

X射线自由电子激光的原理和在生物分子结构测定研究中的应用*

时盈晨 刘海广[†]

(北京计算科学研究中心 复杂系统研究部 北京 100193)

2017-11-01 收到

[†] email: hgliu@csrc.ac.cn

DOI: 10.7693/wl20180703

The principle of X-ray free electron lasers and their applications in biological molecular structure determination

SHI Ying-Chen LIU Hai-Guang[†]

(Beijing Computational Science Research Center, Division of Complex Systems, Beijing 100193, China)

摘要 X射线具有短波长和强穿透能力,利用电子对X射线的散射能够研究材料和分子的精密内部结构。信号的质量高度依赖于X射线发射源。2009年,美国能源部下属的斯坦福线性加速器中心国家实验室建成世界上第一台具有原子分辨率能力的X射线自由电子激光设施(LCLS),从此X射线进入激光时代,人类所能使用的X射线的峰值亮度比最强的同步辐射X射线光源提高了100亿倍。文章简要介绍了X射线自由电子激光的发展历程、产生原理和特性,并结合具体实验研究对如何应用X射线激光研究生物学领域的分子结构和动态变化进行总结。

关键词 X射线自由电子激光,单颗粒散射,串行晶体衍射,时间分辨晶体学成像

Abstract As an electromagnetic wave of short wavelength and high energy, X-rays have been used to probe the fine structure of materials and biological macromolecules for over a century. The first hard X-ray free electron laser (XFEL) facility in the Stanford Linear Accelerator Center national laboratory was commissioned in 2009, and since then its cutting-edge technology has provided highly brilliant, ultrashort pulsed, and fully coherent X-rays, transforming many research fields by expanding the capability of studying fine structures at high temporal and spatial resolutions. The number of XFEL facilities constructed or planned is growing rapidly around the world. In this article, we will explain the principle of an XFEL and summarize the major characteristics of X-rays generated by XFELs. With their unique properties, several milestones have been achieved in the application of XFELs in visualizing the structure and dynamics of molecules in biological sciences.

Keywords X-ray free electron laser, single particle imaging, serial femtosecond crystallography, time-resolved crystallography

* 国家自然科学基金(批准号: 11575021, U1530401, U1430237)资助项目

1 X射线的产生

X射线本质上是电磁波，其波长范围大致从0.01 nm到10 nm，与可见光(400—700 nm)不同，X射线的短波长可以探测物质内部的精细结构，因此自从被伦琴发现以来就被用来观测物质的内部结构。随着人造X射线光源的亮度和稳定性的提高，其应用范围涵盖物理、化学、生物、材料、医学等诸多领域。X射线的产生机制也是基于其电磁波的本性，下面简单比较一下电子射线管、同步辐射设施以及自由电子激光设施的原理和不同之处。

1.1 电子管

X射线最初是因真空管中阴极发出的某种射线遇到玻璃管壁发出荧光而被发现的(因为是未知的，故命名为X射线；又因为伦琴最早发现这种现象，也被称为伦琴射线，Röntgen rays)。直至今日，电子管依然是产生X射线的最普遍的装置，其包含阴极和阳极两个电极，在真空或充气室内工作，电子在其中以高速撞击钨靶，产生X射线韧致辐射(带电粒子与原子发生碰撞减速而产生的辐射)。电子管X射线源的最大问题是效率极低，其99%的能量以热能的形式耗散，并且要求焦点小、强度大，需要对阳极进行冷却。为了降低焦斑的温度，后来出现了旋转阳极X射线管和装有控制栅极的X射线管。但电子管装置无法产生高强度、高偏振、相干性好的X射线束，很大程度上限制了其在生物样品或材料学中的衍射和散射成像的应用。

1.2 同步辐射

在高能物理的研究中，电子在加速器中达到相对论速度，接近光速运动的电子在电磁场中发生偏转时，沿切向方向可以发出电磁辐射。这种电磁辐射的波长范围可以通过控制电子速度和磁场强度进行调制，比如可以使发射的电磁辐射主要为X射线波段等，这种电磁辐射被称为同步辐

射。同步辐射装置一般架设在环形电子同步加速器中，能够产出从远红外到X射线波段的连续谱辐射光，并可以通过应用磁铁阵列实现大范围可调波长、高准直、高偏振性和高亮度的X射线输出。同步辐射设施插入件一般分为两种：强磁场的扭摆器(wiggler)，电子轨道扭曲大，曲率半径小，可以提高同步辐射光子能量；相对弱磁场的波荡器(undulator)，电子轨道扭曲小，曲率半径大，可以提高同步辐射的亮度和相干度^[1]。

目前同步辐射光源共经历三代，第一代是高能物理实验对撞机的兼用机，第二代是在同步辐射专用储存环上建设的专用机，第三代是经过优化能产生高亮度准单色光的基于电子储存环的专用机。现位于上海的同步辐射光源(SSRF)即属于第三代同步辐射光源。利用X射线同步辐射进行的生物大分子晶体学成像是一种非常成功的解析蛋白结构的方法，然而繁复的结晶过程却限制了这种高品质的X射线在解析复杂的生物分子复合体或者具有高度柔性的分子的结构研究方面的应用。为了能够实现对微小晶体或单分子进行成像研究，且最大程度地减小辐射损伤(X射线的非弹性散射给样品注入额外的能量会产生自由电子基，破坏分子结构)带来的影响，一种新的成像模式逐渐发展起来，被称为“损伤前探测法”，也就是利用超短的X射线脉冲对样品进行超饱和攻击，把几万亿的光子压缩到几飞秒的脉冲内发射到样品上，在这么短的时间内，X射线的破坏作用不会完全展现出来，理论上可以突破传统测量方法的极限。类似于武侠小说中极锋利的宝剑，明明已经刺穿敌人，而敌人却还不知道。自由电子激光就是我们在结构生物学探测中的神器。

1.3 自由电子激光

X射线自由电子激光(XFEL)亦被称为“第四代光源”，其亮度更高，全相干，可调谐并且以超短脉冲形式发生。自由电子激光器由直线电子加速器、屏蔽装置和扭摆器构成。为了产生自由电子激光，被加速至准光速的电子束通过周期性横向磁场(扭摆器)产生摆动从而形成正弦状路

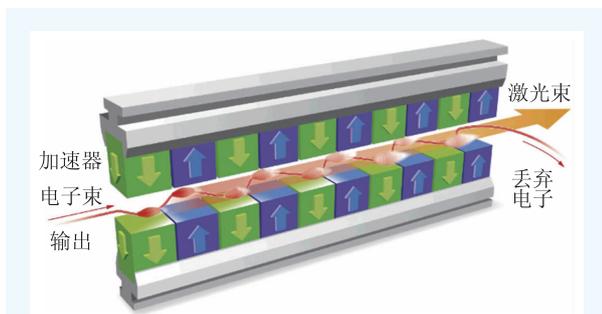


图1 自由电子激光扭摆器原理。从加速器输出的高能电子束输入到扭摆器中，在SASE模式下工作，产生激光(图片根据 European XFEL 资料修改，https://www.xfel.eu/facility/overview/index_eng.html)

径，电子因扭转运动发射电磁辐射(图1)。初始的电磁辐射是低能且不相干的。足够长的扭摆器阵列使得电磁辐射与电子束产生长时间作用，在自放大自发辐射(SASE)模式下电子沿运动方向群聚成为尺寸接近光波波长的周期性束团。在此过程中，束团将自身动能转化为光场能量，增大光场振幅直到饱和输出，得到自由电子激光^[2]。自由电子激光的中心频率取决于输入电子能量和扭摆器参数，频率带宽取决于扭摆器的周期数^[3]。目前美国LCLS、欧洲European XFEL、日本SACLA以及瑞士SwissFEL都是基于自放大自发辐射原理的XFEL装置。

然而，在原生SASE模式下产生的X射线自由电子激光时间相干性较差，也即单色性不好；为了解决这个问题，需要输入相干性较好的种子激光到电子束中(seeding)，并结合谐波产生技术，输出全相干的X射线自由电子激光。常用的谐波产生技术包括高增益谐波产生技术(HGHG)，回声型谐波产生技术(EHGG)以及相位汇聚谐波增强技术(PEHG)。

目前最为先进的欧洲XFEL装置，其输出最短激光波长0.05 nm，脉冲时长10—100 fs，重复频率27 kHz，平均亮度达到 10^{25} photons/(s·mm²·mrad²·0.1%bandwidth)^[4]。位于美国的LCLS，输出最短激光波长0.15 nm，脉冲时长20—120 fs，重复频率120 Hz，平均亮度为 10^{22} photons/(s·mm²·mrad²·0.1%bandwidth)^[5]。高重复频率光脉冲和精细的探测设备导致实验过程中产生巨大的数据流

量，这给数据的传输、存储和处理带来了巨大的挑战。

2 X射线自由电子激光的应用

2.1 自由电子激光的成像原理

XFEL成像的物理原理是光子与物质中电子的相互作用导致样品对X射线激光的散射，而散射的强度分布函数则可以用来反推电子云的分布，也就是样品的三维结构信息。

对于空间中某个取向的单颗粒来讲，其成像过程是相干衍射成像，满足光的远场衍射(夫琅禾费衍射)定律。探测器探测到的图样，是紧贴样品后透射光波场的傅里叶变换强度；由衍射定律可知，空间分辨率决定于光子的最大散射角，也即越靠近探测器边缘，其对应的空间频率越高，同时信号也就越弱。

对于晶体样品来讲，晶格中规律排布的颗粒于某些方向上的散射在传播过程中相干叠加，在探测器上形成布拉格衍射点，依据大量布拉格峰的位置和强度信息可以解析出样品结构。晶体的衍射信号强度与晶体中电子数平方成正比，故一个边长为10个立方体晶胞的晶体的衍射强度为单个晶胞散射强度的100万倍，因此利用晶体衍射，即使非常小的晶体也可以有效地测量高分辨率数据。

2.2 损伤前探测

在生物大分子成像的起步阶段，辐射损伤是限制分辨率极限的重要因素。虽然通过结晶或超低温测量等手段有效地缓解了这一问题，但对于生物分子而言，X射线衍射成像的分辨率极限与辐射剂量成四次方反比关系^[6]，辐射损伤对于高频信号的影响不可忽视。然而Neutze等人在2000年发表在《自然》上的论文中根据动力学模拟的结果提出超短强激光下可以实现损伤前探测的想法^[7]。这个猜想被兴奋的科学家们在德国DESY

和美国 LCLS 的自由电子激光设施开展的实验证实。在这些实验中，飞秒量级的超短 X 射线脉冲都可以在样品被完全电离前产生相干衍射图像，极大地降低甚至避免了辐射损伤对成像的影响^[8, 9]。

图 2 所示是 2006 年 Chapman 等人 在 FLASH(位于德国汉堡的软 X 射线设施)光源进行的首次损伤前探测试验的装置和结果^[8]。实验采用了宽 25 fs、波长 32 nm，功率密度达 $4 \times 10^{13} \text{ W/cm}^2$ 的软 X 射线，激光通过刻有图案的氮化硅薄膜并在 CCD 探测器上成像。研究人员成功地得到了样品损伤前和损伤后成像的数据，并展示了样品受到辐射损伤的情况。

目前典型的利用 XFEL 方法解析结构的装置主要由 XFEL 产生设备、样品注入器、快速读取探测器和数据存储及高性能计算机组成(图 3)。

样品注入器有气相注入和液流注入两种，即通过气体或液流将样品颗粒加速、汇聚并喷射到真空室中与 X 射线作用。为了降低真空中蒸腾降温对样品造成的影响，X 光照射位点需十分接近喷嘴位置，样品注入和光脉冲都可以持续进行，不需要中断。气相注射器将样品汇聚到直径几十微米流速 100 m/s 量级的气体束流内，在喷射过程中可以有效蒸腾样品表面的水分杂质，一般用于单颗粒散射成像；缺点是 X 光的样品击中率很低，目前 LCLS 使用的气相注入器的平均击中率在 1% 到 10% 之间。液相注射器束流最窄可低于 800 nm^[11]，且流速只有 10 m/s 量级，故击中率相对较高，但会产生水的散射噪声，一般用于信号强度较高的晶体衍射成像^[12]。

前文提到，XFEL 成像实验产生的数据量是惊人的。目前一次实验产生的原始数据高达 10—100 TB，而正在建设的 LCLS-II 和上海自由电子激光器预计将会产生每秒百 GB 量级的数据流，如此高密度的信息汇聚将对数据的传输、存储和分析提出巨大的挑战。然而，因为 X 射线脉冲较

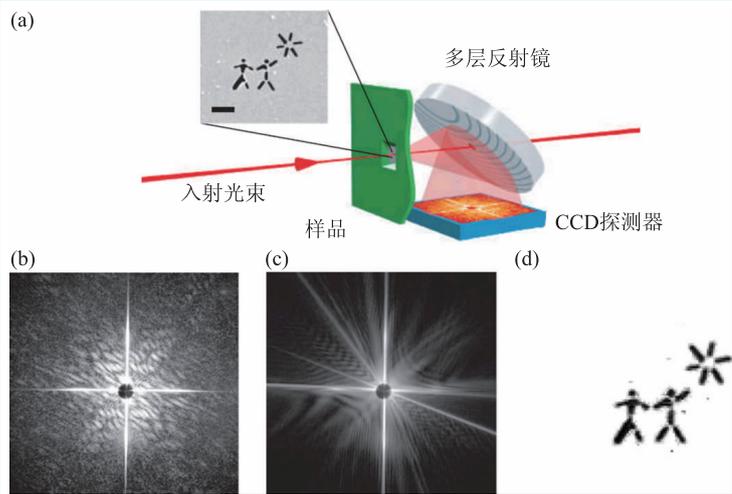


图 2 FLASH X 射线相干衍射实验和辐射损伤效应 (a) 试验装置示意图；(b) 飞秒脉冲的相干衍射成像图。为了让强激光通过，数据中心部位是缺失的；(c) 20 ns 后相似能量和时长的脉冲相干衍射成像图，可以看出首次脉冲对样品造成了辐射损伤；(d) 根据(b)图重建出来的样品。图片修改自文献^[8]

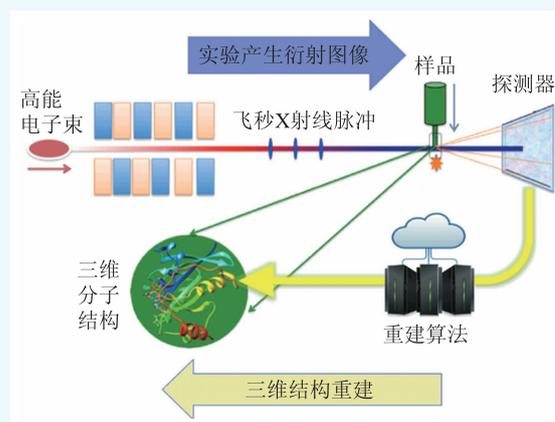


图 3 利用 XFEL 进行“损伤前探测”。图片修改自文献^[10]

低的击中率，在海量的数据中，大部分数据实际没有记录有效信息。对于单颗粒成像，由于光束直径在 0.1 μm 到几个微米的量级，比一般颗粒样品直径稍大，会产生大量多颗粒散射数据；样品表面附着的杂质层(主要是水分子)也无法完全除去，无法避免非样品的散射噪声。同时，探测器故障也会导致很多数据报废无法使用。

为了缩小数据量，首先要进行“命中筛选”(hit-finding)，将记录有效信息的图样保留下来。这个过程一般可以将原始数据减少 90% 左右。在数据处理阶段，则根据需要进一步进行甄选。另外，因为探测器得到的是二维强度图样，无相位

信息,故三维结构的重建算法需要3个步骤:数据挑选、三维重组和相位恢复。

一般来讲,根据样品状态和实验设置的不同,XFEL结构学研究主要有两种具体的方法,分别为单颗粒成像(SPI)和串行晶体衍射成像(SFX),而这两种方法都可以通过与对样品激发装置结合进行动力学研究。到目前为止,单颗粒成像仍然有巨大的挑战,串行晶体学成像法已经较为成熟,具有时间分辨信息的动力学研究也已经取得重要成果。接下来将对这些应用和成果进行简要总结介绍。

2.3 单颗粒成像(SPI)

传统生物学成像手段都需要首先生长大分子晶体,这个过程十分困难,缺少足够的科学指导,很大程度上是基于有限的经验开展的一个艺术创作的过程。更重要的是,在晶体环境中,无法得到生理状态下的蛋白分子结构,也局限了对生物分子动力学和功能的研究。由于XFEL可以产生能够与单个生物颗粒作用的高能窄脉冲,其单颗粒成像模式(Single Particle Imaging, SPI)就提供了这样一种可能,即在不结晶的基础上对单个样品颗粒进行散射成像,利用全同的样品颗粒,在不同的空间取向探测样品结构信息,通过整合拼装并且模型重构得到其在三维空间的电子密度分布,抓取生理环境下的结构特征。

SPI的实验样品一般为体积较大的生物样品或纳米颗粒。生物样品例如拟菌病毒、RDV病

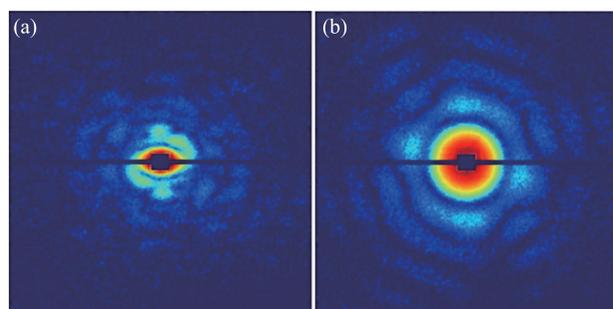


图4 PR772病毒颗粒的单颗粒实验成像图 (a)为非单颗粒散射成像,需要除去;(b)为单颗粒散射成像,用以重建

毒、PR772病毒、T4噬菌体等,纳米颗粒如nanorice、纳米哑铃型颗粒、核壳异构体等。Seibert等人进行的拟菌病毒X射线散射实验是LCLS最早的实验之一^[12],它证实了单颗粒散射的可行性,并进一步表明使用过采样数据可以有效进行二维相位恢复。对实验散射数据进行3D重建的首次尝试是2015年Ekeberg等人完成的^[13],实验样品为直径0.45 μm的拟菌病毒,重建使用了198张散射图,得到125 nm的低分辨率结构,这与结构生物学领域关心的分辨率(比如原子分辨率)有很大的差距。

XFEL的全相干特性使得相干衍射成像(CDI)的技术可以应用到SPI成像中来。对于理想球体的相干衍射理论计算表明^[14],衍射图样的强度与空间频率呈四次反比关系,这导致高频信号与低频信号相比十分微弱。如果只考虑泊松噪声,类似冷冻电子显微镜中使用的平均方法可以有效提高信噪比;但真实情况中,附加噪声如读出噪声、热噪声和漏光等,会在探测器上产生高于一个光子的噪音输出,可以完全淹没远小于一个光子能量的高频实验信号。即使在包含 10^{13} 个光子、直径仅为0.1 μm的窄XFEL脉冲下,理论上也无法得到低于纳米级别的高分辨率数据^[15]。对于高频信号来讲,其强度与入射光强和原子数成正比,所以使用较大的样品颗粒更容易产生有效的实验数据。

图4展示了PR772病毒的单颗粒实验^[16]散射图。数据挑选是重建过程的第一步,即将有效的单颗粒散射成像从海量的、包含有背景散射和多颗粒散射的实验数据中筛选出来。常用的方法一般是基于计算机视觉图像处理的,例如流行嵌入法^[16]和神经网络算法等。

由于在成像过程中样品以自由状态注入,空间取向角未知,故倒空间(傅里叶空间)三维重建的重点在于计算散射图对应的空间取向。在全同颗粒的基本假设下,主要有3种方法。

其一为“共线法”^[17]。同一颗粒两个不同方向的散射强度平面(投影的傅里叶变换强度)一定在倒空间中相交,且交线过中心点。这样,当第

3个散射强度平面加入时，寻找它相对于前两个平面的交线，可以固定3个平面的相对空间取向。“共线法”虽然计算迅速，但噪声对结果的影响很大，特别是高频信号较弱时效果较差。第二种方法为迭代的极大似然法，代表为2009年Loh和Elser发表的EMC算法^[18]。算法在给定的初始粗粒

模型下，评估实验数据与模型的不同角度投影之间的相似度，计算实验数据属于不同空间取向的概率，对模型进行迭代精细化。Ayyer等人发布Dragonfly软件包，完成了对算法的程序实现^[19]。Ekeberg等人对mimi病毒颗粒的三维重建工作即使用了EMC算法^[13]。第三种方法为Ourmazd等人提出的流形嵌入法^[20]。对于同一物体来讲，越相邻角度的投影总是越相似的，故可以利用所有散射数据计算它们的相似矩阵，通过特征值分解等方法将其映射到低维特征空间，与空间旋转群在该空间的特征进行最小误差拟合，得到每张图片对应的空间取向角。当然，流形嵌入算法得到的特征向量也可用于对散射数据进行聚类。虽然这种方法有很多困难，包括空间角增量的映射效果、噪声影响和较高的计算复杂度等，但仍然取得了一些有意义的结果。如图5所示，2017年Ourmazd等人使用扩散映射流形嵌入算法和EMC算法处理PR772病毒的单颗粒散射数据，得到了9 nm分辨率的三维电子密度结构，并且揭示了病毒在生理状态下的连续构型变化^[21]。

在CDI成像中使用的基于投影的迭代算法可被用于单颗粒散射的相位恢复问题。算法的基本思想是将模型在实空间和倒空间之间不断投影，同时施加支撑集等已知约束，使模型在倒空间的强度最大程度地逼近实验得到的三维散射强度。常用的相位恢复算法已经在Maia等人开发的HAWK软件包中被实现^[22]，其中最著名的是Fienup提出的混合输入输出算法(HIO)^[23]和Marchesini提出的收缩算法(Shrinkwrap)^[24]。

单颗粒散射的重建工作是建立在样品全同的假设之上，但一般来讲实验样品的形状无法达到

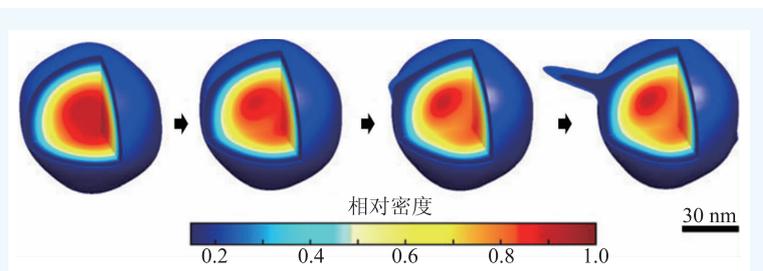


图5 单颗粒成像数据重建出的PR772构型变化。图片修改自文献^[21]

高度均一，存在着尺寸差异和构型差异，例如Kassemeyer等人使用nanorice纳米颗粒的散射数据进行分析时，找到并重建出两个尺寸不同的三维模型^[25]。所以，在处理数据时有效地区分不同形状颗粒的散射图样是十分必要的。对于一些已知结构信息的纳米颗粒，利用精确的模拟数据与实验数据进行对比，可以有效地分析实验数据中颗粒尺寸的分布^[26]；而对于结构复杂的生物样品，借助粗粒化的电子密度模型，也可以通过对比模拟与实验数据在一定准确度下推算颗粒尺寸^[27]。

在单颗粒散射的基础上应用关联函数分析，可进行模型重构和优化。Kam提出的关联函数定义为^[28]

$$C_2(q, \Delta\Phi) = \int I(q, \Phi) \cdot I(q, \Phi + \Delta\Phi) d\Phi, \quad (1)$$

其中， $I(q, \Phi)$ 为探测器探测到的散射谱在极坐标系的强度。对所有空间取向的 C_2 函数进行平均后得到的平均关联函数，包含着单粒子的结构信息，而且有效降低了高斯分布的X光对信号的影响，提高信噪比。以平均关联函数为目标函数，通过逆向蒙特卡罗采样的方法可以迭代获得实空间模型^[29]。近期Kurta等人使用关联函数法成功重建PR772病毒的单颗粒模型，并将这一工作发表在《物理评论快报》^[30]。

目前，单颗粒成像法依然面临着诸多问题和挑战。其一是信号弱，尤其大散射角，也即高频部分的信号，在目前XFEL的亮度下还没有办法达到理想信噪比，故无法得到原子分辨率级别的结构；其次是样品输送困难和命中率低，且实验装置无法避免水等杂质的散射信号对样品信号的干扰，导致实验样品浪费并产生大量的低质量数据。为了解决这些问题，LCLS成立并领导了

“单颗粒成像计划”的国际合作组织(SPI-i), 希望加速SPI技术在XFEL领域的发展。

2.4 串行晶体衍射成像(SFX)

在同步辐射光源进行的大分子晶体学成像(Macromolecular crystallography, MX)是结构生物学最成功的成像方法之一, 至今RCSB蛋白质数据库(PDB)中有89%的结构是使用X射线晶体衍射方法解析出的^[31]。然而, 研究人员需要花费长达数年的时间去优化蛋白的结晶方案, 以令其符合MX实验所需要的晶体标准。而且由于MX的原理是使用单个晶体在不同角度受到反复长时间的辐照, 辐射损伤限制了分辨率极限。XFEL的发展使晶体学成像突破了上述限制, 其高亮度和损伤前探测的特点使得对微小晶体进行衍射成像成为可能, 同时, 超短脉冲还可以用来进行高时间分辨的动力学研究。

串行晶体衍射成像(Serial Femtosecond Crystallography, SFX)使用直径为亚微米到微米的晶体, 晶体被裹挟在连续的液流中注入XFEL光路, 不需要像传统MX一样对晶体角度进行测量。2011年Chapman等人首次报道了使用微晶进行XFEL衍射成像的尝试^[32], 实验在LCLS进行, 使用尺寸0.2—2 μm 的第一类光合作用系统蛋白复合物晶体样本, 在10、70、200 fs的XFEL脉冲下成像。实验用X射线波长6.9 \AA , 单脉冲辐射剂量高达3 G Gy, 最终得到8 \AA 分辨率的重建结

构。该XFEL结构和传统X射线方法得到的结构有着很好的一致性; 同时, 通过比较不同脉冲长度下数据的衍射强度分布, 研究人员在200 fs的数据中发现了明显的辐射损伤。至今为止, SFX已经取得了诸多成果, 例如2012年Boutet等人利用LCLS的硬X射线, 得到了室温条件下1.9 \AA 的溶菌酶结构^[33]; 2015年Suga等人利用日本SACLA装置解析出第二类光合作用系统蛋白复合物的1.9 \AA 结构等^[34]。

在传统的MX成像的重建算法基础上, SFX的数据重建比SPI要成熟得多, 同样需要3个步骤: 命中筛选、三维强度整合和相位恢复。

晶体衍射数据的筛选过程相对简单, 即通过确定是否有明显的布拉格衍射点来判断数据是否为有效衍射数据。考虑到有效衍射可能来自多个不同空间取向的晶体, 或未规则排列的低质量晶体, 或不足以产生强信号的小晶体, 我们需要进一步筛选, 留下可重建的单晶衍射数据。常用的命中筛选软件有Cheetah^[35]、CASS^[36]和cctbx.xfel^[37]等。

三维强度整合分为两个步骤——指标化和方位强度积分。如图6所示, 样品颗粒在理想晶体中规则排布, 但由于晶体尺寸有限(相对MX来讲SFX晶体更小), 其倒易点阵强度不是理想的 δ 函数, 而是类似于单颗粒散射的衍射斑纹。故首先需要在二维衍射图的每一个布拉格点周围一定范围内(ROI)对强度进行积分, 并减去根据ROI之外的区域估计的背景噪声, 作为该点强度值。

三维倒易点阵采用3个指标(h, k, l)作为位置索引, 称为米勒指标, 而指标化就是将二维衍射图的布拉格点映射到三维点阵上的过程。常用的指标化方法只基于布拉格点的位置信息, 包括MOSFLM软件包中实现的基于快速傅里叶分析的自动指标化算法^[38]和REFIX算法^[39], 以及DirAx半自动指标化算法。值得注意的是, 当晶体的晶格对称性高于空间对称性时, 只考虑布拉格点位置信息的指标化过程是存在歧义的, 故

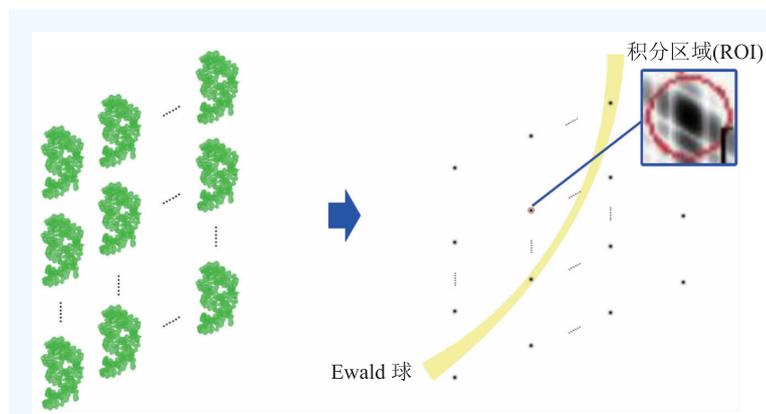


图6 平面晶体及其对应的倒易点阵示意图

必须引入对布拉格点强度的分析。一些算法已经可以处理这个问题，例如使用距离矩阵区分指标化结果的BD算法^[40]和基于极大似然思想的算法^[41]等。

SFX成像的过程实际是对三维倒易点阵布拉格点强度的采样，每一张衍射图只记录了部分布拉格点的部分强度，因为对于每一个布拉格点来说，只有被Ewald球壁覆盖的强度才能被探测器探测(图6)。由于XFEL的带宽较窄，其对应的Ewald球较薄，导致探测到的信息较少。故三维强度整合需要大量的从不同方位探测到的二维布拉格点，在成功指标化后，将相同指标的布拉格点强度累加，逼近真实的倒易点阵强度。这种方法称为蒙特卡罗方位强度积分。

研究发现，有效的后期优化算法(post-refinement)可以减少蒙特卡罗方法所需的衍射图数量，同时提高强度整合的准确性。所谓“后期优化”，即进一步精确每一个二维布拉格点所记录的部分强度。例如White和Sauter等人通过对部分衍射构建几何模型进行精细化^[42, 43]，以及Ginn等人提出的对束流面参数、晶格参数和晶格方位等变量进行迭代优化的方法等^[44]。

SFX的相位恢复方法有分子置换法(MR)、多波长反常衍射法(MAD)和形状变换法等。最常用的方法为分子置换法^[45]，它使用与样品结构相似的蛋白质作为参考模型，在此基础上做微小调整，解析实验样品结构。由于SFX使用的晶体较小，MR方法需要慎重处理晶体外黏附的水等杂质。多波长反常衍射法^[46]利用硒代甲硫氨酸替代样品蛋白的甲硫氨酸，依据硒在不同束流能量下的反常散射对样品结构因子产生的影响，求解相位问题。晶体样品的衍射强度可以表示为

$$I_n(\Delta\mathbf{k}) \propto |F(\Delta\mathbf{k})|^2 |S_n(\Delta\mathbf{k})|^2, \quad (2)$$

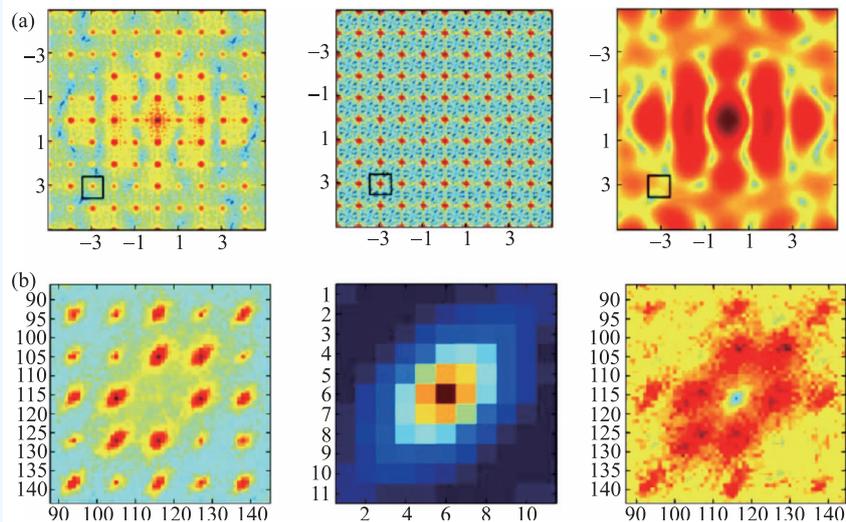


图7 形状变化法相位恢复 (a)模拟数据，从左至右依次为：衍射图样、倒易晶格上的形状变化(调制)函数和调制后得到的结构因子强度；(b)光系统I的SFX实验数据，以相同的顺序排列。图片修改自文献[9]

其中 $F(\Delta\mathbf{k})$ 为与样品结构相关的结构因子，形成布拉格点的有效强度信息；而 $|S_n(\Delta\mathbf{k})|^2$ 导致布拉格点周围形成衍射斑纹，称为形状变化因子。如图7所示，形状变化法^[47]通过实验数据推算 $|S_n(\Delta\mathbf{k})|^2$ ，据此恢复样品的结构因子强度，并对其在CDI成像中常用的直接相位恢复。

以上应用于分子结构解析的重建算法，已经在CrystFEL^[48]、cctbx.xfel^[37, 49]或者Phenix^[50]等软件中加以实现，极大地便利了科研人员的使用。相比于单颗粒成像来讲，串行晶体衍射成像无论是在硬件还是算法上都正在不断趋于成熟，相信在不久的将来，它将成为探索生物大分子高空间、高时间分辨率结构的有力工具。

2.5 时间分辨的动态研究

动态学研究是XFEL应用中最有希望产生可喜突破的分支，它极大地提高了实验所能够达到的时间分辨率极限。由于小晶体对于外界的刺激，例如光激发等，可以发生快速响应，一些样品就可以作为研究其在受到激发后动态变化的理想材料。

在SFX损伤前探测的基础上，时间分辨成像

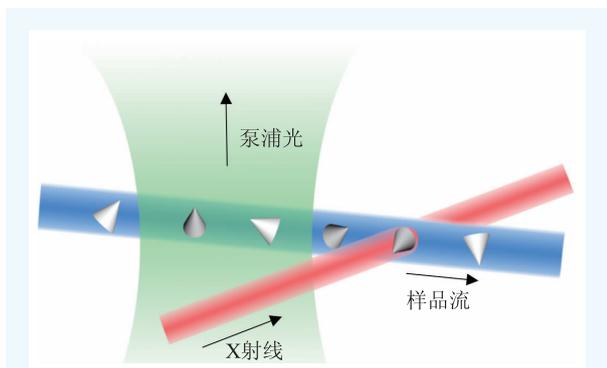


图8 泵浦—探测技术示意图。图片修改自文献[51]

可应用于可逆的以及不可逆的受激过程；而如果使用传统的MX方法，只能探测重复性好的可逆受激过程。基于SFX的时间分辨成像需要蒙特卡罗方法进行三维散射强度整合，而相位恢复和模型重构过程则使用分子置换法(MR)，因为对于很多情况来说，使用的样品基态模型是PDB数据库中已有的蛋白结构。

利用XFEL进行时间分辨成像的最广泛使用的技术是泵浦—探测技术，即在样品注入器和XFEL光路之间加入泵浦光，对注入真空室的样品进行光激发，然后利用X射线成像(图8)。对于连续注入样品的串行晶体衍射来讲，为了保证激发与成像过程的一致性，需要使用宽聚焦的泵浦光，并精确控制泵浦脉冲与XFEL脉冲的时间，以同时提高时间分辨率和受激发样品成像的比例。同时，根据样品颗粒的流速调整泵浦光与XFEL光路之间的距离，可以控制样品从被激发到成像之间的时间间距(以空间换时间)，或者通过调整泵浦脉冲与探测脉冲之间的时间间隔(直接实现时间延时)，都可以实现对样品在激发后不同时间点的观测。当泵浦光后于XFEL发生时，可以探测样品的未激发状态(基态)。同时，由于样品受激前后的衍射数据变化幅度有限，为了降低其他噪声影响，需要额外的探测器探测入射光强度和能量展宽等参数，作为标准化依据。样品流速、泵浦光参数和测量精度等因素共同影响着最终的实验时间分辨率。

首个时间分辨的动态研究是2012年Aquila等人在第一类光合作用系统——铁氧还原蛋白共晶

体上进行的^[51]，研究人员揭示了晶体在被激发5 μs 和10 μs 后发生的变化，并分析了实验结果与光诱导电子转移导致的光合作用系统和铁氧蛋白分离过程的一致性，但遗憾的是，这次尝试未能解析出可支撑结论的蛋白质结构。2014年Schmidt领导的国际合作项目给出了细菌蓝光感光体光敏黄蛋白(PYP)的高空间、高时间分辨率结构^[52]，在10 ns和1 μs 的时间尺度上揭示了PYP受到光激发后向反应中间产物转变的过程中电子密度的变化，并将空间分辨率提升到1.6 \AA 。2016年，Schmidt团队把PYP动力学变化的时间分辨推进到200 fs以内，这是其他方法不能实现的^[53]。

除了泵浦—探测技术外，混合—注入技术发展起来，用以研究不需要光激发、且反应周期通常较长的生物大分子配体结合问题。顾名思义，混合—注入技术即将受体蛋白晶体与配体混合后注入XFEL光路，通过控制混合时间，达到时间分辨成像的目的。2016年Wang等人使用混合—注入法对信使RNA核糖开关与腺嘌呤配体的结合过程进行了研究^[54]，在核糖开关的两种不同构型下得到了相同的反应10 s的中间结合状态结构和反应10分钟的最终结合状态结构，并揭示了配体结合引起的核糖开关晶格结构的变化。作为对比，蛋白质数据库中已知的核糖开关配体结合结构(PDB代码4TZX)与实验解得的最终状态结构有着很好的一致性，证明了混合—注入XFEL时间分辨成像方法的有效性。

3 总结与展望

从2009年至今，XFEL在生物、物理和材料等领域得到了越来越广泛的应用，取得了很多重要的成果，这既得益于光源设施的发展，也得益于高精度、高通量算法和软件的开发。在结构生物学领域，XFEL晶体成像方法发展迅速并且已经趋于成熟，利用其超短脉冲特性发展的高时空分辨率的动态成像研究也取得了许多突破性进展；然而，在单颗粒散射成像方面却依然存在分

分辨率瓶颈。随着能量更高、脉冲更短的欧洲 XFEL 设施和美国 LCLS-II 等光源的建设和使用,新的实验设备和数据分析方法的发展,我们将拥有无限可能,同时也面对很多挑战。为了获得更广泛的应用和更精确的结果,未来 XFEL 的技术突破将集中在这些方面:

(1)提高激光束流能量,同时尽可能降低辐射损伤,提高单颗粒散射的分辨率极限;

(2)控制样品均一性。生物样品具有一定的柔性,这是其实现功能的基础,但是也为数据分析造成了一定困难。样品均一性可以提高结构探测的精度,而必要的样品制备和蛋白质工程的技术可以帮助把分子锁定在某些稳定的形态;

(3)研发有效的样品注入设备,提高命中率并减少样品浪费。比如利用粘稠度高的载体(如 LCP 磷脂分子或者油脂)降低流速,或者把样品固定在对 X 射线透明的底物进行快速扫描,还有利用静电排斥作用的方法产生样品流等方法都比最早开发的液柱喷流法有一定优势;

(4)设计高动态量程和快速记录探测器。这对于发展单颗粒散射和晶体衍射都非常重要,目前的做法是使用双探测器,并且调整为不同光电转换率来增大数据测量范围。将来的 XFEL 产生的

信号将以几十万赫兹的频率出现,配套的探测设备也需要跟进;

(5)设计和开发对快速输出的海量实验数据进行筛选和压缩的硬件和算法。海量数据的筛选越早完成就越有效率,可以降低对后续分析的计算压力。欧洲 XFEL 光源甚至计划在探测器上就完成数据的初步筛选,这无疑是一个很好的出路;

(6)开发更精细快速的重建算法等。计算机技术的发展和超算机群的建设对于支持 XFEL 数据处理至关重要,但是先进的软件也必不可少。在图像分析领域甚至图形重构方面,需要开发新的算法和思路来发展更好的软件。希望当前热门的深度学习能够提供很好的助力。

综上,我们有足够的理由相信,在生物分子结构研究中, XFEL 将继续发挥其特点,提供大量的数据和有价值的信息,尤其是在具有时间信息的动力学研究方面,有独一无二的优势,能够将实验数据与基于物理模型的分子动力学模拟的结果整合起来,最终提供完整的分子电影,满足科学家们的的好奇心,并用于揭示生物分子的功能机理和对生物分子进行理性改造,造福人类。

致谢 感谢孙昌璞院士的建议和鼓励。

参考文献

- [1] Winick H, Brown G, Halbach K *et al.* *Physics Today*, 1981, 34 (5):50
- [2] Milton S V, Yang B X. *Science*, 2001, 292(5524):2037
- [3] Pellegrini C. *Physica Scripta*, 2017, 169:014004
- [4] Ball P. *Nature*, 2017, 548(7669):507
- [5] https://portal.slac.stanford.edu/sites/lclscore_public/Accelerator_Physics_Published_Documents/Forms/AllItems.aspx
- [6] Howells M R, Beetz T, Chapman H N *et al.* *Journal of Electron Spectroscopy & Related Phenomena*, 2009, 170(1-3):4
- [7] Neutze R, Wouts R, Spoel D V D *et al.* *Nature*, 2000, 406(6797):752
- [8] Chapman H N, Barty A, Bogan M J *et al.* *Nature Physics*, 2006, 2 (12):839
- [9] Seibert M M, Ekeberg T, Maia F R *et al.* *Nature*, 2011, 470 (7332):78
- [10] Liu H G, Spence J C H. *Quantitative Biology*, 2016, 4(3):1
- [11] Deponte D P, Mckeown J T, Weierstall U *et al.* *Ultramicroscopy*, 2011, 111(7):824
- [12] Spence J C, Weierstall U, Chapman H N. *Reports on Progress in Physics Physical Society*, 2012, 75(10):102601
- [13] Ekeberg T, Svenda M, Abergel C *et al.* *Physical Review Letters*, 2015, 114(9):098102
- [14] Bohren C F, Huffman D R. *Journal of Colloid & Interface Science*, 1984, 98(1):290
- [15] Huld G, Szoke A, Hajdu J. *Journal of Structural Biology*, 2003, 144(1-2):219
- [16] Reddy H K N, Yoon C H, Aquila A *et al.* *Scientific Data*, 2017, 4:170079
- [17] Bortel G, Tegze M. *Acta Crystallographica*, 2011, 67(Pt 6):533
- [18] Loh N T, Elser V. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2009, 80(2 Pt 2):026705
- [19] Ayyer K, Lan T Y, Elser V *et al.* *Journal of Applied Crystallogra-*

- phy, 2016, 49(4): 1320
- [20] Giannakis D, Schwander P, Ourmazd A. *Optics Express*, 2012, 20(12): 12827
- [21] Hosseinizadeh A, Mashayekhi G, Copperman J *et al.* *Nature Methods*, 2017, 14(9): 877
- [22] Maia F R N C, Ekeberg T, Spoel D V D *et al.* *Journal of Applied Crystallography*, 2010, 43(6): 1535
- [23] Fienup J R. *Applied Optics*, 1982, 21(15): 2758
- [24] Marchesini S, He H, Chapman H N *et al.* *Physical Review B*, 2003, 68(14): 399
- [25] Kassemeyer S, Jafarpour A, Lomb L *et al.* *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2013, 88(1): 042710
- [26] Li X X, Chiu C Y, Wang H J *et al.* *Scientific Data*, 2017, 4: 170048
- [27] Daurer B J, Okamoto K, Bielecki J *et al.* *Iucrj*, 2017, 4(Pt 3): 251
- [28] Kam Z. *Macromolecules*, 1977, 10(5): 927
- [29] Liu H G, Poon B K, Saldin D K *et al.* *Acta Crystallographica Section A*, 2013, 69(4): 365
- [30] Kurta R P, Donatelli J J, Yoon C H *et al.* *Physical Review Letters*, 2017, 119(15): 158102
- [31] Rose P W, Prlić A, Altunkaya A *et al.* *Nucleic Acids Research*, 2017, 45(D1): D271
- [32] Chapman H N, Fromme P, Barty A *et al.* *Nature*, 2011, 470(7332): 73
- [33] Boutet S, Lomb L, Williams G J *et al.* *Science*, 2012, 337(6092): 362
- [34] Suga M, Akita F, Hirata K *et al.* *Nature*, 2015, 517(7532): 99
- [35] Barty A, Kirian R A, Maia F R N C *et al.* *Journal of Applied Crystallography*, 2014, 47(3): 1118
- [36] Foucar L, Barty A, Coppola N *et al.* *Computer Physics Communications*, 2012, 183(10): 2207
- [37] Sauter N K, Hattne J, Grosse Kunstleve R W *et al.* *Acta Crystallographica*, 2013, 69(7): 1274
- [38] Steller I, Bolotovskiy R, Rossmann M G. *Journal of Applied Crystallography*, 1997, 30(6): 1036
- [39] Kabsch W. *Journal of Applied Crystallography*, 1993, 26(6): 795
- [40] Brehm W, Diederichs K. *Acta Crystallographica*, 2014, 70(1): 101
- [41] Liu H, Spence J C. *Iucrj*, 2014, 1(Pt 6): 393
- [42] White T A. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 2014, 369(1647): 20130330
- [43] Sauter N K, Hattne J, Grosse Kunstleve R W *et al.* *Acta Crystallographica*, 2013, 69(7): 1274
- [44] Helen Mary G, Brewster A S, Johan H *et al.* *Acta Crystallogr. D Biol. Crystallogr.*, 2015, 71(Pt 6): 1400
- [45] Rossmann M G. *Acta Crystallographica*, 1990, 46(2): 73
- [46] Hendrickson W A, Ogata C M. *Methods Enzymol.*, 1997, 276: 494
- [47] Spence J C, Kirian R A, Wang X *et al.* *Optics Express*, 2011, 19(4): 2866
- [48] White T A, Kirian R A, Martin A V *et al.* *Journal of Applied Crystallography*, 2012, 45(2): 335
- [49] Sauter N K. *Journal of Synchrotron Radiation*, 2015, 22(2): 239
- [50] Adams P D, Afonine P V, Bunkóczi G *et al.* *Acta Crystallographica*, 2010, 66(Pt 2): 213
- [51] Aquila A, Hunter M S, Doak R B *et al.* *Optics Express*, 2012, 20(3): 2706
- [52] Tenboer J, Basu S, Zatsepin N *et al.* *Science*, 2014, 346(6214): 1242
- [53] Pande K, Hutchison C D M, Groenhof G *et al.* *Science*, 2016, 352(6286): 725
- [54] Stagno J R, Liu Y, Bhandari Y R *et al.* *Nature*, 2017, 541(7636): 242

读者和编者

《物理》有奖征集 封面素材

为充分体现物理科学的独特之美，本刊编辑部欢迎广大读者和作者踊跃投寄与物理学相关的封面素材。要求图片清晰，色泽饱满，富有较强的视觉冲击力和很好的物理科学内涵。

一经选用，均有稿酬并赠阅该年度《物理》杂志。

请将封面素材以附件形式发至：physics@iphy.ac.cn；联系电话：010-82649470；82649029

《物理》编辑部