

蛋白质晶体对 X 射线的吸收效应与衍射数据质量

王耀萍 马星奇 王家槐

(中国科学院生物物理研究所, 北京)

摘 要

本文以实例讨论了在用四圆衍射仪收集蛋白质晶体的衍射数据时, 由于吸收效应, 特别是母液的吸收效应的校正不好, 可以显著影响数据质量, 影响结构的测定。

用 X 射线晶体学方法测定蛋白质分子的三维结构, 其精度取决于收集到的衍射数据的质量。如果有了好晶体, 在用四圆衍射仪收集数据时, 吸收效应往往是决定数据质量最主要的因子。因为蛋白质晶体的外形大多很不匀称, 有些薄片状晶体的宽度和厚度可差一个数量级, 收集数据时, 入射线光源需取不同的方位照射晶体, 这就导致吸收效应的差别。这种差别对于晶体的重原子衍生物尤其严重, 往往相差几倍乃至十几倍。所以仔细地校正吸收效应具有决定性意义。

吸收效应的校正通常采用半经验方法^[1,2]。其中以 North 等人的方法^[1]应用更为广泛, 其要点如下: 把晶体的一个结晶学基轴(譬如 C^* 轴)安装在沿衍射仪测角头 ϕ 轴上。图 1 表示 ool 衍射处于反射位时的几何, 大圆为衍射仪上的赤道面。因此时 ϕ 轴与 C^* 轴一样正好躺在

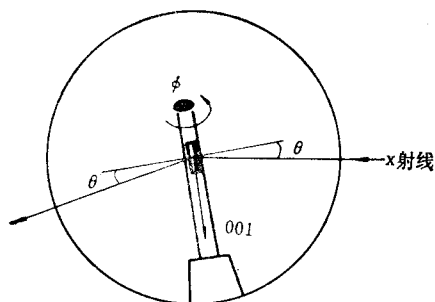


图 1 半经验方法校正吸收示意图

赤道面上, 所以无论 ϕ 轴怎样旋转, ool 衍射永远处于反射位。显然, 当 ϕ 轴旋转时晶体以不同的方位受到 X 射线入射, 因此有不同的吸收效应。测量 ϕ 角在 $0-360^\circ$ 范围内 ool 衍射的不同强度将绘出一条描绘该晶体吸收效应的吸收曲线, 用这条曲线可以校正每个衍射点的吸收效应。

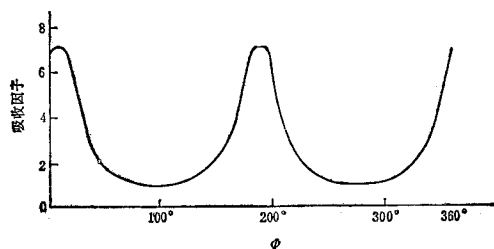


图 2 正交晶型天花粉蛋白碘汞衍生物吸收曲线

图 2 是我们在研究正交晶型天花粉蛋白晶体的碘汞衍生物(以下简称 THg 晶体)时作的吸收曲线。一般情况下, 吸收曲线近于正弦曲线。仔细地校正吸收以后, 可观察衍射等效点的平均偏离因子应当在 5% 左右。

但是, 有许多因素可以严重影响吸收效应的校正。North 等人指出, 晶体外形沿安装方向应当有二重对称性。否则, 吸收曲线的对称性将受影响, 不容易反映每个衍射点的吸收效应。我们的经验表明, 如果照射晶体的人射 X 射线束的强度分布不均匀, 或者晶体在衍射仪

上位置有点偏心,吸收曲线的两个峰间的距离将不正好等于 180° ,且峰高也将明显不相等,所有这一切都将显著影响数据的质量。还有一个极为重要的因素是母液。已知蛋白质晶体必须安装在密闭的薄壁毛细管内,晶体附贴在管壁的一面常常有残留的母液。North 指出^[1],母液会干扰吸收效应的校正,以下我们将详细地讨论这种干扰对于我们收集的 THg 晶体数据质量的影响。

THg 晶体空间群为 $P2_12_12_1$, $a = 38.2 \text{ \AA}$, $B = 76.2 \text{ \AA}$, $c = 79.1 \text{ \AA}$ 。晶体外形通常呈片状,沿 a 方向的线度一般比其余两个方向小 2—3 倍。图 1 为晶体沿 C 方向安装在毛细管内,

而且晶体附贴在管壁的那一面残留有不少母液。由于 THg 晶体的母液内含有 K_2HgI_4 ,所以它对 X 射线的吸收效应不小。作吸收曲线时,在任何方位角上,入射线或衍射线将穿过母液,总的效果都是穿过一次,不影响吸收曲线的对称性。我们获得了一条很漂亮的吸收曲线(图 2)。但是在数据收集过程中发现,很多衍射点的等效点对 hkl 与 $\bar{h}\bar{k}l$ 相差很大,最后全套 $4 \text{ \AA}$ 数据 2152 个点的等效点偏离因子 $R_{sym} = 12.3\%$ 。事实上,校吸收以后等效点之间的偏差不是减小了,反而增大了。表 1 为随机抽出一组衍射点在校吸收前后的等效点对偏离因子。

表 1 校吸收前后两等效点强度偏离因子

指 标	校吸收前(%)	校吸收后(%)	指 标	校吸收前(%)	校吸收后(%)
6 0 0	12	12.3	5 1 3	8.4	9.8
7 1 0	12.8	12.8	2 9 3	2.3	4.0
7 1 1	12.8	13.1	1 10 3	5.0	5.6
5 2 1	17.5	17.8	1 12 3	3.9	4.4
6 2 1	17.0	17.2	1 9 4	4.3	5.3
0 9 1	4.7	4.4	3 13 4	3.3	5.2
0 11 1	4.9	4.6	6 0 5	9.0	12.0
6 1 2	17.6	18.4	5 1 5	2.8	5.2
2 10 2	3.4	4.5	5 2 5	17.4	19.0
2 12 2	5.3	5.8	7 3 5	28.9	30.0
0 15 2	3.8	3.6	0 11 6	2.3	3.0
5 0 3	4.9	7.2	5 0 7	5.5	10.5
1 10 7	0.8	1.2	6 0 8	5.6	10.3
1 12 7	3.0	4.7	1 11 8	2.7	4.9

这表明,尽管吸收曲线很漂亮,但吸收效应的校正却很糟糕。我们将所收集的衍射点分区作统计。凡衍射指标 $4|H| < |K|$ 者划为一区,衍射指标 $|H| > |K|$ 为一区,中间的衍射点 $|H| < |K| < 4|H|$ 为另一区。然后按 L 分层分别统计三个区内等效点的偏离因子。表 2 为统计结果。

统计结果清楚地表明,衍射指标为 $4|H| <$

$|K|$ 区内的数据质量最好, $|H| < |K| < 4|H|$ 区次之,最差的是 $|H| > |K|$ 区。当 L 指标变得很大时,三个区的数据质量都明显下降。这是由于那些点处于高布拉格角区,弱点增多,误差加大,另一原因是由于吸收曲线是用 006 衍射点作的,不能充分反映高层衍射点的吸收状况。至于三个区数据质量的差别则纯粹是母液吸收效应造成的。图 3 显示 HKO 倒易面上三个区

表 2 等效点平均偏离因子的分区统计

L	R($4 H < K $) 点数		R($ H < K < 4 H $) 点数		R($ H > K $) 点数	
0	0.04538	47	0.09258	62	0.14793	40
1	0.04340	94	0.12854	123	0.16178	80
2	0.04665	94	0.10880	121	0.14910	84
3	0.03379	90	0.10583	124	0.14529	82
4	0.03424	92	0.09104	123	0.15499	84
5	0.03667	94	0.12038	124	0.15489	82
6	0.04593	92	0.12469	117	0.22245	78
7	0.05185	86	0.11032	114	0.17557	72
8	0.05913	86	0.11696	108	0.18935	72
9	0.04673	80	0.11447	104	0.16850	68
10	0.07283	78	0.12197	98	0.17841	64
11	0.06687	70	0.12020	90	0.20285	56
12	0.10263	68	0.13136	81	0.24077	54
13	0.05266	60	0.10972	76	0.15446	50
14	0.05905	52	0.11013	68	0.16864	42
15	0.06366	46	0.14039	56	0.23723	38
16	0.11878	40	0.17481	46	0.20283	30
17	0.11237	28	0.11627	36	0.23998	20
18	0.16172	20	0.15455	24	0.13625	12
19	0.13586	10	0.09453	12	0.27726	6

与晶体外形的关系。

图 4 则给出 $|H| > |K|$ 区及 $4|H| < |K|$ 区中衍射点处于反射位时的几何。图中晶体边上的黑阴区代表母液。图 4(a) 左边是 hkl ，右边是等效点 $\bar{h}\bar{k}l$ 。显然，对于 hkl ，入射线与衍射线均不通过母液；对于 $\bar{h}\bar{k}l$ ，入射线与衍射线都通过母液，一共通过两次。前面已提到，作吸收曲线时，入射线与衍射线通过母液，一共一次。就是这种 North 所讨论过的不对称性，使

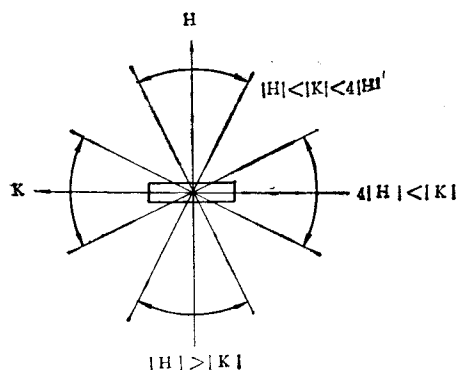


图 3 HKO 倒易面上三个区域与晶体外形的相对关系

hkl 的吸收校正过了头，而 $\bar{h}\bar{k}l$ 则校正不足，导致严重的误差。而图 4(b) 区内， $\bar{h}\bar{k}l$ 也好，其等效点 $\bar{h}\bar{k}l$ 也好，都是入射线与衍射线通过母液，一共一次，与作吸收曲线时的情况一样，所以不同等效点都可以获得很好的吸收校正，等效点偏离因子非常理想。至于 $|H| < |K| < 4|H|$ 区的情况则是上两区之间的过渡。

我们用这套数据计算差值 Patterson 图，试图借以确定重原子位置的努力失败了。后来我们又重新收了一套数据，即将 a^* 轴顺 ϕ 轴安装，由于晶体在另两个方向上线度差不多，所以吸收效应要小一些。在装晶体时，我们尽量使残留母液减少，所得数据质量极佳， $5\text{ \AA}$ 1125 个独立点的 $R_{\text{sym}} = 2.9\%$ 。用这第二套数据确定了重原子位置，两套重原子位置的 Harker 峰和非 Harker 峰都是 Patterson 图上最突出的(待发表)。当我们用这个已知座标参数再回过头去分析第一套数据计算得到的 Patterson 图时，可以看出，有些峰被淹没在背景中。这充分反映了第一套数据误差造成的后果。表 3 是将重原子座标分别用两套数据(都取 $5\text{ \AA}$ 分辨率)去

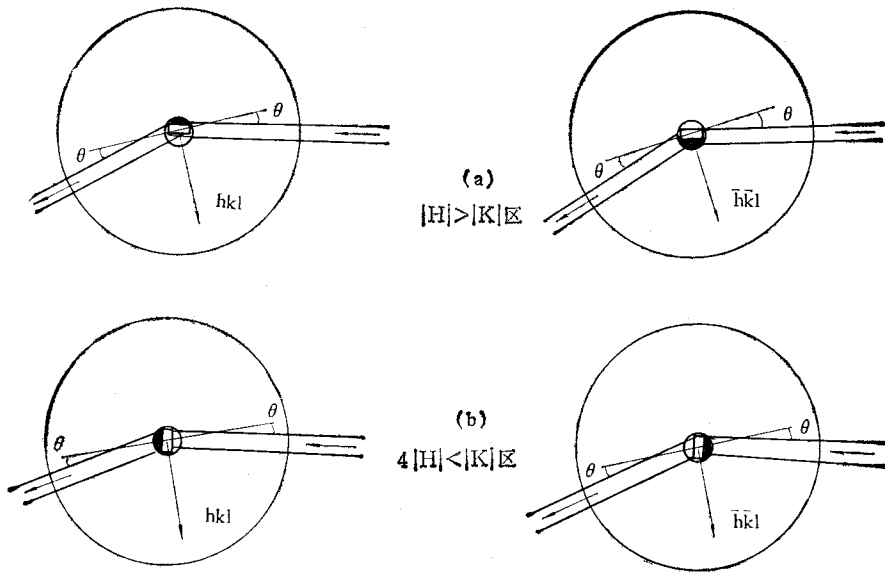


图 4 4(a) $|H| > |K|$ 区两个等效点 hkl 与 $\bar{h}\bar{k}l$ 的衍射几何。
4(b) 为 $4|H| < |K|$ 区等效点衍射几何。

表 3 两套数据的比较

第一套数据	第二套数据
R_{sym} 12.3%(4 Å)	2.9%(5 Å)
R_{FHLB} 0.428(5 Å)	0.406(5 Å)
m 0.613(5 Å)	0.713(5 Å)

注: R_{FHLB} 是重原子参数修正后的偏离因子, m 是反映电子密度图质量的品质因子。

这些参数进一步定量地表明两套质量不一样的数据用于蛋白质结构测定时,其效果不同。而归根到底,是由于吸收校正误差大小的不同,特别是对母液吸收效应处理的好坏造成的。

修正和计算相角时获得的参数。

参 考 文 献

- [1] North, A. C. T. et al.: *Acta Cryst.*, A24, 351, 1968.
- [2] Huber, R. et al.: *Acta Cryst.*, A 25, 143, 1969.

[本文于 1986 年 5 月 2 日收到]



学术活动

第二次全国分子免疫学学术讨论会在南京举行

中国生物化学学会于 1986 年 10 月 6 日至 10 日在南京医学院召开了第二次全国分子免疫学研究进展学术讨论会。会议交流的主要内容有抗体、抗原、补体、淋巴因子、免疫受体、免疫遗传、基因结构、免疫调节网等。出席会议的代表 95 名,收到 102 篇论文,会上宣读了 82 篇。综述报告 7 篇:荣康泰的《人工抗原与化学免疫》、陈诗书的《Ig、MHC、TCR 等多基因族的基本结构》、徐愤的《补体的活化与调节》、侯健存的《粘液上皮细胞 F_{cr} 受体 IgG 运输与刺激粘液分泌亢进》、沈翔珩的《T 淋巴细胞激活和增殖的调控》、赵武述的《白细胞介素研究新进展》、郑昌学的《参加第六次国际免

疫学会议感想和内容介绍》。

参加会议的人数、论文数比 1983 年第一次全国分子免疫学讨论会多近一倍。宣读论文者除了老一辈科学家外,还有许多青年学者,以及正在攻读的博士及硕士研究生。显示出这一新学科的兴旺发达景象。

会议期间,部分代表参观了南京医学院徐愤教授的分子免疫学研究室,并座谈了补体研究的现况和展望,与会者认为为尽快发展我国分子免疫学,在国内也应该互派学者进行交流、取长补短以加速科研进程。

[北京友谊医院、北京市临床医学研究所生化研究室

袁振铎]