则

$$\begin{aligned} q_j = & 1 - \frac{1}{4} \left[\frac{|f_j|^2}{\alpha_{20}} (\rho^2 + \bar{\rho}^2) + \frac{|g_j|^2}{\alpha_{02}} (\sigma^2 + \bar{\sigma}^2) \right] + y + \dots \\ \approx & \exp \left\{ -\frac{1}{4} \left[\frac{|f_j|^2}{\alpha_{20}} (\rho^2 + \bar{\rho}^2) + \frac{|g_j|^2}{\alpha_{02}} (\sigma^2 + \bar{\sigma}^2) \right] + y \right\} \\ = & \exp \left\{ -\frac{1}{4} \left[\frac{|f_j|^2}{\alpha_{20}} (\rho^2 + \bar{\rho}^2) + \frac{|g_j|^2}{\alpha_{02}} (\sigma^2 + \bar{\sigma}^2) \right] \right\} \exp y. \end{aligned} \quad (6)$$

把 (6) 式直接代入 (2) 式, 用传统积分方法曾得到一个严格解. 由于公式极为复杂以及实现计算方面的严重困难, 尚不能看出其实用价值, 这里从略.

把 $\exp y$ 做级数展开, 可以得到不同近似程度的解. 文献 [5] 用的是一次近似解

$$q_j \approx \exp \left\{ -\frac{1}{4} \left[\frac{|f_j|^2}{\alpha_{20}} (\rho^2 + \bar{\rho}^2) + \frac{|g_j|^2}{\alpha_{02}} (\sigma^2 + \bar{\sigma}^2) \right] \right\} (1 + y). \quad (7)$$

看来 (7) 式相当粗略, 结果如何应视其高次项能否不计而定. 本文将对此加以考察.

把 $\exp y$ 展到二次项

$$q_j \approx \exp \left\{ -\frac{1}{4} \left[\frac{|f_j|^2}{\alpha_{20}} (\rho^2 + \bar{\rho}^2) + \frac{|g_j|^2}{\alpha_{02}} (\sigma^2 + \bar{\sigma}^2) \right] \right\} \left(1 + y + \frac{1}{2} y^2 \right). \quad (8)$$

通过对无穷级数中最重要的修正项 $\frac{1}{2} y^2$ 的引入, 来了解 (7) 式的可靠程度.

把(8)式代入(2)式,应用联合概率定积分定理^[7]得到双波长四个归一化结构因子的联合概率密度为

$$\begin{aligned} P^0(R, \bar{R}, S, \bar{S}; \phi, \phi, \phi, \phi) = & A \exp[2y_1 R \bar{R} \cos(\phi + \phi - \eta_1) \\ & + 2y_2 S \bar{S} \cos(\phi + \phi - \eta_2) + 2y_3 R S \cos(\phi - \phi - \eta_3) \\ & + 2y_4 \bar{R} S \cos(\phi + \phi - \eta_4) + 2y_5 R \bar{S} \cos(\phi + \phi - \eta_5) \\ & + 2y_6 \bar{R} \bar{S} \cos(\phi - \phi - \eta_6)], \end{aligned} \quad (9)$$

上式中 A 为归一化常数。

在已知四个归一化结构因子的模的情况下,条件概率

$$P^0(Q_i | R, \bar{R}, S, \bar{S}) = \frac{1}{K_i} \exp[z_i \cos(Q_i - \xi_i)] \quad i = 1, \dots, 6. \quad (10)$$

上式中的 $1/K_i$ 为归一化常数。

(10)式同文献[5]结果的不同之处在于它包含了 $\exp y$ 的二次近似项。这个含有更高次近似项的公式能否对文献[5]的公式在实际上有修正作用,可用于考察 $\exp y$ 级数展开式的收敛性快慢以及文献[5]的简明公式的实用价值。

三、计算结果

全部计算工作是在 Panasonic 微机机上以双精度完成的。采用的晶体是药物小结构柯卡因甲基碘化物^[8,9],空间群为 $P 2_1 2_1 2_1$, 分子的化学式为 $C_{18}H_{21}NO_4I$, $z = 4$, 全晶胞含四个反常原子碘, 92个正常非氢原子。949个密勒指数不含零的 $|E| > 0.3$ 的 Friedel 对参加计算。对文献[5]的一级近似公式和本文的二级近似式(10),做了对比计算。表1给出由(1)式定义的六个结构不变量的有关结果。表1中的统计平均值一律为对全部 Friedel 对的总体平均值。 $\langle Q_i \rangle_{\text{真}}$ 是供参考用的真实值之平均。 ξ'_i 为文献[5]的结果, ξ_i 为本文公式的结果。 $\Delta \xi'_i$ 及 $\Delta \xi_i$ 代表有关量的绝对误差。单位为度。双波长为 $\text{CrK}\alpha(E_H)$ 和 $\text{CuK}\alpha(G_H)$ 。

表 1

Q_i	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6
计算结果						
$\langle Q_i \rangle_{\text{真}}$	61.88	29.65	17.06	46.71	46.72	17.07
$\langle \xi' \rangle$	60.75	28.29	16.27	44.54	44.54	16.27
$\langle \xi \rangle$	60.68	28.32	16.32	44.55	44.55	16.32
$\langle \Delta \xi' \rangle$	14.04	6.40	3.81	10.18	10.50	4.15
$\langle \Delta \xi \rangle$	14.08	6.39	3.82	10.16	10.49	4.14

上面的结果再次有力地证明文献[5]的非常简便的近似公式,对小分子而言是很可信的,以致于更复杂的二级近似式(10)不论对平均值,还是对绝对误差均不能再给出进

1) 文中所涉及的 y_i 和 $\eta_i (i = 1, \dots, 6)$, z_i 和 $\xi_i (i = 1, \dots, 6)$ 的定义可向作者直接索取。

一步的改进。

(10) 式中的 ξ_i 与 $R, \bar{R}, S, \bar{S}$ 有数学上的函数关系, 这种关系的强弱是人们关心的另一个问题。因为 Hauptman 对具有反常散射单波长双相角结构不变量的简单情况, 得出一个很不寻常的结论: 它只与化学组成有关^[3]。这就是说结构如何或结构因子的模及其观测误差将不影响其概率估计值, 按他的公式只对概率的高低有影响。如果这是事实, 这将为双相角结构不变量的实际应用提供很大的方便。本文的 (10) 式比 Hauptman 的公式复杂得多, 如果他的结论对这种复杂情况也能成立, 则 ξ_i 应该对 $R, \bar{R}, S, \bar{S}$ 呈现弱的函数关系。由于数学关系的复杂性, 只能通过计算来估计这种函数关系的强弱, 这是本文第二部分计算内容。

我们把参加计算的 $|E|$ 分别加 5%, 10%, 20% 直到 30% 和 50% 的随机误差, 然后重复上面叙述过的计算。结果发现它们对 ξ_i 和 $\Delta\xi_i$ 都没有实际影响, 看来 Hauptman 的结论对双波长情况也是成立的, 即 ξ_i 与 $R, \bar{R}, S, \bar{S}$ 呈现极弱的函数关系。如果事实是这样, (10) 式的计算问题将得到极大的简化, 在实际应用中会带来方便。

在 (10) 式中如果 ξ_i 与 $R, \bar{R}, S, \bar{S}$ 呈现极弱的函数关系, 则它们基本上是由含有反常散射的原子散射因子 f_j 与 g_j 决定的 ($j = 1, \dots, N$)。它们是与 $\sin\theta/\lambda$ 有关的, 即 ξ_i 是 $\sin\theta/\lambda$ 的函数。按照这种分析, 又考虑到文献 [5] 的公式既简便又可信, 本文建议了一种快速计算 Q_i ($i = 1, \dots, 6$) 的简化方案。

选取 $\sin\theta/\lambda$ 均匀分布于 0.1—1.0 范围内的若干个 Friedel 对, 对这几个 Friedel 对用文献 [5] 的近似公式计算 ξ_i ($i = 1, \dots, 6$) (Q_i 的概率估计值), 做出 6 条 $\xi_i - \sin\theta/\lambda$ 曲线。这 6 条曲线可用做所有衍射点的插值工作曲线。图 1 给出四条独立曲线 (Q_1, Q_2, Q_3 和 Q_4) 计算结果。出人意料的是它们竟是非常理想的四条直线。绘制这些曲线的数据由表 2 给出。

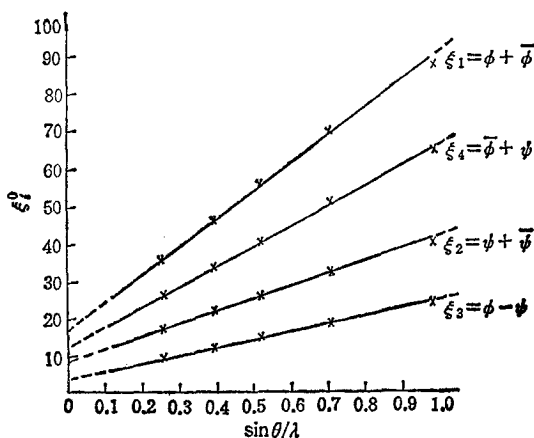


图 1

表 2

(h, k, l)	$\sin\theta/\lambda$	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6
(1, 1, 6)	0.1263	26.54	12.85	6.445	19.39	19.37	6.452
(2, 3, 4)	0.2531	35.80	17.11	9.045	26.23	26.23	9.045
(4, 4, 1)	0.3927	46.41	21.87	12.07	33.99	33.99	12.07
(6, 4, 7)	0.5141	55.58	25.94	14.73	40.70	40.70	14.73
(8, 6, 4)	0.7012	69.49	32.20	18.87	51.02	51.02	18.87
(10, 10, 3)	0.9820	87.60	40.82	24.35	65.01	64.82	24.44

四、结 论

1. 高次近似的计算支持文献[5]的简化公式。
2. 双波长双相角结构不变量的概率估计值与晶体的化学组成有关, 与结构因子的模呈极弱的函数关系。
3. 用文献[5]的简明公式制作 $Q_i - \sin \theta / \lambda$ 工作曲线的近似计算方法, 是快速可靠的。

- [1] H. Hauptman, *Acta Cryst.*, **A31**(1975), 680.
- [2] C. Giacovazzo, *Acta Cryst.*, **A33**(1977), 933.
- [3] H. Hauptman, *Acta Cryst.*, **A38**(1982), 632.
- [4] H. Hauptman, *Acta Cryst.*, **A38**(1982), 289.
- [5] H. Hauptman, Guo Dongyao, *Mol. Struc. Inter. Symp. Beijing*, (1986), p. 195.
- [6] Guo Dongyao, H. Hauptman, *American Cryst. Assoc.*, **14**(1986), 46.
- [7] Guo Dongyao, H. Hauptman, *American Cryst. Assoc.*, **13**(1985), 45.
- [8] M. Shen, J. R. Ruble, *Acta Cryst.*, **B31**(1975), 2706.
- [9] 郭东耀, H. Hauptman, *科学通报*, **23**(1987), 1791.

TWO-WAVELENGTH NEIGHBORHOOD PRINCIPLE OF TWO-PHASE STRUCTURE INVARIANTS

SUN HONG-LIN ZHANG GANG GUO DONG-YAO

(Theoretical Chemistry Institute of Jilin University, Changchun)

ABSTRACT

An approximate result of first neighborhood is given in ref [5]. In consideration of the importance of strict treatment about this problem, a new formula is derived. According to the computational comparison between the two formulas, it can be concluded that the approximate method in ref [5] is very simple, convenient and reliable. It is also concluded by adding random errors to the magnitudes of normalized structure factors that Hauptman's conclusion, namely, the two-phase structure invariants are only dependent on chemical composition and independent of the structure factors, is still correct in the two-wavelength case. Considering this fact, we suggest a fast computational method of the two-wavelength structure invariants which may be useful in practical macrostructure crystallography.