

双相角结构不变量的 Neighborhood 原理

——单个 Friedel 相角对的概率估计

郭 东 耀

(吉林大学理论化学研究所, 长春)

H. Hauptman

(布法罗医学基金会, 美国)

通过 Friedel 相角对统计性质的研究, 我们理解到按概率公式^[1]

$$\varphi_h + \varphi_{\bar{h}} \simeq -\xi, \quad (1)$$

$-\xi$ 的统计平均值基本上能反映真实的 Friedel 相角对取正值的倾向性及其统计平均值本身。这启发人们进一步去探求单个相角对这个最引人关注的问题。

一、一级 Neighborhood 方法

公式(1)是假定在已知 $|E_h|$ 及 $|E_{\bar{h}}|$ 这两个模的条件下, 推导出来的概率式。它可能用于估计双相角结构不变量的前提是

$$A = 2|E_h||E_{\bar{h}}|X/(1 - X^2) \quad (2)$$

比较大, 即条件概率分布的方差比较小。以上是一级 Neighborhood 的理论结果。

以(2)式为判据, 用(1)式来计算药物小结构柯卡因甲基碘化物 (cocaine methiodide) 时^[2], 不论在判别 $\varphi_h + \varphi_{\bar{h}}$ 的符号(正值)或估计数值方面, 结果都是满意的。在此基础上, 着

表 1

衍 射 号 数	正确符号数	错误符号数	$\langle \varphi_h + \varphi_{\bar{h}} \rangle$	$\langle -\xi \rangle$	$\langle \text{绝对误差} \rangle$
1—100	82	18	1.4	1.9	1.3
101—200	96	4	3.2	3.1	1.4
201—300	100	0	5.1	4.4	1.6
301—400	99	1	5.7	5.6	1.8
401—500	98	2	5.3	6.1	2.0
501—600	99	1	5.7	6.5	2.0
601—700	99	1	5.9	6.7	1.9
701—800	99	1	6.5	7.0	1.9
801—900	99	1	6.3	7.3	2.2
901—1000	99	1	7.5	7.8	2.2
总体情况	970	30	5.3	5.6	1.8

本文 1987 年 3 月 26 日收到。

* 有关实验工作曾受到美国 NSF 基金 No. CHE 8203930 的资助。

重对分子生物大结构蛋白细胞色素 C₅₅₀ 的 Pt 衍生物进行了计算^[1]。CuK_α 辅射晶胞内 4 个 Pt, 4 个 Fe, 24 个 S 为反常原子; 812 个 O, 680 个 N, 2548 个 C 为正常原子。其它有关计算细节见本文 (I)。1000 个 $|E| > 2.0$ 的 Friedel 对的计算结果列于表 1。其中的角度用度表示。由表可见, Hauptmann 的一级 Neighborhood 公式 (1), 能以很高的可靠程度判断 $\varphi_h + \varphi_k$ 的符号, 但计算分子生物大结构 $\varphi_h + \varphi_k$ 的数值时误差相当大。如果想用于含有实验误差的未知结构时, 情况会更差些。

另外, 换用 CrK_α 波长时也得到类似的结果。在 1000 对数据中, (1) 式可以对 964 个符号做出正确的判断, 错的只有 36 个。

二、高级 Neighborhood 理论

为了改进分子生物大结构的 $\varphi_h + \varphi_k$ 的估计值, 高级 Neighborhood 理论看来是有益的。只要对三相角结构不变量的联合概率分布进行对 Φ_1 和 Φ_2 积分, 得到的新条件概率分布就是双相角结构不变量的一种高级 Neighborhood 公式。在满足 $h + k + l = 0$ 的条件下, E_h, E_k, E_l 的联合概率分布为:

$$P = \frac{R_1 R_2 R_3 R_1 R_2 R_3}{(2\pi)^{12}} \int_{\Phi_1=0}^{\infty} \int_{\Phi_2=0}^{2\pi} \rho_1 \rho_2 \rho_3 \rho_1 \rho_2 \rho_3 \\ \times \exp\{-i[R_1 \rho_1 \cos(\theta_1 - \Phi_1) + R_2 \rho_2 \cos(\theta_2 - \Phi_2) + R_3 \rho_3 \cos(\theta_3 - \Phi_3) \\ + R_1 \rho_1 \cos(\theta_1 - \Phi_1) + R_2 \rho_2 \cos(\theta_2 - \Phi_2) + R_3 \rho_3 \cos(\theta_3 - \Phi_3)]\} \\ \cdot \prod_{i=1}^N q_i d\rho_1 d\rho_2 d\rho_3 d\rho_1 d\rho_2 d\rho_3 d\theta_1 d\theta_2 d\theta_3 d\theta_1 d\theta_2 d\theta_3,$$

式中 R 及 Φ 分别代表有关的 $|E|$ 及相角 φ 。完成 12 重积分, 并引入

$$D_1 = \Phi_1 + \Phi_1, D_2 = \Phi_2 + \Phi_2, \quad (3)$$

$$R_1 = |E_h|, R_1 = |E_k|, R_2 = |E_k|, R_2 = |E_k|, R_3 = |E_l|, R_3 = |E_l|, \quad (4)$$

$$x_\mu = \frac{2R_\mu R_\mu X_\mu}{1 - X_\mu^2}, \tau_\mu = I_1(x_\mu)/I_0(x_\mu), \mu = 1, 2, 3, \quad (5)$$

然后把联合概率 P 对 Φ_1 及 Φ_2 积分, 就可以得到当给定 $R_1, R_1, R_2, R_2, R_3, R_3$ 时 D_1 和 D_2 的条件概率分布 $P(D_1 D_2 | R_1 R_1 R_2 R_2 R_3 R_3)$ 。

$$P(D_1 D_2 | R_1 R_1 R_2 R_2 R_3 R_3) = k \cdot \exp\{x_1 \cos(D_1 + \xi_1) + x_2 \cos(D_2 + \xi_2)\} \\ \cdot I_0 \left\{ \frac{2Y}{(1 - x_1^2)(1 - x_2^2)(1 - x_3^2)} \right\}, \quad (6)$$

其中 k 是归一化常数, 其它参数可以从文献 [1] 中找到。而 Y 的定义为

$$Y^2 = Z_0^2 [R_1^2 R_2^2 R_3^2 + \tau_3^2 R_1^2 R_2^2 R_3^2] + Z_1^2 [R_1^2 R_2^2 R_3^2 + \tau_3^2 R_1^2 R_2^2 R_3^2] + Z_2^2 [R_1^2 R_2^2 R_3^2 + \tau_3^2 R_1^2 R_2^2 R_3^2] \\ + Z_3^2 [R_1^2 R_2^2 R_3^2 + \tau_3^2 R_1^2 R_2^2 R_3^2] + 2\tau_3 R_1 R_1 R_2 R_2 R_3 R_3 [Z_0^2 \cos(D_1 + D_2 - \xi_3 - 2\xi_0) \\ + Z_1^2 \cos(D_1 - D_2 + \xi_3 + 2\xi_1) + Z_2^2 \cos(D_1 - D_2 - \xi_3 - 2\xi_2) \\ + Z_3^2 \cos(D_1 + D_2 + \xi_3 - 2\xi_3)] + 2Z_0 Z_1 [R_1 R_1 (R_2^2 R_3^2 + \tau_3^2 R_2^2 R_3^2) \cos(D_1 - \xi_0 + \xi_1) \\ + \tau_3 R_2 R_2 R_3 R_3 (R_1^2 + R_1^2) \cos(D_2 - \xi_0 - \xi_1 - \xi_3)] \\ + 2Z_0 Z_2 [R_2 R_2 (R_1^2 R_3^2 + \tau_3^2 R_1^2 R_3^2) \cos(D_2 - \xi_0 + \xi_2) \\ + \tau_3 R_1 R_1 R_3 R_3 (R_2^2 + R_2^2) \cos(D_1 - \xi_0 - \xi_2 - \xi_3)] \\ + 2Z_1 Z_3 [R_1 R_1 R_2 R_2 (R_3^2 + \tau_3^2 R_3^2) \cos(D_1 + D_2 - \xi_0 - \xi_3)]$$

$$\begin{aligned}
& + \tau_3 R_3 R_3 (R_1^2 R_2^2 + R_1^2 R_4^2) \cos(\zeta_0 - \zeta_2 + \xi_3)] \\
& + 2Z_1 Z_2 [R_1 R_1 R_2 R_2 (R_3^2 + \tau_3^2 R_4^2) \cos(D_1 - D_2 + \zeta_1 - \zeta_2) \\
& + \tau_3 R_3 R_3 (R_1^2 R_2^2 + R_1^2 R_4^2) \cos(\zeta_1 + \zeta_2 + \xi_3)] \\
& + 2Z_1 Z_3 [R_2 R_2 (R_1^2 R_3^2 + \tau_3^2 R_4^2 R_4^2) \cos(D_2 - \zeta_1 - \zeta_3) \\
& + \tau_3 R_1 R_1 R_3 R_3 (R_2^2 + R_4^2) \cos(D_1 + \zeta_1 - \zeta_3 + \xi_3)] \\
& + 2Z_2 Z_3 [R_1 R_1 (R_2^2 R_3^2 + \tau_3^2 R_4^2 R_4^2) \cos(D_1 - \zeta_2 - \zeta_3) \\
& + \tau_3 R_2 R_2 R_3 R_3 (R_1^2 + R_4^2) \cos(D_2 + \zeta_2 - \zeta_3 + \xi_3)].
\end{aligned} \quad (7)$$

(6) 式的优点在于可以利用更多的归一化结构因子的模。它还有如下三个特点:

1. 公式 (6) 和 (1) 存在着内在联系。由于 $\exp\{x_1 \cos(D_1 + \xi_1) + x_2 \cos(D_2 + \xi_2)\}$ 项的存在, 用 (1) 式估计出的 D_1 及 D_2 , 可以看成 (6) 式的近似值。
2. 由于 Y 项的存在, 使由 (1) 式计算出来的概率值得到修正。
3. 因为 (6) 式同三相角结构不变量 h, k, l 有关, 对于一个指定的 Friedel 对 h 及 $\bar{h}$, 可能找到许多组 h, k, l ; $i = 1, 2, \dots, p$, p 为组数。于是单个相角对 $\varphi_h + \varphi_{\bar{h}}$ 可以有 p 个估计值。这意味着可能通过统计平均的方法来改善单个相角对的估计值。

三、高级 Neighborhood 的计算

计算工作的核心问题是求 (6) 式的极大值。如果概率分布 $P(D_1 D_2 | R_1 R_1 R_2 R_2 R_3 R_3)$ (缩写为 $P(D_1 D_2)$) 的极大值在 D_{1m} 及 D_{2m} 处, 可把 D_1 分布的方差定义为

$$\text{Var } D_1 = \int_{D_{1m}-\pi}^{D_{1m}+\pi} (D_1 - D_{1m})^2 P(D_1 D_{2m}) dD_1 / \int_{D_{1m}-\pi}^{D_{1m}+\pi} P(D_1 D_{2m}) dD_1. \quad (8)$$

如果某个双相角结构不变量可以找到 p 个估计值, 则统计平均的估计值为

$$\langle D_{1mi} \rangle = \sum_{i=1}^p (D_{1mi} / \text{Var } D_{1i}) / \sum_{i=1}^p (1 / \text{Var } D_{1i}). \quad (9)$$

本文只用小结构柯卡因甲基碘化物做了计算方法上的探索, 以便为大结构的计算提供经验。 $P(D_h D_k)$ 在二维平面上的分布对 $|E|$ 比较敏感, 这反映了对方差的影响。图 1 是用 $|E_{1,5,1}| = 1.57, |E_{1,5,1}| = 1.73, |E_{1,5,7}| = 1.59, |E_{1,5,7}| = 1.42, |E_{2,10,6}| = 1.60, |E_{2,10,6}| = 1.69$ 计算出来的。图 2 是用 $|E_{1,5,1}| = 1.57, |E_{1,5,1}| = 1.73, |E_{1,6,10}| = 0.38, |E_{1,6,10}| = 0.46,$

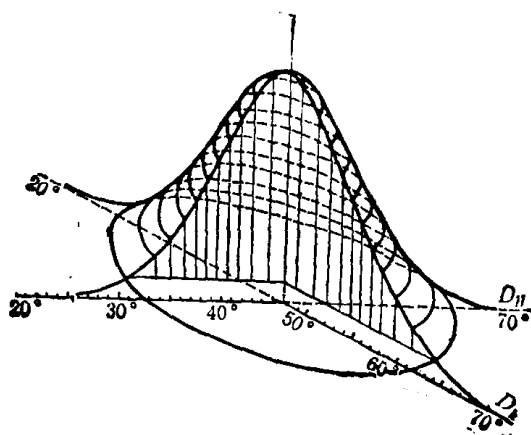


图 1

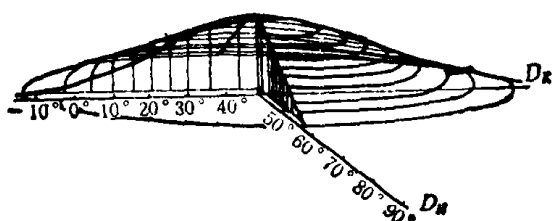


图 2

$|E_{1,1,9}| = 0.41$, $|E_{2,1,9}| = 0.53$ 计算出来的。

个别相角对的计算结果示于图 3 和图 4。由于在以前的所有理论工作中,并未特殊地考虑某些反射的密勒指数中含有 0 的这个细节,它可能导致对这种点结果不佳,至少在计算中也要留心这个问题。我们选取 $|E| > 1$ 的衍射点进行计算,结果按平均方差由小到大排队;从顶部取一批衍射点,分成密勒指数不含 0 的 (84 个相角对) 示于图 3, 含有 0 的 (66 个相角对) 示于图 4。

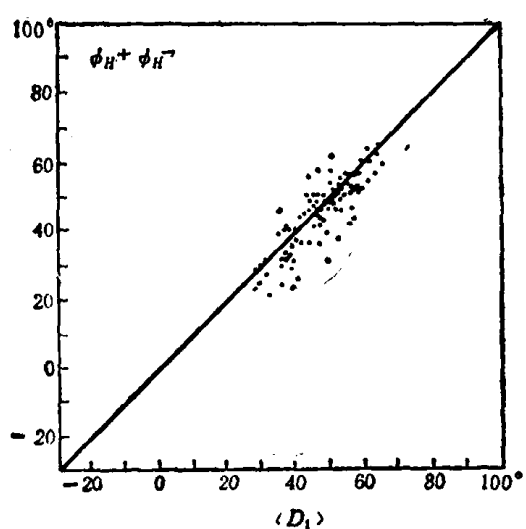


图 3

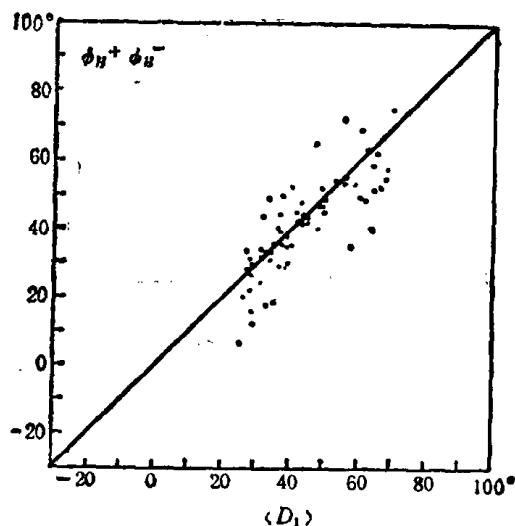


图 4

对这个药物小结构,个别相角值的对角分布规律相当明显,密勒指数含 0 的结果相对差些。应指出,在用一级 Neighborhood 方法判别分子生物大结构的 Friedel 相角对的符号时,虽然在一千个符号中只有三十几个是错的,但其中二十来个是密勒指数含有 0 的。这就提醒我们还有必要补充这方面的理论工作。

参 考 文 献

- [1] Hauptman, H., *Acta Cryst.*, **A38**(1982), 632.
- [2] Shen, M. and Ruble, J. R., *Acta Cryst.*, **B31**(1975), 2706.
- [3] Timkovich, R. and Dickerson, R. E., *J. Biol. Chem.*, **251**(1976), 4033.