

晶体衍射分析方法

郝 权 张蔚哲

(香港大学医学院)

在自然界中，同样是化学元素碳，石墨漆黑而柔软，金刚石却剔透而坚硬。同样是化学元素磷，白磷剧毒而易燃，红磷却无毒而且更加稳定（图 1）。这是由于他们微观上的原子排列方式不同而造成宏观上化学和物理性质的差异。同为胰岛素，人胰岛素和牛、猪的胰岛素其氨基酸组成有着明显的差异，但是却执行相同的生物化学功能（图 2）。这是由于他们有共同的结构特征。大量的科学事实表明，物质的宏观性质是源于他们本身的微观结构。探索物质的结构与性质之间的关系，是凝聚态物理、结构化学、材料科学、

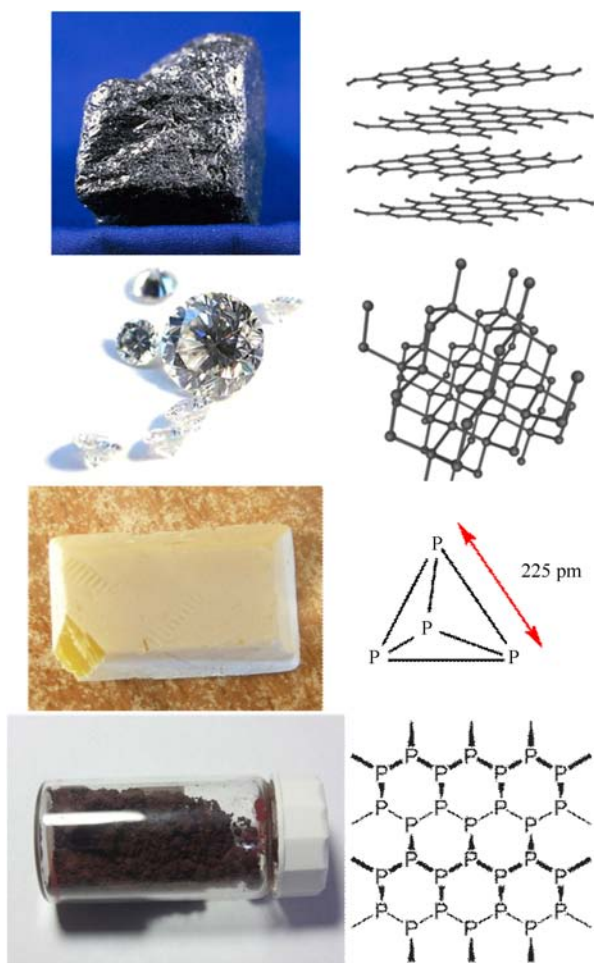


图 1 自然界中的同素异形体以及他们的分子结构，从上到下分别是石墨、金刚石、白磷、红磷（图片来自中文维基百科）

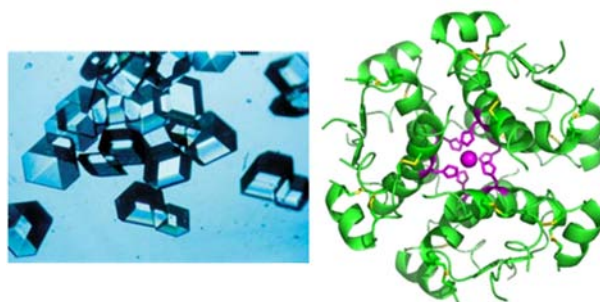


图 2 胰岛素晶体和胰岛素六聚体结构（图片来自维基百科）

分子生物学等许多学科的一个重要研究内容。晶体是原子、离子或者分子按照一定周期性在空间中有序排列，形成的具有一定规则的几何外形的固体。晶体结构分析是在原子尺度上测量晶态物质微观结构的主要手段，它与上述众多学科有着密切的联系，是物理学中的一个小分支。它主要利用了晶体中的物质是周期性有序排列的这一特性，利用晶态物质对 X 射线、电子以及中子的衍射效应来测定物质的微观结构。以 X 射线晶体学为例，X 射线可以和分子中的电子发生相互作用产生散射，但是单分子对 X 射线的散射过于微弱而且很难测量，而晶体中含有数量巨大的方位相同的分子，这些分子的散射效应叠加起来就可以产生足以被探测到的信号，因而晶体就是 X 射线信号的放大器。从这些散射信号中可以分析得到电子密度的空间分布图，从而得到原子的位置信息即结构信息。除了常见的钻石、石英、食盐、金属和合金等晶体之外，有机物和生物大分子也能形成晶体。对有机和生命物质的研究可以算是晶体结构分析最辉煌的成就了。晶体结构分析服务于许多不同的学科，同时许多学科的发展反过来也会对晶体结构分析产生深刻的影响。另一方面，晶体结构分析有自己独立的学科，它本身的发展又对所服务的学科起着促进作用。

晶体结构分析的方法主要有两大类，一类是以 X 射线衍射为代表的衍射分析方法，另一类是以电子显微镜技术为代表的显微成像方法。电子显微成像也可以认为是前后两个相继的电子衍射过程，因

此，可以说衍射分析是晶体结构分析的核心。用衍射分析方法测定晶体结构的理论依据，在于晶体结构和它的衍射效应之间存在着互为傅里叶变换（Fourier Transformation）（式（1））的关系：

$$\rho(r) = \frac{1}{V} \sum_H F_H e^{-2\pi i H \cdot r} \quad (1)$$

这里说的衍射效应，是指晶体向各个方向发出的衍射波的振幅和相位，在式（1）中用结构因子 F_H 来表示， H 表示衍射指标， r 则表示晶体中原子的位置， $\rho(r)$ 表示电子密度的空间分布， V 表示晶胞体积。结构因子是物质与入射的 X 射线相互作用的数学表达，衍射指标是一种符号系统，它可以将衍射点在倒易晶格中的位置用简单唯一的三个整数 (hkl) 表示。一旦测量或者计算出衍射效应的振幅和相位，就能够用逆向傅里叶变换计算出晶体结构。虽然衍射效应的振幅能够在衍射实验中测量出来，但是目前还没有普遍实用的方法可以直接测量由晶体发出的衍射波的相位。因此想要解出晶体结构，必须事先找回“丢失”了的（无法直接测量的）相位。这就是晶体学中的相位问题，它一直是研究晶体结构分析的关键问题（图 3）。用于解决相位问题的方法主要有四种：直接法、同晶置换法、反常散射法和分子置换法。

同晶置换法

早期用于解决小分子晶体相位问题的方法是所谓尝试法。其要点是：先根据已经掌握的线索猜想出一个结构模型，再从这个结构模型计算出一组理论衍射强度，然后同实验所得的衍射强度作比较并据此对模型进行修改。上述步骤须经多次反复，直至理论和实验的衍射强度得以吻合。小分子晶体中原子数量少，衍射分辨率高，能够达到原子级分辨率（高于 1.3 Å），因此可以在不需要考虑初始相位的情况下用尝试法测定晶体结构。而生物大分子中含有数量巨大的原子数

目，需要尝试的参数太多，再加上生物大分子晶体往往难以达到原子级分辨率，尝试法在此时就无能为力了。这样，寻找更有效的解决相位问题的方法便提上了日程。早在 20 世纪中后期，英国的布拉格（W. L. Bragg）和考克（J. M. Cork）为解决相位问题分别提出了所谓的重原子法和同晶置换法。原子的衍射强度和它的原子序数的平方成正比，因此哪怕是混在成千上万的轻原子（原子序数较小）中，单独一个重原子（原子序数较大）对衍射强度的贡献也很容易被观察到。重原子法的大意是：假定晶体中含有少数重原子，而它们的位置是已知的，这时就可以通过 Patterson 法或直接法计算出重原子对相位的贡献并将其当作全体原子相位的一个初始估计。用这样的相位配以实验测得的衍射振幅就可以近似的计算出一幅代表晶体结构的电子密度空间分布图。同晶置换法的要点则是：如果能够制备出待测晶体的重原子衍生物，而且衍生物的晶体与母体晶体是同晶型，这时如果已知重原子的位置，就可以根据母体和衍生物两者在衍射强度上的差异来推算相应的衍射相位。对于蛋白质晶体来说，除了蛋白质大分子在空间中周期性有序排列之外，大分子间还存在孔洞和通道，通常在这些通道中填满了水分子或者结晶母液中的其他溶剂分子。因此，如果将蛋白质晶体浸泡在含有重原子离子或者含有重原子的小分子的溶液中，重原子就有可能通过这些通道进入蛋白质内部，并且在一定的位置置换溶剂分子，形成该晶体的重原子衍生物。这种置换一般不会引起大分子结构和整个晶体晶型的明显变化，形成较好的重原子衍生物。重原子位置的计算方法见后文直接法章节。理论上如果只获得一组重原子衍生物的数据，经过计算其解并不是唯一的（相位模糊），因此通常会结合多对不同的重原子衍生物的数据来计算，得到更

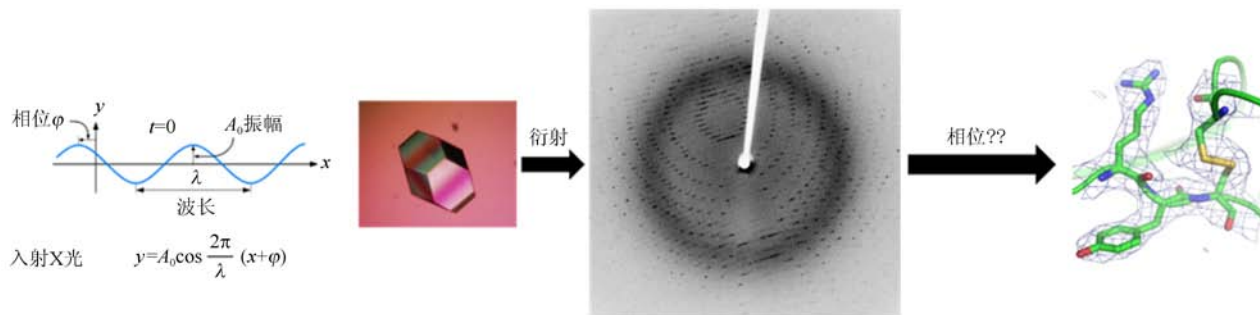


图 3 晶体结构解析（图片编辑自维基百科及网络图片）

加精确的相位 (MIR)。同晶置换法是最早被提出的解决生物大分子相位问题的方法,世界上第一个蛋白质晶体结构就是用同晶置换法解析的,作为这项工作的代表人物,肯德鲁 (J. C. Kendrew) 和佩鲁茨 (M. F. Perutz) 获得 1962 年诺贝尔化学奖。现在同晶置换法仍被用于解析全新结构。同晶置换法的难点在于如何制备同晶型的重原子衍生物。

反常散射法

反常散射法与同晶置换法都属于重原子亚结构法,也叫做实验法。反常散射法利用了重原子对特定波长 X 射线的吸收这一特性,也是通过确定重原子位置来推算全体原子的衍射相位。我们已经知道, X 射线入射晶体都会产生散射,其本质是 X 射线和原子中的电子发生相互作用,电子受到 X 射线的激发,在原子核周围受迫振动,形成了发射电磁波的次波源。当 X 射线携带的光子的能量足够高,等于电子的跃迁能级时,会激发电子跃迁到激发态,这些受激电子在跃迁回低能级时会释放光子,这部分散射的 X 射线的相位较入射 X 射线有延时,这种现象被称为反常散射,此时 X 射线的波长 (或能量) 被称为原子的 X 射线吸收边。原子的 X 射线吸收边是原子的一种固有属性,不同原子的 X 射线吸收边一般不相同。如果能够调节 X 射线的波长,记录吸收边前后产生不同衍射结果,就可以据此推算出重原子位置,并计算出相位,这就是多波长反常散射法 (MAD)。使用这个方法的前提是 X 射线的波长必须根据重原子的特性加以调节,一般实验室用的 X 光机是固定波长 X 光机,无法满足这个方法的使用条件,所以反常散射法就要利用同步辐射的可变波长光源来完成。一般应用反常散射法时会在表达体系中加入硒代甲硫氨酸 (SeMet, 硒和硫在元素周期表中是同族元素,它们有相似的化学性质,硒代甲硫氨酸就是硒原子取代了甲硫氨酸中的硫原子) 来取代甲硫氨酸,使蛋白质形成过程中带入重原子硒 (Se),接下来蛋白质晶体会在硒的吸收边附近进行衍射实验,利用 MAD 方法计算相位。反常散射法是现在解析全新结构的主要方法,它的难点在于制备硒代蛋白质,有许多蛋白质尤其是复杂的哺乳动物蛋白很抗拒反常散射原子的取代。

随着方法和技术的不断发展,现在也可以只采集单波长的反常散射数据来计算相位了,这就是单波长

反常散射法 (SAD),它结合了电子密度修正以及直接法 (见下文) 的方法来破除单波长情况下存在的相位模糊问题。除了硒原子,蛋白质中天然存在的硫原子也有反常散射现象,领域内已经有一些利用硫原子的反常散射求解蛋白质晶体相位问题的报道。但是硫原子的 X 射线吸收边波长比较长,反常信号很弱,对数据的收集处理以及相位求解的算法都有极高的要求。

分子置换法

分子置换法是现在使用最广泛的蛋白质晶体相位解析方法,在蛋白质数据库中有大约 60% 的晶体结构是用分子置换法解析的。我们知道不同的分子结构会导致不同的衍射图样,或者说衍射强度分布特点。反过来,在分子结构和晶体排列上完全相同或相似的晶体,在衍射图样上也必然出现相同或相似的特征,这就是分子置换法解决相位问题的基础。如果有两个分子结构上相同或者相似,其中一个的结构已经被解析,那么就可以将这个已知结构作为初始模型带入待测晶体的衍射数据中,设法找出它旋转和平移的可能位置,从而得到未知待测结构的初始模型。该初始模型经过进一步的建模和修正后就能得到完整且准确的最终结构。随着蛋白质结构的增加,人们发现天然蛋白质的折叠方式是有限的,序列上具有同源性的蛋白质往往也具有相同的折叠方式。因此,一般来说只要未知蛋白能够在蛋白质数据库里找到同源的已知结构,就能在取得衍射数据之后利用分子置换法快速解析结构。由于不需要制备重原子衍生物或者硒代蛋白,分子置换法是最经济的解决相位问题的方法。但是应用分子置换法需要有已知的同源结构,因此它并不适合解析很新的蛋白结构。除了解析未知的同源结构,分子置换法还被用于测定非晶体学对称、酶和抑制剂的复合物中酶和抑制剂相互作用等方面。

直接法

无论是同晶置换法还是反常散射法都依赖于已知的重原子位置,确定重原子位置主要有两种方法, Patterson 法和直接法。Patterson 法利用了衍射振幅的平方为系数计算傅里叶级数,这样一个级数是晶体中电子密度分布函数的自卷积,在一定的条件下可以从中提取有关晶体中重原子位置的信息。直接法利用了三相位结构不变量 ($\varphi_{-h} + \varphi_k + \varphi_{h-k}$) 符合一定概率分布的特点来推导相位信息,在满足一定的相位关系的

约束条件下，由一组衍射的振幅“直接”推定他们自己的相位。直接法通常用于解析小分子或者小蛋白质（小于 1000 个非氢原子）的结构，它要求晶体衍射分辨率达到原子级别。在蛋白质晶体学中，直接法通常用于解析重原子亚结构，或者结合单对同晶置换或者单波长反常散射破解相位模糊问题，可以说直接法是实验法相位解析的基础。随着直接法的新算法以及它在标准化、自动化方面的不断发展，在 20 世纪 70 年代，直接法终于在小分子晶体结构分析中取代了 Patterson 法而占据统治地位。直接法的成功，成十倍的提高了晶体结构分析的能力和效率。它有力的推动了结构化学的发展并促成了药物设计的创立。为此，直接法的两位先驱豪普特曼 (H. Hauptman) 和卡勒 (J. Karle) 于 1985 年获得诺贝尔化学奖。中科院物理所的研究人员，从 20 世纪 60 年代起就从事着直接法的研究，并在其后的几十年里让这一方法持续地焕发出活力，该研究的领导者是范海福院士。除了在传统领域，直接法在电子显微镜学、非公度晶体、大分子晶体学方面也有应用。

(1) 从 X 射线晶体学到电子显微学

高分辨电子显微学是研究固体材料微观结构的重要手段。许多材料因其晶粒太小或缺陷严重而不便使用 X 射线分析，但却适于用电子显微镜观察。X 射线晶体学中的直接法，实质上是一种特殊的图像处理方法。在高分辨电子显微学中引入直接法，可创立新的图像处理技术，大量减少实验工作量和计算过程。1994 年，这一方法成功地用于处理一张高 Tc 超导材料 Bi-2212 的高分辨率电子显微像，使其分辨率从 2 Å 提高到 1 Å。这是由晶体衍射分析催生的直接法在电子显微学图像处理中的应用。

(2) 从周期性晶体到非公度晶体

通常晶体结构分析都假定晶体具有严格的三维周期性，但实际晶体的原子往往存在元素置换、缺位或偏离平均位置等缺陷。基于衍射效应的晶体结构分析，只给出大量晶胞的平均结果。如果这种缺陷的分布本身具有周期性，就形成所谓调制晶体。缺陷分布的周期若为晶体周期的整倍数，即形成公度调制结构或称超结构；缺陷分布的周期若非晶体周期的整倍数，则形成非公度调制结构，它是晶体缺陷长程有序分布的一种形式。国际上用于测定非公度调制结构的

流行方法均在某种意义上属于尝试法，因此有必要建立一种更客观、直接的方法去代替尝试法。1986 年姚家星、范海福等在物理所建成了一个含有作者原创方法的直接法软件包 SAPI (SAPI 自左至右是 Structure Analysis Program with Intelligent control 的缩写；自右至左则是 Institute of Physics, Academia Sinica 的简称)。SAPI 是当时世界上唯一的、能够自动解析“超结构”的晶体学软件包，曾被国外两家分析仪器公司（日本的 RIGAKU 公司和美国的 MSC 公司）采用为单晶体结构分析的主程序。非公度调制结构，就其整体而言在三维空间不具备严格的周期性，但它可表示为一个 n 维 ($n>3$) 周期结构的三维“超截面”。为了在 n 维空间中求解晶体结构，需将现有的晶体结构分析方法从三维空间推广到多维空间。郝权、刘一苇、范海福在 1987 年首先将直接法推广到多维空间，建立了测定非公度调制结构的直接法相位推演理论。1992 年物理所将此方法成功地通过电子衍射分析，测定高 Tc 超导材料 Bi-2223 晶体的非公度调制结构。2003 年，物理所还将用于电子显微学图像处理以及用于从头测定非公度调制晶体结构的直接法，综合到程序包 VEC 中。该程序自从在网上 (<http://cryst.iphy.ac.cn>) 发布以来，已有来自 66 个国家和地区的三千多人下载。范海福所在的科研团队因晶体赓对称性和超结构求解的研究以及 SAPI 软件包的开发等工作，获得 1987 年国家自然科学基金二等奖。

(3) 从小分子晶体到生物大分子晶体

蛋白质的晶体结构分析是结构生物学的重要实验基础。结构未知的蛋白质可分为两类，一类是本身结构未知，但有结构已知的同源类似物可供参照；另一类则是“完全未知”，即没有结构已知的同源类似物。测定前者的主要方法是“分子置换” (MR) 法，测定后者的主要方法是“多对同晶型置换” (MIR) 法和“多波长反常散射” (MAD) 法。MIR 和 MAD 的共同缺点是对试样制备有特殊要求，且实验量和计算量都较大，遇到试样不易制备或者易受辐照损伤的情况就不便使用。因此，用单对同晶型置换 (SIR) 法或单波长反常散射 (SAD) 法来代替是合乎逻辑的出路。但从 SIR 或 SAD 的实验数据不能唯一确定衍射相位，多数情况下每一个衍射的相位都有两个可能的解。这就是所谓的“相位模糊”问题，它成为使用

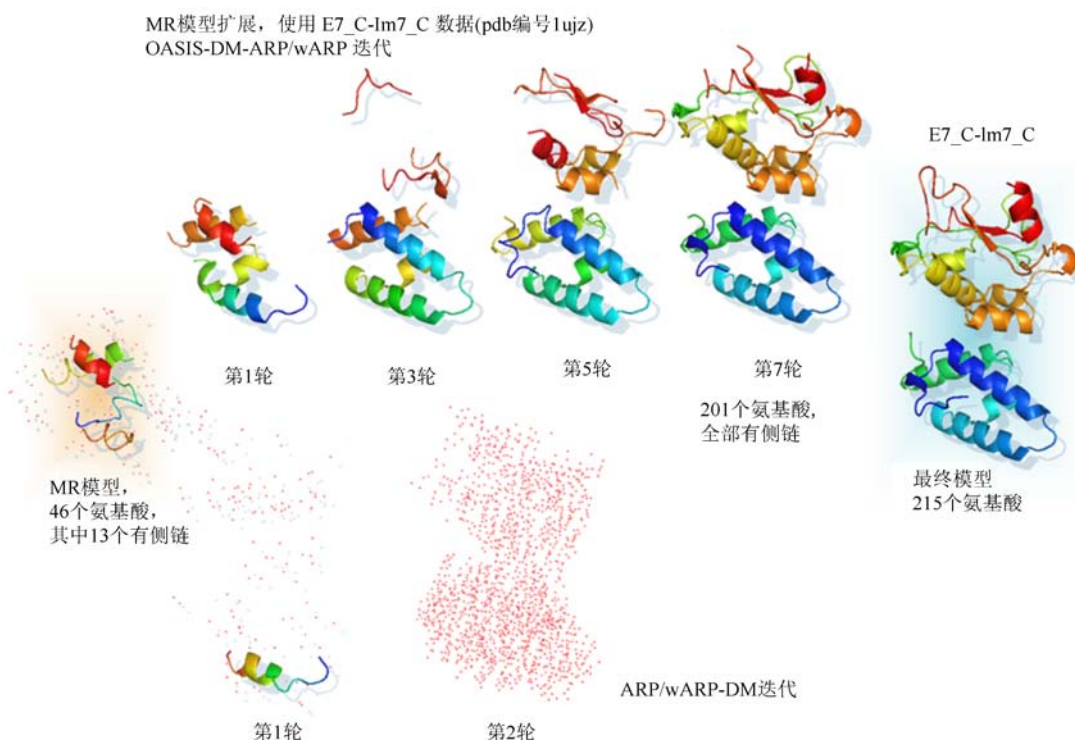


图4 基于直接法的双空间分子置换结构模型迭代扩展 [Acta Cryst. (2007) D63, 793-799]OASIS 结构碎片双空间迭代扩展方法的一个实例, 蛋白是 E7_C-Im7_C, 初始模型只有 46 个氨基酸, 其中只有 13 个氨基酸有侧链, 在经过若干轮双空间迭代后模型长度扩展到 215 个氨基酸, 迭代过程展示为上排图片; 而传统方法则不能合理有效的扩展初始模型, 展示位下排图片

SIR 或 SAD 方法必须设法克服的障碍。范海福在其 1965 年的第二篇论文中提出用直接法中的“分量关系式”（由范海福提出的一种变形 Sayre 方程）破解 SIR 或 SAD 的相位模糊问题。1990 年, 该研究组在国际上首次用直接法推定一套 2.0 Å 分辨率的 SAD 数据的相位, 获得可跟踪解释的电子密度图。1995 年, 他们进一步提出用直接法和电子密度修饰法协同处理蛋白质的 SAD 数据, 并用 3.0 Å 分辨率的 SAD 数据证实这种方法可解出蛋白质 streptavidin 的晶体结构。这个结构原本需要用 3 倍于 SAD 的 MAD 数据求解。1999 年, 他们在国际上首次用直接法帮助从一套 2.1 Å 分辨率的 SAD 数据解出一例原属未知的蛋白质结构。2000 年, 基于该研究组的方法编写的程序 OASIS (One-wavelength Anomalous Scattering and single Isomorphous Substitution) 首次发行, 并被国际上广泛使用的蛋白质晶体结构分析程序库 CCP4 正式采用, 成为其中唯一的用于推演 SAD 或 SIR 衍射相位的直接法程序。2004 年, 他们提出 SAD 或 SIR 衍射相位的“双空间迭代”方法, 将原有方法的功效提高了好几倍, 使直接法在蛋白质晶体结构分析中, 从相位推

演的环节进一步渗透到自动建模的环节。2007 年, 他们又提出无须 SAD 或 SIR 信息的“结构碎片双空间迭代扩展”方法。该法将直接法与测定蛋白质结构中使用最广的分子置换法相结合, 显著提高了它的功效 (图 4 为 OASIS 结构碎片双空间迭代扩展方法的一个实例, 蛋白是 E7_C-Im7_C, 初始模型只有 46 个氨基酸, 其中只有 13 个氨基酸有侧链, 在经过若干轮双空间迭代后模型长度扩展到 215 个氨基酸, 迭代过程展示为上排图片; 而传统方法则不能合理有效的扩展初始模型, 展示为下排图片)。范海福早年的合作者林政炯在该项研究中起了关键的作用。范海福作为其科研团队的代表获得 1996 年第三世界科学院物理学奖以及 2006 年陈嘉庚数理科学奖。

直接法技术的发展, 是整个晶体学衍射结构分析的缩影, 可以预期, 随着晶体结构分析法在理论和技術上的不断发展, 它将在更广的领域、更深的层次发挥更大的作用。

致谢: 感谢范海福先生和古元新先生提供本文写作素材。