

# 生物大分子结构的X射线衍射分析

## (一) X射线衍射分析原理

王 大 成      梁 栋 材

(中国科学院生物物理研究所)

核酸、蛋白质(包括酶)等生物大分子是构成生命的主要物质基础。研究它们的结构与功能,构成了当前分子生物学的中心内容,也是现阶段认识生命现象的中心环节。与非生命物质比较,生物大分子在结构上的一个重要特点,就是其三维结构对其功能的极端重要性。所谓三维结构,是指分子在三度空间中的立体结构,亦即组成分子的所有原子的特征空间排布,也常称为空间结构。越来越多的研究事实使人们认识到,没有特征的三维结构就没有复杂的蛋白质和核酸的功能。

测定生物大分子的三维结构相当困难。这是因为,它们中的大多数原子间的距离,都在 $1.5\text{ \AA}$  ( $1\text{ \AA} = \text{亿分之一厘米}$ )左右;要测定这样精细的结构,必须拥有可以达到这样高分辨率的方法和技术。迄今为止,X射线衍射分析是较好地满足这一要求的唯一技术,因此成为目前测定生物大分子三维结构的主要方法。这一技术利用X射线穿过晶体发生的衍射现象与晶胞内原子的空间位置有严格的关系,以数学物理方法从晶体的衍射像推引出分子的真实像,分辨率可以达到原子水平,且成像是三维实体的。由于应用这一技术,要求被研究材料必须是晶体,所以也常称其为晶体结构(或晶体衍射)分析。这一技术的成功应用,已使上百个蛋白质和少数核酸的结构在原子水平测定,一些病毒和脂蛋白体系在低分辨率得到研究,对阐明一些蛋白质和酶的作用机理起了重要作用(关于这方面的详细介绍见本刊1980年第5、6期)。本文首先介绍它的一般原理,再介绍结构测定的实际过程,以及我国科学工作者用以研

究胰岛素晶体结构所取得的结果。

### 衍射现象与物质结构

光学衍射产生于光的波动性。可能你看见过,在碧波荡漾的湖面上放一块中心有一小孔的木板,水波穿过小孔后变成了以小孔为中心的圆形波纹,这表明穿过小孔后水波的传播并不严格限制在孔径规定的直线范围内,而是漫射到了被遮挡的其它部位。这就是衍射现象。事实上,任何光波当作用到与其波长大小差不多的物体上时,都会产生衍射。这是因为,光是一种电磁辐射,当它作用于某种物质时,它的周期递变的电场会迫使组成物质的带电粒子(电子、原子核)跟着振动(图1a),这些粒子作为新的辐射源发出次生射线(图1b),似乎入射光在许多方向弥散开来,这就是散射。这种散射波

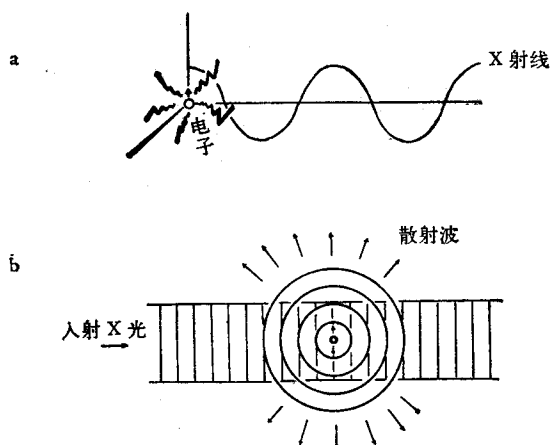


图1 X光与原子外层电子的作用

(a) 电子受X射线作用而振动; (b) 入射线的平面波和电子散射的球面波。

的主要部分与入射光有相同的波长，当入射波长与散射体大小相当时，它们彼此发生叠加(干涉)，于是在入射方向以外的一些方向加强(亮)，在另一些方向彼此减弱(暗)(图2)。如用一张感光底片记录下来，就会得到明暗相间，强弱有致且具有特征几何分布的花样，这就是常称的衍射花样 (diffraction pattern) (图4)。

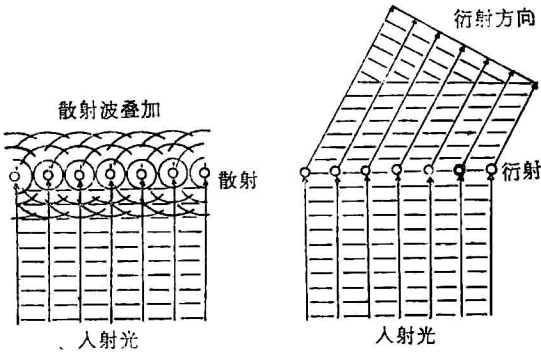


图 2 散射波的叠加——衍射原理示意

任何对象的衍射花样完全地决定于它自身的结构。图3是可见光通过一透明小孔产生的衍射，它是明暗相间的圆环；图4a是一个平行四边形格子结构产生的衍射花样。如果有一个近似六角形的平面分子，它的衍射花样如图4b所示。显然，有什么样的结构，就对应地产生一种特征的衍射强弱分布。对可见光来说，

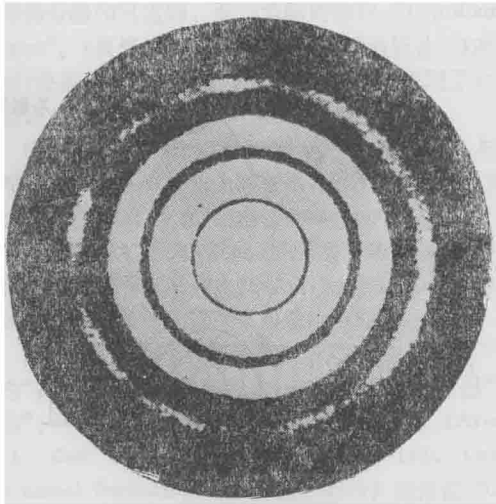


图 3 一个小孔的衍射

(选自 Blundell, T. L. and Johnson, L. N.: *Protein Crystallography*, 1979.)

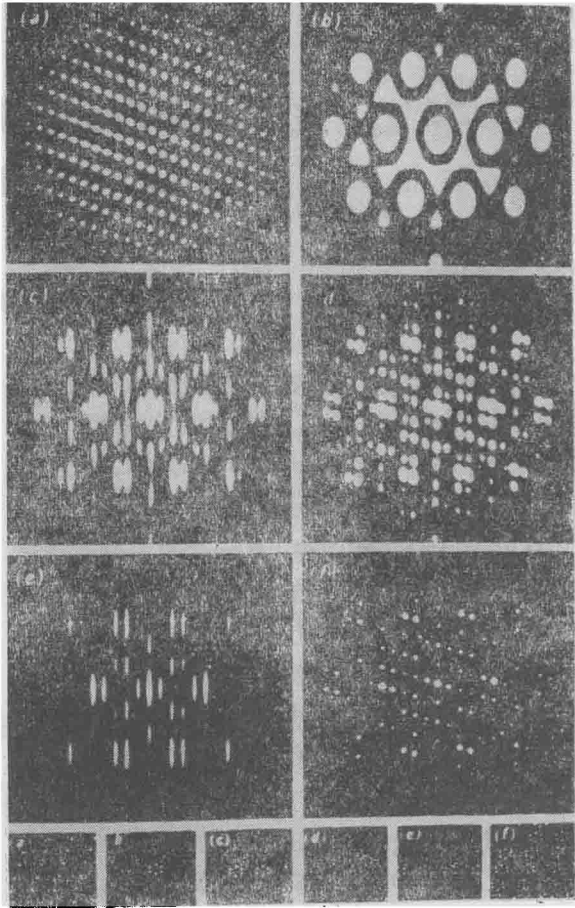


图 4 不同结构所产生的光学衍射花样

(选自 Taylor, C. A. and Lipson, H.: *Optical Transforms*, 1964.)

- (a) 一个格子，(b) 一个单分子，(c) 二个分子，
- (d) 四个分子，(e) 一系列分子，(f) 一个小晶体。

由于它与物质的作用较强，所以有可能观察和记录到单个分子的衍射，称为分子的傅里叶变换或分子变换 (molecular transform)。图4就是结构示于底部的分子傅里叶变换的光学模拟。所以，光学衍射分析的研究材料，并不一定要是晶态。从图还可看出，当有多个相同分子规则重复时，衍射花样的强度彼此加强，分布逐渐从连续变化(a)过渡为离散点(f)。对于X光，由于它与物质相互作用很弱，所以不可能观察到单个分子的衍射变换。但如果大量同一分子规则排列起来(如图4f)，则任何一个分子的散射都会被所有其它分子的散射所加强，从而得到可观察和记录的衍射花样。晶态物质正

好满足这一要求。在一个单晶体中,数目极大的(10的十数次方)分子构成晶胞,晶胞以极为规则的方式在三维方向重复排列,从它才可以得到足够分析用的X射线衍射效应。所以,用于X射线衍射分析的研究材料,必须是晶体。晶体的大小(即所含分子数的多少),内部排列的规则程度(有序程度),乃至晶体的外形,都直接影响其衍射花样的质量。

### X 射线衍射与晶体结构

一个晶体在结构上与非晶态物质的基本不同点,就在于它具有三维的周期重复性。为了便于理解,下面的讨论将用二维的情况作代表。将许多小的相同平行六面体沿 X、Y、Z 三个方向规则地排列起来,就构成了一个典型的晶格结构(图 5),如在每一晶格放上一个或数个同一分子或分子的一部分,就构成一个实际的晶体结构;这些晶格在二维平面上,就是一个由平行四边形组成的网格(见图 6c)。显然,任何晶体的衍射效应都包含两部分内容:格子的衍射和格子内所含分子的衍射。一套平行线的衍射就是一行与这些平行线相垂直的点(图 6a、b),而这些点间的距离是与平行线间的距离成反比的。对由两套平行线组成的平面网格,它的衍射就是相应两行点的乘积。将一行点的原点顺序放在第二行点的各点上,就得到这一乘积的结果(图 6c)。所以一个平面网格的衍射仍为一网格结构,但后一网格点间的距离与产生

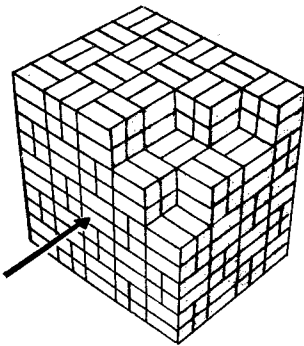


图 5 晶体的三维周期重复性——一个四方晶格示意  
图中还显示了对称性的存在。

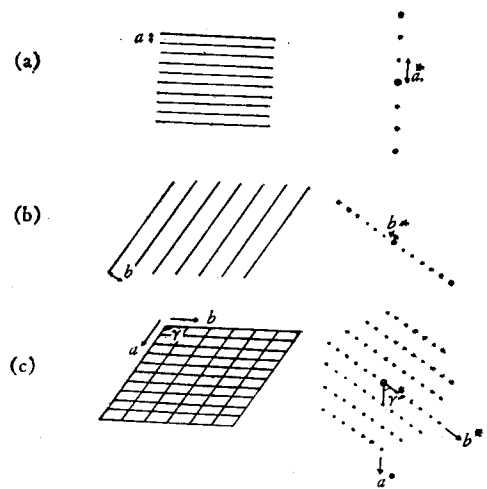


图 6 一套平行线和网格(左)的衍射变换(右)  
〔根据 Holmes, K. E. and Blow, D. M.:  
*The Use of X-ray Diffraction in the Study of Protein and Nucleic acid Structure*,  
1966.)

衍射的网格的线间的距离是成反比例的。在图 6c 所示的情况下就有以下关系:

$$a^* = \frac{1}{a \sin \gamma}, \quad b^* = \frac{1}{b \sin \gamma}$$

这里  $\gamma$  是两行线间的夹角。显然,晶体存在的正空间与衍射效应存在的衍射空间,具有倒易关系。因此,常把衍射空间称为倒易空间,把衍射所产生的网格称为倒易格子。倒易格子和倒易空间是晶体学中至关重要的概念,极为有用的工具。X 射线衍射分析的基本任务,就是把可观察和记录的倒易空间的衍射结构,变换为至今尚极难直接观察的正空间中的晶体结构。晶格中的分子的衍射产生特定的衍射强度分布(图 4),这种衍射作用于每个倒易格子点,并赋予各点由分子变换所产生的强度。换言之,分子衍射在一个倒易点位置有强射线,这一倒易点就是一个有强衍射出现的点;而分子衍射在一个倒易点位置只有弱射线,这一倒易点就是一个只有弱衍射出现的点。图 7 是用旋进像机拍摄的猪胰岛素三方晶体的零层衍射照片,可以清楚看到衍射花样的格子特征和每个格子点上的不同强度。所以,一个分子晶体的衍射花样是一具有特征衍射强度分布的三维格

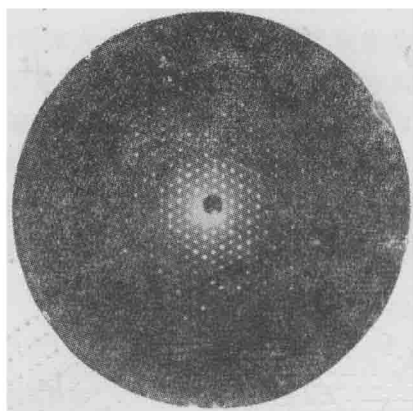


图7 猪胰岛素三方二锌晶体的旋进衍射照片  
( $h\lambda\theta$ ,  $\mu = 20^\circ$ )

$$Q(x, y, z) = \sum_m \sum_n \sum_p A_{mnp} \times \exp[\pm 2\pi i(mx + ny + pz)] \quad (1)$$

式中  $m, n, p$  为整数,  $A_{mnp}$  是级数的振幅, 常称为傅里叶系数。凡符合这种类型的过程, 统称为傅里叶变换。由于它可以定量而完善地处理包括一系列光学变换在内的实际问题, 遂成为一种广泛使用的有力工具。从衍射强度重建晶体结构, 就是与上述过程类似的傅里叶变换。

我们知道, 任何原子都可看作包围着原子核的一定数目电子的凝聚体——电子云, X射线的主要作用对象是这些电子, 而不是由质子和中子组成的核。所以, 一个晶体可以看作是一定空间中的电子云, 它的结构就是这些电子云在晶格中独特的疏密分布(电子云密度)。一种物质分子在特定的晶体中, 有一种特征的三维空间的电子云密度分布; 因此, 求得这一特征

的电子云密度分布, 就可重建该物质分子的结构(实际是一球体), 其格子的几何性质决定于晶体中晶格的类型和空间排布, 其强度分布决定于晶格中分子的结构。相应地, X射线衍射分析一般地可大体分为两步, 第一步, 测量晶体衍射的各种几何参数而不计及衍射的强度, 由这些参数可以求得该晶体所属的晶系、晶胞参数、空间群(晶格类型和对称因素)、晶胞中独立单位(不对称单位)中所含有的分子数等, 常称为X射线晶体学研究。第二步, 测量衍射点的强度, 导出分子结构, 通常称为X射线晶体结构分析。一般, 分子的每一部分对所有衍射花样都有贡献, 因此, 必须测量所有衍射点的强度。

如何从衍射强度重建分子结构呢? 这是一个复杂过程, 但它的原理, 正如一张唱片可以通过唱机重建歌唱家发声时空气密度分布的过程一样。(如图7e和e')。早在19世纪初, 法兰西数学家傅里叶就发现, 任何复杂的波形都可由一系列简单的正弦波加和而成(图8)。声波也是如此。一个良好的声学系统就是将原声中所含有的各个正弦波不多不少地, 完全地传递给听众, 从而加和重建出原有的声波。如果用数学工具来做这件事, 就是傅里叶加和。它的形式如下

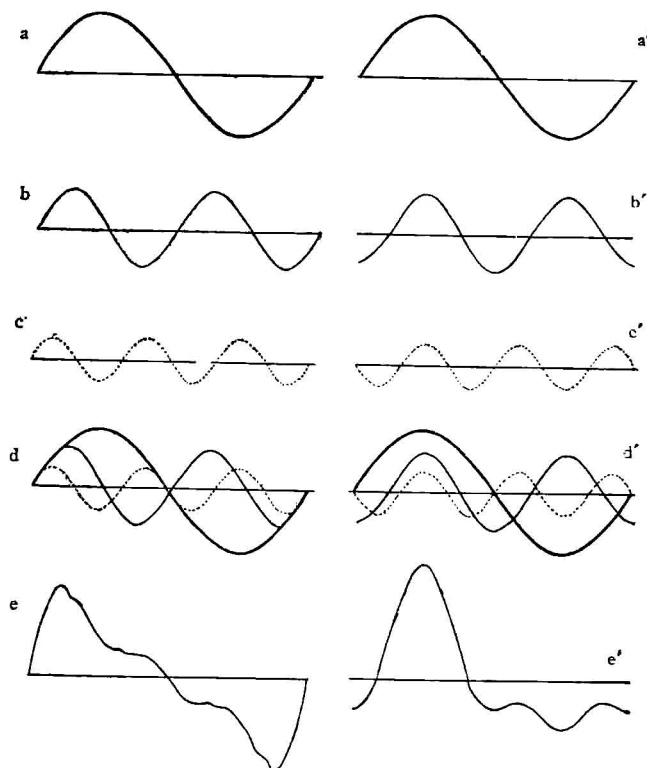


图8 谐波的加和——傅里叶变换

$a, b, c$  三个正弦波在(d)处加和得到复杂的波形(e)。  $a', b', c'$  三个正弦波与  $a, b, c$  相同, 但在加和时的相对位置(相位)不同, 于是得到与(e)不相同的合成波(e')。

构。X 射线衍射分析的最后结果,就是晶体(实际上只是晶胞独立单位范围内)的一套三维电子密度分布图,它就是该晶体及其所含分子的像,常称 X 射线像。从一种物质的 X 射线像,就可直接得到该物质的分子结构。因而,一个晶体就相当于其疏密分布是由电子产生而不是由空气粒子形成的三维声波。这一复杂的三维复合波可以完全分解为简单的正弦波。晶体的 X 射线衍射花样中的每一个衍射点,正是一个这样的正弦谐波。它的数学表达式如下:

$$F_{hkl} = |F_{hkl}| \cos \alpha_{hkl} + i |F_{hkl}| \sin \alpha_{hkl} \quad (2)$$

式中,  $hkl$  是该衍射点在倒易空间中的坐标位置(如正空间中的  $XYZ$  一样),称为衍射指标;  $|F_{hkl}|$  是该衍射波的振幅(称为结构振幅),  $\alpha_{hkl}$  是其相位。  $F_{hkl}$  称为结构因子,这是 X 射线衍射分析中一个最基本的量,可以说,整个复杂分析的最根本任务就是求得晶体衍射产生的全部  $F_{hkl}$ 。显然式中要求的基本量就是结构振幅  $|F_{hkl}|$  和相应  $\alpha_{hkl}$  两项。  $|F_{hkl}|$  很容易直接从衍射斑点的强度(黑度)求得。但求  $\alpha_{hkl}$  没有任何简便的方法,因而成为整个分析困难的主要来源(见下段)。从对声波的分析我们已经了解,一旦我们求得了全部的组成谐波——结构因子  $F_{hkl}$ ,我们就可以用傅里叶变换把它们加和起来,得到三维合成波——晶体空间中的电子密度分布。它的函数形式如下

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} \times \exp[-2\pi i(hx + ky + lz)] \quad (3)$$

式中,  $\rho(xyz)$  表示晶体空间中  $(xyz)$  位置处的电子密度,  $F_{hkl}$  是结构因子,  $hkl$  是衍射指标,加和是分别对  $h$ 、 $k$ 、 $l$  三个方向进行的,因而是三维的。比较(3)和(1),可以看出形式是完全一样的,电子密度分布函数就是一以结构因子为系数的傅里叶加合(Fourier Synthesis)。

### 相 位 问 题

为了算得电子密度分布,就必须求得所有结构因子(式(2));要计算结构因子就必须知道被晶体衍射出来的所有衍射波(相应于衍射照

片上的所有衍射斑点)的振幅  $|F|$  和相位  $\alpha$ 。  $|F|$  与衍射线的强度  $I$  有简单的平方关系

$$I = K |F|^2 \quad (4)$$

式中  $K$  是由一系列物理因素决定的常数,可由实验获得。衍射强度  $I$  可通过各种记录方法测定,于是  $|F|$  可求得。但是,目前尚无任何实验的方法可以记录和测量衍射线的相位  $\alpha$ ,这是 X 射线衍射分析在实际应用中的根本困难,晶体结构测定中大多数复杂而艰难的工作都是为了克服这一困难。

这里所说的相位,是指在各个衍射线作为正弦谐波相加时彼此间应有的正确相对位置关系,有时也称相对相位。图 7 把这一问题表现得很清楚。  $a, b, c$  与  $a', b', c'$  是对应地波长和振幅完全相同的三个正弦波,但从它们波峰或波谷的相对位置可以清楚看到,  $a$  与  $b$  与  $c$  的相对位置关系,同  $a'$  与  $b'$  与  $c'$  的相对位置关系是不相同的,也就是它们有不同的相对相位。当它们彼此加和时,  $a + b + c$  的合成波  $e$ ,就与  $a' + b' + c'$  的合成波  $e'$  完全不同。因此,要从分解开来的组成谐波忠实无误地重建原来的合成波,就必须要知道它们正确的相对相位。实际上,所有波的傅里叶加和都有这一问题,幸运的是,我们的听觉对相位差异极不敏感,它不能区分图 7 中的  $e$  和  $e'$ ,而是将它们作为同一音调接受。所以,在声波的重建过程中,无须去专门解决相位问题。但在用傅里叶变换去重建分子结构时,必须求得参加合成的各组成波的正确相位。

目前只有激光全息术可以记录相位。“全息”的含义正是指可以得到包含振幅和相位在内的全部信息,因而可用全息照片进行直接的立体造像。遗憾的是,至今尚没有可供使用的 X 光激光,也就得不到 X 光全息照像。目前在记录 X 射线的晶体衍射效应时,只能得到振幅这一半信息,而完全丧失了相位信息。力求获得失去的衍射波的相位信息,就是 X 射线衍射分析的中心课题,也是这一技术应用的根本困难所在。在 X 射线衍射分析领域内存在的各种方法,实质上就是解决相位问题的不同方法。尽

管在 30 年代就已经可以很好地记录蛋白质晶体的衍射效应了,但由于相位问题未获解决,科学家面对大量的衍射数据犹如面对一部难解的天书。50 年代初,对蛋白质晶体衍射的解析也不过只有诠释甲骨文的水平,这一领域在当时被视为“科学中的艺术”。直到 50 年代中叶,英国 Perutz 等人提出用同晶置换法来解决蛋白质晶体衍射的相位问题, Kendrew 等人在 60 年代测定肌红蛋白和血红蛋白的晶体结构获得成功,整个状况才有了根本的转变。目前,解决相位问题的方法已有多种,如试误法、向量空间法、反常散射法、分子置换法、直接法等,但对生物大分子来说,最成熟和最有效的方法仍然是多对同晶置换法。由于这一方法配合着反常散射法的广泛应用,已使大量蛋白质和一些核酸、病毒的晶体和分子结构得到阐明,并使生物大分子的 X 射线衍射分析逐步从“科学中的艺术”变为研究生物大分子结构的“常规武器”。

### X 射线衍射分析的特点和局限

用 X 射线衍射分析来获得物质分子的结构,与一个光学显微系统的成像过程完全类似。

图 9 是这两个系统的简要类比。它们的成像有两个阶段:先形成衍射像,后由衍射像形成显微放大像。不同的是在光学显微系统中这两阶段是由透镜系统连续实现的;而在 X 射线衍射分析中这两阶段是分离的。因为 X 光的穿透性极强,至今没有任何透镜系统可以将它会聚成像,只能借助数学计算来实现。这里的傅里叶变换的作用就相当一会聚透镜,人们有时称它为“数学透镜”。光学系统的极限分辨率都受所用光源波长的限制,不能超过入射光波长的二分之一。普通光学显微镜用可见光和紫外光作光源,理想分辨率一般超不过 2000 Å。X 射线衍射分析中用的 X 光(CuK<sub>α</sub> 辐射)的波长是 1.54 Å,分辨率可接近甚至高于 1 Å。X 射线像显示的是分子的三维整体,包含着化学的和空间的全部结构信息,这也是其它同类技术所不能的。

X 射线衍射分析要求研究对象必须是晶体,这极大地限制了它的研究范围。已知的蛋白质(包括酶)不下数千种,能结晶的只占极少数,迄今测定的结构仅 50 来种,同时,这也使它所得到的精细三维结构信息只能是静态的,或更严

(下转第 74 页)

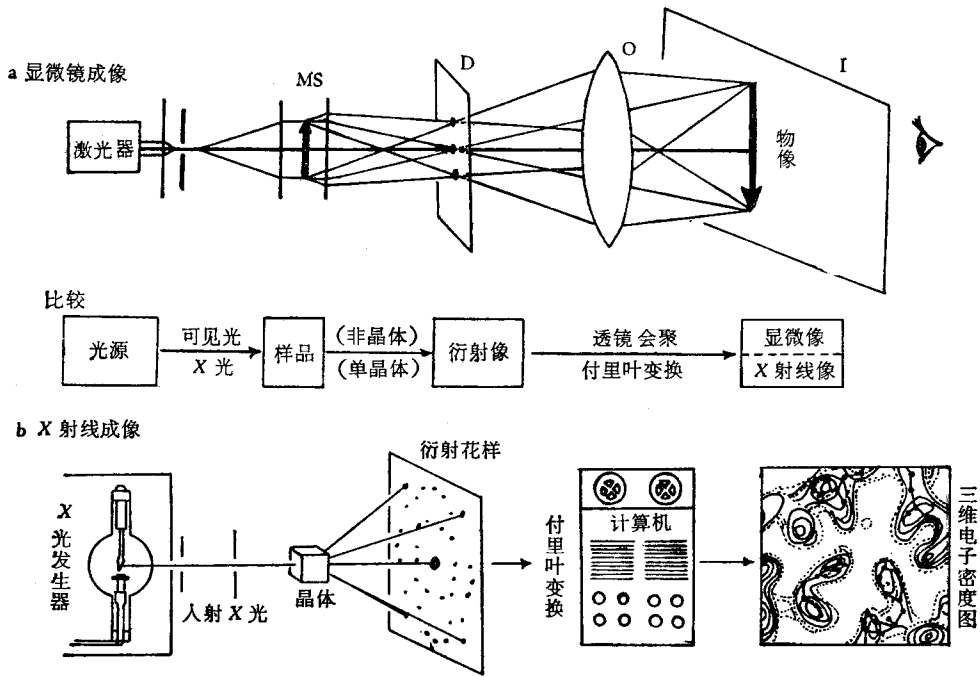


图 9 X 射线衍射分析与光学显微放大的比较

水→用缓冲液平衡柱

洗柱时流速可接近或略高于正常实验时情形。

#### 四、应用举例

大麦  $\beta$ -淀粉酶的纯化: 40ml 提取液(25% 饱和度的硫酸铵的 0.01M 磷酸盐缓冲液 pH 6.8)。柱床体积 30ml, 流速 25 毫升/小时, 洗脱时先用 85 毫升样品缓冲液淋洗, 再用硫酸铵浓度降低及乙二醇浓度上升的梯度洗脱(终浓度分为 0% 及 50%)。

疏水胶在图 7 中为辛基—交联—琼脂糖 4B, 图 8 为苯基交联琼脂糖 4B, 二者都得到较好的分离效果。

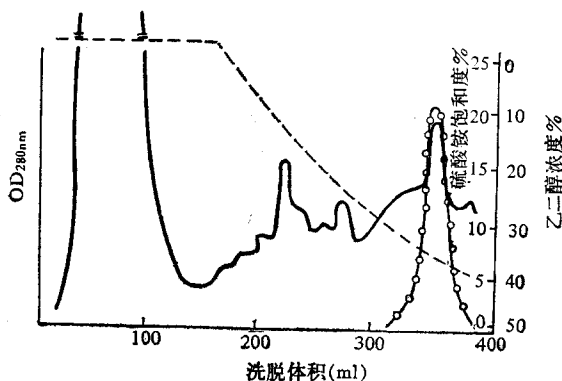


图 7 大麦  $\beta$ -淀粉酶的疏水层析图

— 280nm 下光吸收值  
○—○ 大麦  $\beta$ -淀粉酶  
—— 饱和度硫酸铵梯度及乙二醇浓度梯度

大量的研究工作表明: 在蛋白质分子表面存在着疏水区域, 抗原与抗体免疫结合的分子基础是抗原分子的疏水区域与抗体分子超变区的疏水氨基酸之间的相互作用。金色葡萄球菌细胞壁表面的蛋白 A ( $S_pA$ ) 的分子中的疏水片段 B, 发现可与琼脂糖偶联, 因此可用疏水层析方法纯化 IgG 的  $F_c$  片段<sup>[11]</sup>。应当提到的

(上接第 36 页)

格说,是在时间和空间上的平均结构;它可以测定一个变化过程之前和之后的结构,经过努力也可能测定化学上“凝固”起来的一些稳定中间态的结构,但不能探测动态过程,不能跟踪变化速率。特别对于生物机体,晶态不是其自然状

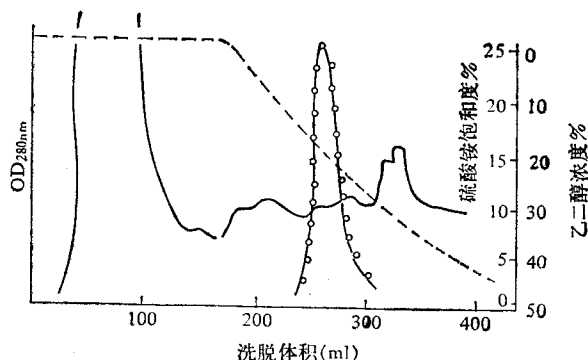


图 8 大麦  $\beta$ -淀粉酶在苯基交联 4B 琼脂糖柱上的纯化效果

— 280nm 下光吸收值  
○—○ 纯化后  $\beta$ -淀粉酶  
—— 硫酸铵梯度及乙二醇浓度

是,疏水作用有时是十分强烈的,所以不少工作中利用这种技术实现酶的固定化而应用在工业中。总之疏水层析正好利用了生物大分子的结构特点,因而有广泛应用的可能性。

#### 参考文献

- [1] Rosengran, J. et al.: *Biochem. Biophys. Acta*, **412**, 51—61, 1975.
- [2] Lewin, S. et al.: *Method of Protein Separation*, Vol. 2, 233—243, 1976.
- [3] Hoffmann-Ostenhof, O. et al.: *Affinity Chromatography*, 141—158, 1977.
- [4] Pahlman, S. et al.: *J. Chromatography*, **131**, 99—108, 1977.
- [5] Shaltiel, S. et al.: *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 3397—3401, 1975.
- [6] Axen, R. et al.: *Nature*, **214**, 1302—1304, 1967.
- [7] Shaltiel, S. et al.: *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 778—781, 1973.
- [8] Er-el, Z. et al.: *Biochem. Biophys. Res. Commun.*, **49**, 383—390, 1972.
- [9] Conner, B. J., et al.: *B. B. A.*, **532**, 122—136, 1978.
- [10] Conner, B. J.: *J. B. C.*, **250**, 3283—3291, 1981.
- [11] Hans Bisswanger and Eva Schminke-off: *Multifunctional Proteins*, 312—315, 1980.

[本文于 1981 年 5 月 18 日收到]

态,且结晶所需要的条件(如浓的盐溶液或有机溶剂)也常与生理条件很不相同,这就要求我们在应用 X 射线衍射分析结果来阐释生命活动时取谨慎态度。近年来一些新的方法(如低温结晶学)和技术(如高强度高准直性同步辐射 X 光源)的应用,正在逐步改善这一状况。