

直接法中起始套反射的优化选择方法^{*1)}

郑 启 泰

(中国科学院生物物理研究所)

提 要

应用直接法测定晶体结构,通常多半采用记号附加法与多解记号附加法进行。在相角测定中,起始套反射的确定成为测定结构成功与否的关键。目前多半采用收敛图方法确定起始套反射。本文则利用衍射指标的相关关系,提出确定起始套反射的优化选择方法,并用一个已知结构例进行了验证。

一、概 述

应用直接法测定晶体结构,通常是由若干个反射(包括固定原点、对映体的反射和若干个附加反射)组成起始套,然后通过迭代方式导出全部未知反射的相角。因此,如何确定起始反射就成为直接法研究中的一个重要课题。记号附加法^[1]虽然确定了起始套选择的一般原则,但是,由于记号附加法存在着下述这样一些不足,即

在最初的推导步骤中必须依靠单一的相角关系式,而它们却常常包含着错误的信息;要承认导出的记号间的关系;起始套反射的选择尚需依赖于晶体对称群的投影有否对称中心等。

这就限制了它的应用效力。为克服这些弱点,Germain, Woolfson 提出了多解记号附加法(MULTAN)^[2,3],取代记号而赋于一组多解值($\pm\pi/4, \pm 3\pi/4; 0, \pi, \pm\pi/2$),并采用加权正切公式来推导未知相角值,

$$\begin{aligned}\tan \phi_h &\approx \frac{\sum_{h'} W_{h'} W_{h-h'} |E_{h'} E_{h-h'}| \sin(\phi_{h'} + \phi_{h-h'})}{\sum_{h'} W_{h'} W_{h-h'} |E_{h'} E_{h-h'}| \cos(\phi_{h'} + \phi_{h-h'})} \\ &= \frac{S_h}{C_h},\end{aligned}\quad (1)$$

式中 $W_h = \tanh[\sigma_3 \sigma_2^{-3/2} |E_h| (S_h^2 + C_h^2)^{1/2}] = \tanh\left(\frac{1}{2} \alpha_h\right)$,

$$\alpha_h = 2\sigma_3 \sigma_2^{-3/2} |E_h| (S_h^2 + C_h^2)^{1/2}, \quad \sigma_n = \sum_{j=1}^N Z_j^2.$$

MULTAN 的确立把直接法的应用推向了一个新的发展阶段。看来,求解晶体结构的直接

* 1979年2月20日收到。

1) 本文曾在1978年第二届全国物质结构会议报告过。

法,似乎已可以做为一个客观的数学方法处理了,仿佛不再涉及到使用的“技巧”问题了。

然而,使用 MULTAN 求解晶体结构失败的例子,促使人们重新估价这一结论。Lessinger 分析了 MULTAN 的失败例子^[4],曾详细地论述了导致直接法失败的原因,并且提出用增加起始套反射数目的方法来避免失败。但是,有时在增加了起始套反射数目之后,也仍然无法解出结构,这就提出了一个究竟应该怎样选择起始套反射的问题,因为单纯地增加起始套反射的数量,并不一定能确保方法的成功使用。

起始套反射的选择是个由多方面因素交织的问题^[5,6],它是目前运用直接法测定中等复杂程度有机分子结构成功与否的关键。

二、问 题

应用直接法测定晶体结构的过程,是以起始套反射为始,逐轮进行迭代直到全部未知相角导出并获得相容(收敛)的结果时为止。在起始十余步的推导中,常常隐含着一定的错误信息(有时甚至是以强 Σ_2 关系式出现),它左右着全过程的成功与失败(表 1)。对于中心对称结构,如果欲导出的一个反射的符号是由两个 Σ_2 关系式表示的,而且两者具有一致的结果,若令 p_1, p_2 表示各自的概率, p_+ 表示导出符号的总概率,则

表 1

N (晶胞中原子数)	十步中无错误的概率	十步中至少有一次错误的概率
25	0.998	1.000
50	0.962	1.000
100	0.834	0.987

$$p_+ = \frac{p_1 p_2}{1 - p_1 - p_2 + 2p_1 p_2}, \quad (2)$$

当 $p_1 = p_2 = 0.80$ 时, $p_+ = 0.94$;

当 $p_1 = p_2 = 0.85$ 时, $p_+ = 0.97$ 。

这表明:一对一致的但联系为弱的(概率值偏低)符号关系式,能给出一个强的符号指示(M. M. Woolfson, 1968)。而在应用(1)式推导相角时,一对表示不一致的相角关系(其中有一个是错误的),其结果是相应的权函数 W 值变小,从而减少了它在循环迭代中的作用。这说明,倘能在相角推导的初始阶段,使每个欲导出的相角是同两个(或更多) Σ_2 关系式联系,那么就有可能在逐轮迭代中,消除起始步骤中错误信息的干扰,从而获得正确的相角。

怎样才能实现这一目的?用增加起始套反射数目的方法,有可能达到这一目的(但也并非总是这样),但是每增加 n 个反射,解的套数就要以 4^n (至少是 2^n) 倍增加,这是个可怕的数字!那么能否在不增加或少增加起始套反射的条件下实现这一目的呢?

这就是反射指标优化选择方法企图回答的问题。

三、方 法

当 X 射线通过晶体时, 实现一次傅里叶变换, 隐含晶体结构信息的衍射线, 分布在以衍射指标 (H, K, L) 表示的衍射空间中. 对于一个中等大小的有机分子, 其衍射点数目约为 2000—4000 个左右. 运用直接法测定结构, 一般是选择大于一定 $|E|$ 值的一组反射, 其数目约为独立原子数目的 5—7 倍(最多 10 倍), 在这样一个范围内, 衍射指标分量 (H, K, L) 的分布, 仍然存在着一致分布的特征^[7], 同时也存在着某些分量出现频率较大的差异, 表 2 所列是一个中心对称结构 ($C_{14}H_{16}O_6$), 空间群为 $P2_1/c$; 表 3 所列是一个非中心对称结构 ($C_{21}H_{32}O_5$), 空间群为 $P2_1$.

表 2 $s.g. P2_1/c$, $|E| \geq 2.0$, 118 个衍射点

出现次数 指标分量类型	指标	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
H		12	8	6	13	14	19	11	7	5	16	4	—	1						
K		7	19	19	22	22	18	7	2											
L		4	6	10	13	2	9	5	7	5	8	4	8	8	7	7	6	5	3	

 $H_{\max} = 12$ $K_{\max} = 7$ $L_{\max} = 17$ 表 3 $s.g. P2_1$, $|E| \geq 1.9$, 80 个衍射点

出现次数 指标分量类型	指标	0	1	2	3	4	5	6	7	8	9	10	11
H		4	8	8	8	3	7	11	12	10	6		
K		9	7	6	3	11	14	8	5	5	5	4	3
L		8	9	8	11	9	14	8	7	3			

 $H_{\max} = 9$ $K_{\max} = 11$ $L_{\max} = 8$

首先考虑对于由等同原子组成的结构, 当导出的相角是由单一的 Σ_2 关系式决定时的概率 $p_{h,h'}$ (M. M. Woolfson, 1955)

$$p_{h,h'} = \frac{1}{2} + \frac{1}{2} \tanh \left[\frac{|E_h E_{h'} E_{h-h'}|}{N^{1/2}} \right], \quad (3)$$

	$ E_h E_{h'} E_{h-h'} \approx 20$			$ E_h E_{h'} E_{h-h'} \approx 15$		
N	25	50	100	25	50	100
$P_{h,h',\max}$	0.997	0.996	0.982	0.997	0.986	0.952

计算结果表明, 若选择 $|E| \geq 2.0$ 的值域, 对于上述情况, 由单一 Σ_2 关系式决定的相

角概率是十分接近的,因此,在这种情况下就应该尽可能选择那些包含 Σ_2 关系式多的反射做为起始套反射,以保证运用 Σ_2 关系式迭代后能导出更多的未知相角。

其次再考虑,在相角推导过程中,怎样实现衍射指标的一致分布。

对于给定的晶体衍射指标,若以 $H_{\max}$, $K_{\max}$, $L_{\max}$ 分别表示衍射指标分量在给定范围内的最大值,那么在起始套反射的起始迭代中,若要同时取得 $0-H_{\max}$, $(K_{\max}, L_{\max})$ 的诸分量值,至少需要几个 $H(K, L)$ 值才可以呢?

还要注意到,以 Σ_2 关系式联系的衍射指标对称性为 Laue 对称性,而指标分量则只存在 $\bar{1}$ 对称性(即 $\pm H$, $\pm K$, $\pm L$ 同时存在)。

例如,当 $H_{\max} = 5$, 为取得 $0, 1, 2, 3, 4, 5$ 的指标分量,可以指定 $0, 2, 2, 3$; 或 $0, 1, 1, 4$ 四个值为起始套反射的 H 分量值。又如,当 $H_{\max} = 20$ 时,一种可能的选择是:以 $0, 3, 6, 8, 9, 10, 10$ 七个值做为起始套反射的 H 分量值。对于具体的某个晶体结构而言,存在着由其结构特点而决定的晶体衍射指标、衍射强度的分布规律。因此,如果两个结构,即使具有相同的极大衍射指标分量值,它们衍射指标、强度的分布规律却仍然是不同的,所以,我们不准讨论只按衍射指标分量极大值来确定指标分量值分布的一般规律。不过,对于给定的晶体结构,我们可以采用尝试组方法来获得指标分量的一致分布结果(如上所列)。

上例还表明,指标分量极大值的增加,将导致起始套反射数目的增加。因此,在求解有机小分子晶体结构时,按照在给定的 $|E|$ 值范围内指标 H, K, L 具有一致分布的性质,合理地选择出一定数目的反射,就有可能既满足直接法中对起始套反射选择的一般原则,又能实现在起始迭代后获得一致分布的衍射指标。同样的讨论也适用于 $K_{\max}$, $L_{\max}$ 。

这样的选择结果(即所谓“优化选择”)就能够减少迭代轮数,从而实现在相角推导早期,使每个新导出的相角获得多于一个的 Σ_2 关系式指示,以期获得正确的相角。

优化方法选择起始套反射,包括以下内容:

(1) 按给定结构的不对称单位中的独立原子数目,选择 5—7 (最多 10) 倍的 $|E|$ 值范围;

(2) 计算出相应的 Σ_2 关系式;

(3) 对所得 $|E|$ 值,按衍射指标的奇偶性分为八个奇偶组;

(4) 对每一奇偶组,计算诸反射的 $\{|E_h|N\}$, N 为 h 反射相应的 Σ_2 关系式的数目,并以 $\{|E_h|N\}$ 值大者做为选择起始反射的初选条件;

(5) 在给定的最大值范围内,按指标分量对八个奇偶组中 $\{|E_h|N\}$ 之大者进行尝试迭代,以期获得指标分量的一致分布结果——即优化选择。本步可获得初选的起始套反射;

(6) 由于尚存在指标分量出现频率的差异,在初选的起始套反射中,有时需加入 1—2 个反射,通常的做法是附加由 Σ_1 关系确定的反射,若缺少具有高概率指示的 Σ_1 关系,那么,起始套中尚需包含一个(偶偶偶)型的反射,或者是选择指标分量中有一个(或两个)为零的反射;

(7) 起始套反射之间不能存在 Σ_2 型相容关系。

表 4 给出了一个验证优化选择方法的已知结构例子, $C_{26}H_{28}O_{12}$, $pn21a^{[4]}$.

表 4

	起 始 套 反 射			数 目	指标分量分布情况	结构测定结果
I	6 0 0 12 0 0 0 0 8	6 8 1 0 9 7 1 3 3	2 5 3 3 7 5	8	— — —	成 功
II	同 I	同 I	2 5 3 3 7 5 4 9 4	9	— — —	成 功
III	0 0 4 6 0 0 12 0 0 0 0 8	2 5 3 7 11 8 5 17 7	4 9 4 8 10 2	9	2* — 3 — 13	失 败
IV	同 III	同 III	4 9 4 8 10 2 10 11 5	10	2 — 3 — 13	失 败
V	同 III	1 3 3 6 8 1 2 9 7	2 10 2 4 12 7	9	9 14 —	成 功
VI	同 III	同 V	2 10 2 4 12 7 4 9 4	10	9 14 —	失 败
VII	同 III	6 8 1 1 3 3 2 9 7	4 9 4 8 10 2	9	4 14 — 15 — 16 17	失 败
VIII	同 III	同 VII	4 9 4 8 10 2 10 15 2	10	14 — 16 — 17	失 败
注	由 Σ_i 式确定 的反射	确定原点与对 映体的反射	附加反射		H K L * 为未能组合 的指标	

在所列八种起始套反射的选择方案中, I, II 给出正确的结构测定结果, 其起始套反射指标分量符合优化选择方法. III 中 K 指标未优化, 结构测定失败. IV 虽增加了 (10, 11, 5) 反射, 但其 K 指标与原起始套中 (7, 11, 8) 重复, 未能改变 K 指标优化状况, 故结构测定仍失败. V 表示指标基本优化情况, 结构测定成功. VI 中增加的反射 (4, 9, 4), 其指标分量与 V 起始套的指标分量重复, 虽然未改变指标优化情况, 但这一重复, 却增加了起始步骤中错误信息的权重而导致结构测定失败. VII 的 K 指标未优化, 结构测定失败. VIII 虽然增加了 (10, 15, 5) 反射, 但只部分改善了 K 指标优化状况, 故结构测定仍失败.

四、讨 论

在起始套反射的指标优化选择方法中,是把 $|E|$ 值与 Σ_2 关系式做为一个联合品质因素来考虑的^[8],这样做的目的,是为了保证在相角推导的初始阶段,使导出的相角中的多数是用两个(或更多)的 Σ_2 关系式确定。

在收敛法中,按 $\langle \alpha_h^2 \rangle^{1/2}$ 值筛选的起始套反射,其指标分量的组合值,有时要落在指标分量的封闭集合域之外,这就要减少待推反射出现的机会;而且收敛法在临近终端的结果中,在 $|E|$ 值大小与 Σ_2 关系式多少之间并未进行判断。

起始套反射的优化选择方法是希望给出一个减少失败的可行方法,但并不是说不优化的指标选择就一定不能给出正确的相角推导。

优化方法的使用,有可能提高多解相角套中正确解的品质因素值,而对于记号附加法的使用,有可能获得明确的记号关系,从而扩大它的应用效力。

对于复杂的有机小分子结构,起始套的数目必须很大^[9],从指标优化的观点看,乃是随着结构复杂程度的增加,相应衍射指标值增大的缘故。此外,从指标优化观点看,即使一个分子量不大的小分子,倘若其某一个方向的晶胞参数很大,此时在指定的分辨率范围,如 $1 \text{ \AA}$,仍需选择一个较大数目的起始套反射,以保证相角推导的顺利进行。

指标优化方法,对于低级晶系的相角推导可能是更为有利的。

在应用分量关系式解决 Patterson 法中轻原子的多解问题时^[10,11],指标的优化选择方法提供了一种选择实分量 (A_h) 与虚分量 (B_h) 起始符号的可行方法。

范海福、裘宗沪同志给予作者许多有益的讨论,谨致谢意。

参 考 文 献

- [1] J. Karle, I. L. Karle, *Acta Cryst.*, **21** (1966), 849.
- [2] G. Germain, M. M. Woolfson, *Acta Cryst.*, **B24** (1968), 91.
- [3] G. Germain, M. M. Woolfson, *Acta Cryst.*, **B26** (1970), 274.
- [4] L. Lessinger, *Acta Cryst.*, **A32** (1976), 538.
- [5] S. E. Hall, *Acta Cryst.*, **A34** (1978), 38.
- [6] P. H. Stothart, *Acta Cryst.*, **A34** (1978), 421.
- [7] А. И. Князьковский «Теория структурного анализа», Москва (1957).
- [8] Olga Kennard *et al.*, *Proc. R. Soc. Lond.*, **A325**, (1971), 401;
范海福等,物理学报, **27**(1978), 554.
- [9] M. M. Woolfson, *Acta Cryst.*, **A33** (1977), 219.
- [10] 范海福、郑启泰,物理学报, **27**(1978), 169.
- [11] 郑启泰,物理学报,待发表.

AN OPTIMAL SELECTION METHOD ON STARTING REFLEXION SET IN DIRECT METHOD

ZHENG QI-TAI

(Institute of Biophysics, Academia Sinica)

ABSTRACT

In the direct method of crystal structure determination, one usually applies the symbolic addition method and multiple tangent-formula method. For determination of the phase, the selection of starting reflexion set is a key problem for successful analysis of the crystal structures. The convergence method has been applied for determination of starting reflexion set. This paper presents an optimal selection method on starting reflexion set on the basis of the relations between the diffraction indexes and gives an example of examining a known structure.