

生物大分子结构的 X 射线衍射分析

(二) 蛋白质和核酸三维结构的测定

王大成 梁栋材

(中国科学院生物物理研究所)

将 X 射线衍射分析技术应用于蛋白质和核酸三维结构测定,经历了相当长的时间。1934 年,英国晶体学家 Bernat 与 Crowfoot 首次摄得一个蛋白质晶体(胃蛋白酶)的 X 射线衍射照片。次年, Crowfoot-Hodgkin 又获得了胰岛素晶体的衍射像。当时使人们十分激动,因为这些结果充分证明,与生命现象密切相关的蛋白质分子在晶格中是完全有序排列的,因而肯定可以用 X 射线衍射技术来测定它们的结构。1937 年, M. F. Perutz 选择了血红蛋白晶体结构测定作为研究课题。当时,因为还没有一般的方法来解决相位问题,测定只含有少数原子的小分子结构还是十分困难的任务,而血红蛋白分子却含有上万原子,作出这一选择要何等的决心和信念!随后的 15 年,是漫长的寻求方法的时期。直到 1954 年, Perutz 及其同事发现,重原子(如 Hg)可以同晶置换的方式进入蛋白质晶体的晶格中,改变原来晶体的衍射效应,从而可用以解决相位问题。随后几年,方法和技术得到进一步发展和完善,建立了多对同晶置换法。1961 年, Kendrew 等人运用这一方法首次测定了肌红蛋白的晶体和分子结构;从而开辟了一个蛋白质结构测定的新时期。自那时以来,已有大量蛋白质结构得到阐明,其中绝大多数都是运用同晶置换法测定的。这些结果使我们对蛋白质本身的认识和一些与蛋白质有关的重要生物学问题的了解,都达到了一个全新的深度。

核酸结构的测定,采用与蛋白质结构测定

相同的方法。但由于结晶困难,使得它的进展比蛋白质更缓慢得多。目前只有 $tRNA^{Phe}$ 、 $tRNA^{Asp}$ 、 $tRNA^{Gly}$ 等的结构在中等以上分辨率测定,一个左旋 Z-DNA 分子片断的结构在高分辨率阐明,尽管为数不多,但对核酸类物质的认识已产生重要影响。非单晶的纤维衍射分析至今对核酸类物质的研究起着重要作用,大家熟知的右手双螺旋 DNA 结构模型主要就是据此提出的。但纤维衍射提供的信息很有限,一般不能直接导出分子的三维结构,限于篇幅,本文将不予详述。

多对同晶型置换法

多对同晶型置换法 (multiple isomorphous replacement method) 是迄今测定蛋白质和核酸分子和晶体结构的基本方法。1937 年 Robertson 和 Woodward 首次将此法引入酞菁染料结构测定,他们用较重的原子置换酞花菁分子中较轻的原子(如用硒替换硫,用碘替换溴),而不改变分子在原来晶格中的排布,也不改变分子的结构。这样制备的含重原子的晶体称为同晶型重原子衍生物。所谓“同晶型”,是要求重原子衍生物中的电子密度分布,除在置换位置有变化以外,其余应与原来的晶体(以下称母体或母体晶体)中的完全一样。以共价键的方式置换蛋白质分子中个别原子相当困难,只有对一些可脱除的金属原子方可做到(如用铅或汞替换锌)。但蛋白质晶体中却有大量的溶剂通道,通过扩散,重原子可以“进驻”,并接近蛋白质分子

表面的溶剂位置,从而与母体分子的某些特定部位的基团结合(通常采取络合、静电吸引、偶极子作用或范德华力作用等方式),同时又不影响母体分子及晶体的结构,严格说来这是同晶“加成”,而非“置换”;但对被取代的无序溶剂分子来说,这的确是置换。

多对同晶置换法就是借助多个同晶型重原子衍生物来解决蛋白质晶体衍射波的相位问题的方法。由于重原子象蛋白质分子那样有序地处于晶格中,会使母体晶体衍射波的强度发生显著变化,利用强度变化量就可以推出相应母体晶体衍射波的相位值。图 1 示出了两个比较典型的重原子衍生物产生的衍射强度变化。从这些照片可以清楚看到,母体与衍生物间衍射点分布的格式未变,只是强弱花样已经显著不同了。

用几何图解的方法,可以清楚地了解利用同晶型重原子衍生物求解衍射波相位的原理。一个正弦波可以用矢量来表征,波的振幅就是矢量的大小,波的相位就是矢量的方向(图 2a)。如果只知道该矢量的大小(波的振幅 $|F_P|$)而不知道它的大小(波的相位 α_P),画在平面上就是一个以此大小为半径的圆(图 2b)。所以,相位值未知的蛋白质母体的衍射波的几何表达,就是以其振幅大小 $|F_P|$ 为半径的一个圆。所谓相位问题,就是要确定从圆心指向圆周的矢量的正确位置。当获得一个重原子衍生物后,就有了三个矢量:母体的衍射波 F_P ,衍生物的衍射波 F_{PH} 和重原子的衍射贡献 f_H 。显然,它们之间有下列关系:

$$F_{PH} = F_P + f_H$$

这在几何上就是一个闭合三角形的关系(图 2c)。从衍射强度测量我们可以分别求得母体和衍生物衍射波的振幅 $|F_P|$ 和 $|F_{PH}|$,同时通过衍生物衍射强度的变化寻找出重原子在晶胞中的位置之后,可以单独计算出 f_H 的大小和方向。因此,我们可以画出分别以 $|F_P|$ 和 $|F_{PH}|$ 为半径的两个矢量轨迹圆,要使它们满足前面列出的闭合三角形关系, $|F_{PH}|$ 圆的圆心必须在 $-f_H$ 矢量的端点上(图 3a)。一般情况下, $|F_{PH}|$ 圆必然与 $|F_P|$ 圆相交两点(H 和 L)这就是母体衍射波相位值的两个解 α_1 和 α'_1 (图 3a)。如果有第二个重原子衍生物,可以得到第二套 $|F_{PH_2}|$ 和 f_{H_2} ,从而可以画出以 $|F_{PH_2}|$ 为半径的第三个圆。此圆与母体圆又交两点 H 和 G,其中一点是真解,它必然也是 $|F_{PH_1}|$ 与母体圆相交两点中之一。换言之,从三圆共交的点 H 就可得到母体衍射波相位解 α 。所以一般说来,至少要有两个同晶型重原子衍生物方可获得母体衍射波相位值的唯一

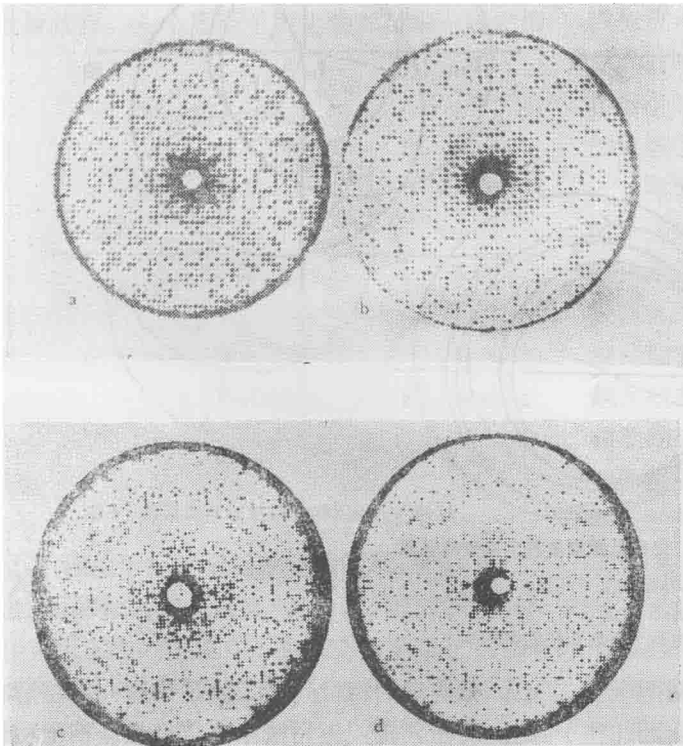


图 1 蛋白质母体和重原子衍生物的照片
(旋进相机, $\mu = 15^\circ$)

(a) 溶菌酶母体 (b) 对-氯汞苯磺酸溶菌酶衍生物 (c) 磷酸化酶母体 (d) 乙基汞硫代水杨酸磷酸化酶衍生物。[取自 Blundell, T. L. and Johnson, L. N. (1976): *Protein Crystallography*.]

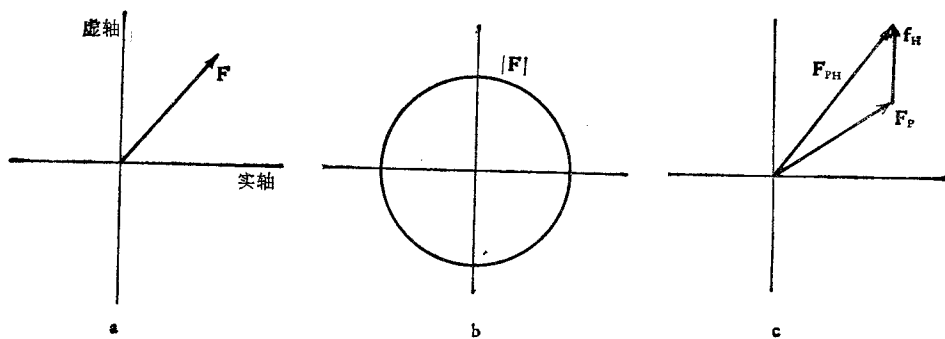


图 2 衍射波的矢量表达

- (a) 振幅和大小均已知的衍射波 F ;
- (b) 只知振幅 $|F|$ 而相位不定的衍射波;
- (c) 蛋白质母体, 重原子衍生物和重原子衍射波 F_P , F_{PH} 和 f_H 之间的闭合三角形关系。

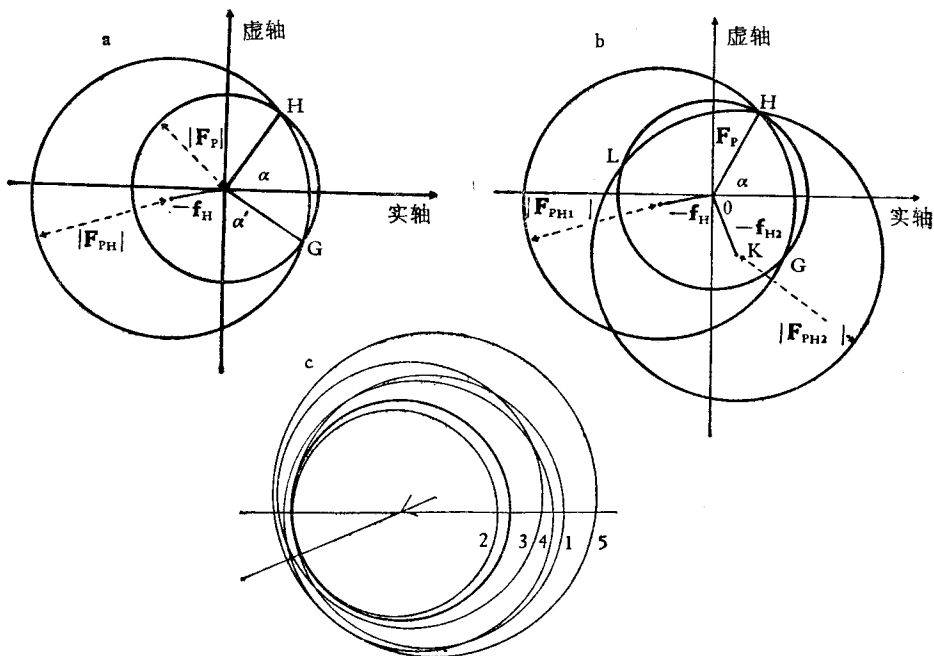


图 3 用同晶置换法测定相位的原理

- (a) 单对同晶, 相位值有双解 α 和 α' ;
- (b) 多对同晶, 相位值获唯一解 α ;
- (c) 肌红蛋白一个反射相位的实际测定。

解。在实际中, 由于各种误差的存在, 用两个重原子衍生物并不一定能得到一个交点, 而往往需要多个才能得到较为准确的解, 称为多对同晶置换法。图 3c 是肌红蛋白用 5 个重原子衍生物求解相位的实际情况。所以, 重原子衍生物的质量越好, 数量越多, 结果也越准确。上述实际工作过程, 称为 Harker 图解法。肌红蛋白

6 Å 分辨率的相位问题, 就是用这一办法解决的。现在, 普遍应用概率来求得最佳相位和最可几相位值。

一些其它方法

其中一些可能有很好的发展前途, 但迄今还只起着辅助作用。现择要介绍如后。

1. 反常散射法(anomalous scattering method)

如果晶体中的电子处于完全自由状态或仅由一种原子组成,则晶体对X射线的衍射空间总是具有中心对称的,这就称为衍射强度的中心对称定律(Friedel定律)。但实际上,原子中的电子都处于束缚状态,其散射能力比自由电子或大或小,相位值也不同,二者之间存在着一个修正量。在这修正量中的一部分,其散射波的相位超前原子散射波的相位 90° ,这一部分称为反常散射校正量。如果晶体中存在吸收边接近入射光频率的重原子,就会产生明显的反常散射效应。其结果就使上述衍射强度中心对称定律遭到破坏,并可用于帮助测定相位。由于反常散射效应的存在,一个重原子衍生物的衍射球的正半球内的衍射振幅 $|F_{PH}^+|$ 就与负半球的衍射振幅 $|F_{PH}^-|$ 不同,结果各自相当于一个衍生物,如图4所示。所以,原则上说,利用反常散射效应,一个重原子衍生物即可唯一确定相

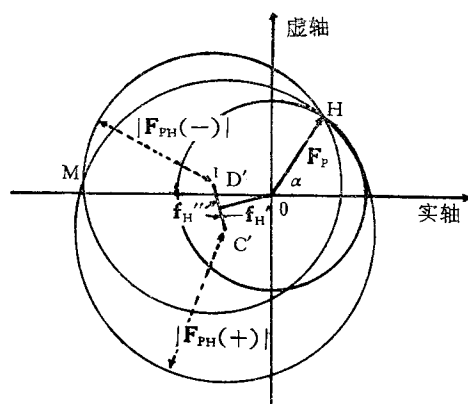


图4 利用反常散射效应确定相位的原理

位。但由于相对于衍射振幅 $|F_{PH}|$ 的大小而言,反常散射量一般比较小,因而误差比较大,在实际工作中往往有不少困难。事实上,至今虽有一些单对同晶加上反常散射测定结构的实例,但目前这还不是一个普遍适用的方法。

反常散射法是同晶置换法的一个重要补充和辅助手段。实用中的同晶置换途径,都是以同晶置换加上反常散射法作为基础的。近年来,由于同步辐射的应用,可以获得单色性极好的各种波长的X光光源,便于选择适合于各种

原子的吸收限的X光频率,使得反常散射法已有可能发展成为独立解结构的方法。最近美国Hendrickson只用蛋白质母体晶体,利用反常散射成功地解出了两个蛋白质的结构,其中一个只含三对二硫键作反常散射中心,另一个仅含三个铁原子。

2. 差值电子密度法(difference Fourier method) 一个蛋白质的结构一经解出,它与某些分子结合形成的复合物的结构、它的同族蛋白质的结构及其某些类似物的结构,常常能以完全相同的密堆积方式(空间对称群)结晶,这时它们的结构测定通常无需再经同晶置换,而只用差值电子密度法就可解决。

1964年,Stryer等人首次用此法研究叠氮离子与肌红蛋白的结合。由于这一方法极为简便易行,随后广泛用于研究抑制剂和类底物与大量蛋白质的结合,特别对于探知许多酶的活性位置,对了解酶的作用原理,发挥了重要作用。同时,它可用于测定与蛋白质表面结合的小分子的结构,用作探测母体蛋白质构象变异的灵敏手段。

3. 分子置换法(molecular replacement method) 这是利用晶胞(即晶体中最小的重复单元,晶体就是由这些最小单元在三维方向上周期重复所构成)的对称独立单位中具有结构的类似性或等同性,来帮助测定结构的方法。它越来越多地应用于下述研究中:当一个基本蛋白质的结构已经测定,在对那些结晶类型(空间对称群)不同的它的类似物以及它的同源“家族”进行系列研究时,其余的结构往往无需同晶置换途径,而只用分子置换法即可测定。

一个晶体中具有局部对称性和不同晶体具有结构类似性,一定会在衍射花样中反映出来。如前文所述,晶体的衍射花样是由分子的傅里叶变换和晶格的变换两部分合成的,所以,分子的结构类似,其衍射花样必然有某种类似。分子置换法就是利用这种类似性来帮助测定蛋白质结构的方法。它一般包括下列三个步骤:

(1) 测定独立分子或分子的亚基在一个晶胞中或在不同晶型的晶胞之间的相对取向。这

可由一个矢量函数 (Patterson 函数) 的系统考查来完成,称为旋转函数。

(2) 对给定的坐标系及其原点, 确定亚基的平移量。这也可用 Patterson 函数作系统检测来解决,称为平移函数。

(3) 建立和求解分子置换方程。通过求解相位, 结合结构修正技术, 测定结构。

显然, 应用分子置换法的基本条件是, 或在不对称单位内部, 或在不同晶型的晶胞之间, 必须具有分子 (或亚基) 结构的相同或类似。而且, 一般只能在相关系列中至少有一个结构已经较精细地测定时, 方可用分子置换法获得完整的结构解。它的优越之处在于, 可以回避同晶置换途径的冗长而繁杂的实验和计算过程。

结构测定的一般步骤

多对同晶置换法测定蛋白质结构, 一般要经历下列步骤:

1. 培养蛋白质晶体 获得足够大的蛋白质单晶体是进行结构测定的先决条件。许多蛋白质在生化研究中就可得到微晶, 然后精心培养, 使它以单晶形式长大到约一毫米线性大小, 方能进行较满意的 X 射线衍射分析。足够的大小, 一定的厚度, 优良的衍射强度和较高的分辨率, 这些都是晶体质量的重要指标。要获得可用的单晶体, 必须掌握好影响晶体生长的因素, 诸如溶解度、pH、温度以及溶剂成份 (缓冲体系与离子强度) 等。很多蛋白质只能提供很小的量, 因此近年来十分注意微量生长的方法和技术, 其中气相扩散法得到较广泛运用。迄今, 对蛋白质晶体生长的规律所知尚不多, 经验和机遇仍是成功的重要因素。

2. 制备重原子衍生物 得到蛋白质晶体后, 要设法将重原子引入, 制备同晶型重原子衍生物。质量好的重原子衍生物必须符合两方面要求: (1) 引入的重原子对蛋白质分子的构象和在晶体内的堆积破坏极小, 具有足够高的同晶度; Crick 和 Magdeff 曾计算, 在 $3 \text{ \AA}$ 分辨率, 晶胞参数 0.5% 的变化就会产生 15% 的衍射强度变化。衍生物晶胞尺寸的变异, 一般不应大

于此量。(2) 重原子结合专一, 位置集中, 占有率高。对一些特定的蛋白质, 可以用一些专一性很强并能很好控制的方法 (如用重原子标记的酶的专一性抑制剂, 引入特定的功能基团作为重原子的结合位置等) 制备衍生物。但对大多数情况, 普遍使用的是浸泡扩散法或共晶法, 这基本上还是一种类似药物筛选的过程。图 5 是重原子试剂对-氯汞苯磺酸进入肌红蛋白晶体形成重原子汞衍生物的情况, 带有重原子的小分子正好结合在两个分子之间的合适的空间里, 并被蛋白分子表面的相应基团所结合。

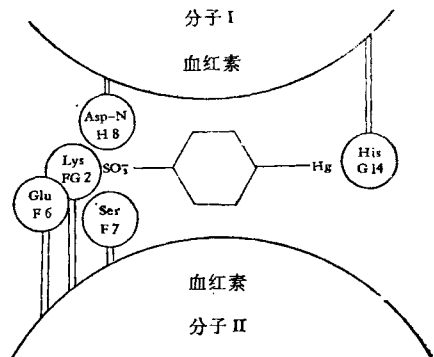


图 5 肌红蛋白与对-氯汞苯磺酸的结合

[取自 Watson, H. e. et al.: *J. Mol. Biol.*, 8, 166, 1964.]

3. 收集和校正处理衍射数据 高质量的衍射数据是获得良好结果的基础, 必须予以充分重视。蛋白质晶体衍射波, 数量大, 强度弱。其收集方法有两大类: 照相记录和衍射仪记录。六十年代以衍射仪法占优势, 由于它测量精确, 易于自动化, 所以得到广泛应用。主要设备是带计算机控制的线性衍射仪和四圆衍射仪, 至今四圆衍射仪仍在广泛使用。图 6 是其原理和机械装置示意图, 通过图中所示的四个圆周运动, 晶体可达到任何所需的取向位置, 从而对入射的 X 光产生衍射, 并被准确定位的探测器 (计数管) 收集。进入七十年代, 由于计算机控制的光密度计的出现, 底片感光的衍射强度测量问题获得解决, 照相法又重新获得生命力。由于底片作为一个面探测器, 记录迅速, 特别对分子量较大的生物大分子的研究带来一

系列好处,近年已越来越广泛地应用起来。

收集到的数据,必须经过对一系列实验因素(偏极化因子,角因子,温度因子、吸收因子、不同晶体的统一因子等)的校正和处理。其中,吸收大概是最重要的系统误差,也是数据处理中最困难的问题。

4. 测定衍生物中重原子的位置 从衍生

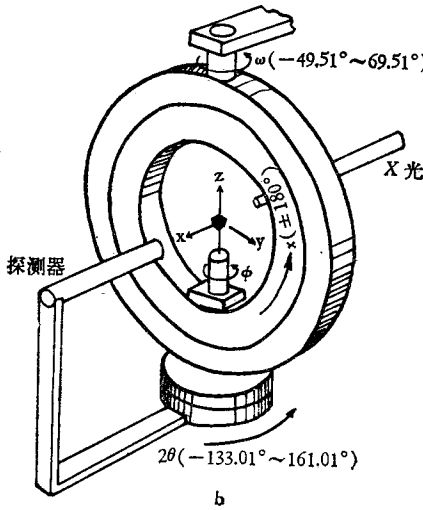
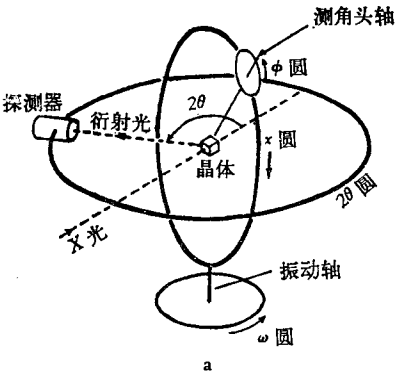


图 6 四圆衍射仪原理示意
(a) 晶体和探测器的四圆运动;
(b) 机械装置。

物与母体间衍射强度的差异测定重原子的位置,一般用矢量分析法,其基本的数学工具就是 Patterson 函数。同晶置换和反常散射数据的联合使用,常能得到最佳结果。

5. 计算衍射波的相位 现在计算上多用同晶置换加上反常散射的联合概率法,从相位概率密度分布函数求得最佳相位 ϕ_b 和最可几相位 ϕ_M ,用以计算电子密度。

6. 电子密度图的计算和诠释 求得蛋白质衍射波的相位值后,即可与通过强度测量获得的振幅合并,用“数学透镜”傅里叶综合,求算三维的电子密度图。这就是 X 射线衍射分析的

直接结果,对它进行仔细的诠释和分析,就可得到分子的三维结构信息。三维电子密度图的分析是一个复杂的问题,必须综合利用各方面知识。一般,在低分辨率 ($dm > 5 \text{ \AA}$),只能得到分子的大小和形状,在中等分辨率 ($4 \text{ \AA} > dm > 2 \text{ \AA}$),可获得肽链的走向和侧链的空间取向;在高分辨率 ($dm < 2 \text{ \AA}$) 可获得一些原子水平

的信息,有的报道甚至已看到氢原子。目前,少数蛋白质结构分析的分辨率已达 $1.2 \text{ \AA}$ 。

7. 修正(精化)结构

通过上述同晶置换途径测定的结构,严格说来,准确度都很低。一般情况下,结构的误差因子(结构模型与衍射实验数据之间的偏离因子 R)常常达到 40% 左右,原子坐标误差 $1 \text{ \AA}$ 左右。无论从结构的观点还是从结构与功能关系研究的要求,都需使结构进一步精化。运用各种数学物

理方法提高结构精度的过程,称为结构的修正。蛋白质结构修正的主要困难在于,分子太大,原子太多,从而使得衍射数据与要修正的结构参数相比较,显得不足。最近几年发展了一系列蛋白质结构修正技术,包括带立体化学制约的差值傅里叶法、最小二乘法、能量极小化修正等,已取得了显著成效。目前,通过修正,结构分析的分辨率已可达 $1.2 \text{ \AA}$,精度已可达 $0.1 \text{ \AA}$ 左右,而且可以确定晶体环境中相当量的水和溶剂分子的位置和结构,从而能够深入研究蛋白质分子与其所在环境的关系。

【待续】