

# 起始相位的受限排列

范海福 韩福森

(中国科学院物理研究所)

1980年10月30日收到

## 提 要

在含重原子晶体结构的测定中,重原子贡献的相位是一个非常重要的信息.利用重原子相位作电子密度逼近或者 Karle 迭代是求解含重原子晶体结构的重要手段.但是,当重原子的贡献不占主导地位,或者当重原子分布的对称性高于整个结构的对称性时,上述方法就不易奏效.这时的重原子信息是否还可以用来大大简化结构分析的过程?这是一个值得探讨的问题.本文提出利用重原子的相位来限制直接法中起始相位的排列方式.这样可以成倍地缩减起始相位排列的种数.这一方法同作者提出的另一方法相配合可以使直接法在测定含重原子晶体结构方面的效能大为提高.

## 一、引 言

重原子法是测定含重原子晶体结构的主要方法.其要点就是用重原子贡献的相位代替结构因子的相位来近似地计算电子密度图,然后从中寻找有关轻原子的信息.这样就要求重原子贡献的相位接近于结构因子的相位.当重原子的贡献不占主导地位时,上述条件就不易满足. Woolfson<sup>[1]</sup> 以及 Sim<sup>[2]</sup> 提出在计算电子密度图时使用权函数以抑制由于重原子相位偏离结构因子实际的相位所产生的误差.但是加权的办法并没有改善相位本身,因此它的作用有一定局限.另外,当重原子分布的对称性高于整个结构的对称性时,用重原子相位代替结构因子相位有时会引进赝对称性从而使轻原子产生多解问题.遇此情况,使用权函数是无济于事的.为了在不能简单地用重原子相位代替结构因子相位的情况下,还可以利用重原子信息来简化结构分析的过程,本文提出了一种简便而有效的方法,将重原子信息同 MULTAN 法<sup>[3]</sup>结合起来使之得到更充分的利用.

## 二、原 理

MULTAN 法的主要特点之一在于对每一个未知相位的起始衍射点给定若干个可能的相位,然后分别考察由每一种相位排列推引出来的结果,择其优者以之计算  $E$  图. 一般说来,起始衍射点的数目越多,相位推引和修正的过程也越顺利和可靠.但是起始衍射点数的增加会使相位排列的套数按指数关系上升,所需计算时间将随之激增.如果能设法利用重原子信息对起始相位的可能范围加以限制,就有可能使相位排列的套数成倍地缩减.这样就可以大大节省计算时间,从而使 MULTAN 法有可能处理更为复杂的晶体

结构。

Sim<sup>[2]</sup> 曾对重原子相位同结构因子相位之间的偏差作过概率估算, 他得到如下表达式:

$$P(\xi) = \int_{-\xi}^{\xi} \exp(x \cos \xi) d\xi / 2\pi I_0(x), \quad (1)$$

式中  $P(\xi)$  是结构因子相位同重原子相位之差落在  $\pm \xi$  范围内的概率;  $I_0(x)$  是零阶的第一类变型贝塞耳函数, 即

$$I_0(x) = \frac{1}{\pi} \int_0^{\pi} \exp(x \cos \xi) d\xi;$$

$$x = 2|F||F_H| / \sum_1^{N_L} f_L$$

这里  $F$  是结构因子,  $F_H$  是重原子对结构因子的贡献,  $f_L$  是轻原子的散射因子,  $N_L$  是晶胞中轻原子的数目。

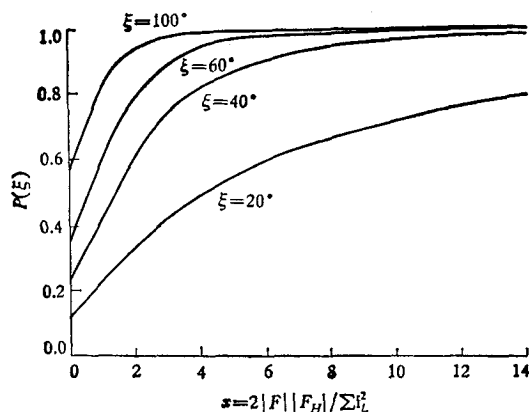


图 1

图 1 示出  $\xi$  取各种不同数值时  $P(\xi)$  对  $x$  的关系。由图 1 可知, 要使结构因子的相位近似地等于重原子贡献的相位, 就需要有相当大 (接近于 10) 的  $x$ 。另一方面, 只要  $x \geq 3$ , 结构因子相位同重原子相位之差超过  $100^\circ$  的可能性就只有 2% 左右。因此, 一个由 MULTAN 程序选出的起始衍射点, 如果相应的  $x \geq 3$ , 其相位的可能变化范围就不应定为  $360^\circ$ , 而应该限制在重原子相位  $\pm 90^\circ$  左右的范围内。由此可见, 用通常

的重原子法, 要想由重原子相位计算出一张可以辨认出轻原子的电子密度图就必须有相当数量的、 $x$  接近于 10 的衍射点。但是要想在 MULTAN 法中利用重原子信息来限制起始相位的变化范围, 则只要有几个  $x \geq 3$  的衍射点就够了。此外, 如果重原子的对称性较之整个晶胞的对称性, 具有更短的平移周期或者多一个对称中心, 这时用重原子相位代替结构因子相位, 不管  $x$  值有多大都会引进赝对称性, 给结构分析带来困难。但如果只用重原子相位来限制起始相位的排列方式则可避免这种情况。

### 三、实 例

#### 1. 溴化五味子醇乙

化学式:  $C_{23}O_7H_{27}Br$

空间群:  $P2_12_12_1$

晶胞:  $a = 19.25 \text{ \AA}$ ,  $b = 15.99 \text{ \AA}$ ,  $c = 7.28 \text{ \AA}$ ;  $z = 4$ 。

用 MULTAN 78 程序自动求解时, 选出 3 个相位不确定的起始衍射点, 共得 16 种可

表 1

衍射指标	18, 4, 0	13, 3, 1	3, 3, 1
$x$	7.87	11.50	10.14
重原子相位	$0^\circ$	$270^\circ$	$90^\circ$

表 2

相位组合编号	$\varphi_{18, 4, 0}$	$\varphi_{13, 3, 1}$	$\varphi_{3, 3, 1}$	综合品质因子
1	0	45	45	1.165
2	0	135	45	1.058
*3	0	225	45	2.001
*4	0	315	45	= 3 号
5	180	45	45	0.829
6	180	135	45	0.718
7	180	225	45	0.635
8	180	315	45	0.635
9	0	45	315	= 3 号
10	0	135	315	= 3 号
11	0	225	315	= 3 号
12	0	315	315	1.005
13	180	45	315	0.880
14	180	135	315	0.525
15	180	225	315	1.008
16	180	315	315	1.178

能的相位排列。表 1 示出这三个点的指标、 $x$  值及重原子贡献的相位。表 2 示出 16 种相位排列及其综合品质因子。由于这三个衍射点的  $x$  值都比 3 大得多, 因此完全有理由将它们的相位限制在重原子相位  $\pm 90^\circ$  范围之内。如此限制之后, 可能的相位排列就只剩下表 2 中第 3 和第 4 两套。这两套恰巧也是综合品质因子最佳的两套。

## 2. 噻吩甲酰三氟丙酮-镧-胺基吡啶

化学式:  $C_{37}O_8N_2F_{12}S_4H_{26}Pr$

空间群:  $Pca2_1$

晶胞:  $a = 19.06 \text{ \AA}$ ,  $b = 11.86 \text{ \AA}$ ,  $c = 19.53 \text{ \AA}$ ;  $z = 4$ 。

用 MULTAN 78 程序自动求解时, 选出了四个相位不确定的起始衍射点(表 3), 共得 40 套可能的相位排列。利用重原子限制之后, 只剩下第 10, 11, 30, 31 四套, 其中包含了综合品质因子最佳的第 30 套(见表 4)。

表 3

衍射指标	8, 3, 13	1, 2, 4	8, 6, 13	1, 1, 2
$x$	15.59	14.56	12.04	9.39
重原子相位	$180^\circ$	$0^\circ$	$0^\circ$	$180^\circ$

表 4

组合编号	$\varphi_{8,3,13}$	$\varphi_{1,2,4}$	$\varphi_{3,6,13}$	$\varphi_{1,1,2}$	综合品质因子
1	45	45	36	27	1.8424
2	135	45	108	81	1.8815
3	225	45	180	135	1.4363
4	315	45	252	189	1.8553
5	45	45	324	243	= 1号
6	135	45	36	297	= 3号
7	225	45	108	351	= 4号
8	315	45	180	45	= 1号
9	45	45	252	99	1.7513
*10	135	45	324	153	= 3号
*11	225	45	36	207	= 1号
12	315	45	108	261	= 3号
13	45	45	180	315	= 4号
14	135	45	252	9	1.6659
15	225	45	324	63	= 3号
16	315	45	36	117	= 3号
17	45	45	108	171	= 1号
18	135	45	180	225	= 9号
19	225	45	252	279	= 3号
20	315	45	324	333	1.2878
21	45	315	36	27	= 9号
22	135	315	108	81	= 1号
23	225	315	180	135	= 2号
24	315	315	252	189	= 3号
25	45	315	324	243	1.6289
26	135	315	36	297	= 4号
27	225	315	108	351	= 3号
28	315	315	180	45	= 14号
29	45	315	252	99	= 14号
**30	135	315	324	153	= 2号
*31	225	315	36	207	= 1号
32	315	315	108	261	= 9号
33	45	315	180	315	= 3号
34	135	315	252	9	= 4号
35	225	315	324	63	= 1号
36	315	315	36	117	= 2号
37	45	315	108	171	= 2号
38	135	315	180	225	= 1号
39	225	315	252	279	= 4号
40	315	315	324	333	= 3号

上述两个例子表明,用重原子相位限制起始衍射点的相位变化范围,可以成十倍地缩减相位排列的套数。实际应用时可以按照下面的步骤进行:

- (1) 用 Patterson 法求出重原子的坐标。这个步骤可以由电子计算机自动地执行<sup>[4]</sup>。
- (2) 用 MULTAN 法选出起始衍射点,注意使规定原点及对映体的衍射点的相位不得同重原子坐标的要求相矛盾。

(3) 按 MULTAN 法求出起始衍射点的各种相位排列. 对于  $x \geq 3$  的起始衍射点, 凡遇其起始相位超出重原子相位  $\pm 90^\circ$  的范围时, 将相应的相位排列弃置不用.

(4) 用带重原子校正项的正切式<sup>[5]</sup>作相位的扩展和修正. 遇晶体中所含重原子较重时, 尤其是当重原子分布的对称性高于整个晶体的对称性时, 重原子的校正项是不能忽略的.

#### 四、讨 论

要利用重原子相位来限制起始相位排列的方式, 其必要的条件就是在起始套中必须含有若干个  $x \geq 3$  的衍射. 因此上述方法的适用范围有多大将取决于重原子含量可以低到何种程度而仍能保证有足够数量的  $x \geq 3$  的衍射点.

按文献 [2] 的定义

$$x = 2|F||F_H| / \sum_1^{N_L} f_L.$$

将式中的结构因子  $F$  换成归一化结构因子  $E$ , 可得

$$\begin{aligned} x &= \left( 2 \sqrt{\sum_1^N f^2} \sqrt{\sum_1^{N_H} f_H^2} / \sum_1^{N_L} f_L \right) |E||E_H| \\ &= \left( 2 \frac{\sqrt{\sum_1^{N_H} f_H^2}}{\sqrt{\sum_1^N f^2}} / \frac{\sum_1^{N_L} f_L}{\sum_1^N f} \right) |E||E_H|, \end{aligned}$$

其中  $N = N_H + N_L$  是晶胞中原子的总数. 令

$$f_i \sim \hat{f}Z_i,$$

其中  $Z_i$  是第  $i$  种原子的原子序;  $\hat{f}$  是平均的单位原子散射因子. 再令

$$\sigma_1 = \sqrt{\sum_1^{N_H} Z_H^2} / \sqrt{\sum_1^N Z^2},$$

$$\sigma_2 = \sqrt{\sum_1^{N_L} Z_L^2} / \sqrt{\sum_1^N Z^2}.$$

则

$$x = \frac{2\sigma_1}{\sigma_2^2} |E||E_H| = \frac{2\sigma_1}{1 - \sigma_1^2} |E||E_H|.$$

在全衍射点中  $x \geq x_{\min}$  的概率为

$$\int_{x_{\min}}^{\infty} P(x) dx = \int_0^{\infty} \left[ \int_{x_{\min}}^{\infty} \sigma_2^{2/\sigma_1|E|} P(|E|; |E_H|) P(|E_H|) d|E_H| \right] d|E|,$$

其中  $P(|E_H|)$  是重原子归一化结构因子绝对值的概率分布;  $P(|E|; |E_H|)$  是当重原子

贡献的绝对值为  $|E_H|$  时, 整个晶胞的归一化结构因子绝对值  $|E|$  的条件概率分布. 根据文献[6],

$$P(|E|; |E_H|) = \frac{2|E|}{\sigma_2^2} \exp\left[-\left(\frac{|E|^2 + \sigma_1^2|E_H|^2}{\sigma_2^2}\right)\right] I_0\left[\frac{2\sigma_1|E||E_H|}{\sigma_2^2}\right].$$

通常使用直接法时, 只对  $|E| \geq |E|_{\min}$  的衍射点感兴趣. 在全体衍射点中, 同时满足

$$|E| \geq |E|_{\min} \text{ 和 } x \geq x_{\min}$$

的概率为

$$P_2 = \int_{|E|_{\min}}^{\infty} \left[ \int_{\sigma_2^2 x_{\min} / 2\sigma_1 |E|}^{\infty} P(|E|; |E_H|) P(|E_H|) d|E_H| \right] d|E|.$$

注意到在全体衍射点中  $|E| \geq |E|_{\min}$  的概率为

$$P_1 = \int_{|E|_{\min}}^{\infty} \left[ \int_0^{\infty} P(|E|; |E_H|) P(|E_H|) d|E_H| \right] d|E|,$$

可得在  $|E| \geq |E|_{\min}$  的衍射点中  $x \geq x_{\min}$  的衍射点所占的百分比应为

$$P = P_2/P_1.$$

当空间群为  $P_1$  时, 若晶胞中只含一个重原子, 则

$$P(|E_H|) = \delta(|E_H| - 1)^{[6]};$$

若晶胞中含有两个重原子, 则

$$P(|E_H|) = \frac{2}{\pi\sqrt{2 - |E_H|^2}}^{[6]},$$

图 2 示出在上述两种情况下, 当  $x_{\min} = 3$  以及  $|E|_{\min} = 1.5$  时  $P$  同  $\sigma_1^2$  之间的关系.

对于约含 100 个独立原子的晶体结构, 通常约可有四、五百个  $|E| \geq 1.5$  的衍射点. 如果  $x \geq 3$  的衍射在其中占 3%, 其总数就可有十几个. 一般说来, 这就有足够的机会使用本文提出的方法.

从图 2 可知, 要满足这样的条件, 对于晶胞中只含一个重原子的情况, 要求  $\sigma_1^2 \geq .25$ ; 对于晶胞中含有两个重原子的情况, 则只要求  $\sigma_1^2 \geq .15$ . 后者只相当于在每 700 个氮原子中加入一个汞原子. 由此可见, 本文提出的方法基本上适用于通常可能遇到的各种含重原子的晶体, 甚至有可能适用于含重原子的较小的生物大分子晶体.

作者对英国约克大学 M. M. Woolfson 教授有益的讨论特表谢意.

### 参 考 文 献

- [1] M. M. Woolfson, *Acta Cryst.*, **9** (1956), 804.
- [2] G. A. Sim, *Acta Cryst.*, **12** (1959), 813.
- [3] G. Germain, P. Main and M. M. Woolfson, *Acta Cryst.*, **A**, **27** (1971), 368.
- [4] 韩福森、范海福, 物理学报, **30** (1981), 224.
- [5] 韩福森、范海福、古元新, 物理学报, **30** (1981), 111.
- [6] R. Srinivasan and S. Parthasarathy, "Some Statistical Applications in X-Ray Crystallography" Pergamon Press Oxford, (1976), Chapter 3.

## RESTRAINED PERMUTATION OF THE STARTING PHASES

FAN HAI-FU HAN FU-SEN

*(Institute of Physics, Academia Sinica)*

### ABSTRACT

The heavy atom phases are extremely important in solving crystal structures containing heavy atoms. Successive Fourier approximation and Karle recycling procedures starting with the heavy atom phases are two major methods of solving structures containing heavy atoms. However, difficulties may some times happen when the contribution of heavy atoms is too small to dominate sufficient number of phases or if the distribution of heavy atoms possesses a higher symmetry than that of the whole structure. It is worthy to investigate in these cases whether or not the heavy atom information can still be used to simplify the process of structure determination greatly. In this paper a method is proposed that the heavy atom phases may be used to restrict the permutation of the starting phases leading to a great reduction in number of the starting phase sets. This method incorporating with that proposed by the authors can improve greatly the efficiency of direct methods in solving structures containing heavy atoms.