

用X射线自由电子激光拍摄分子电影*

刘海广[†]

(北京计算科学研究中心 复杂系统研究部 北京 100193)

2021-08-20收到

[†] email: hgliu@csrc.ac.cn

DOI: 10.7693/wl20210906

Making molecular movies with X-ray free electron lasers

LIU Hai-Guang[†]

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

摘要 作为生命活动的基本单元,蛋白质等生物大分子的功能与分子结构和动力学密切相关。X射线衍射和电子显微技术能够确定原子分辨率的分子结构,但是分子在“工作”期间的结构演化还需要大量的方法研究。X射线自由电子激光(XFEL)因其特有的超高亮度飞秒脉冲,成为捕捉动力学过程中分子结构变化的利器。文章介绍了基于XFEL的“泵浦—探测”时间分辨实验方法,通过捕捉蛋白质分子被激发之后特定时刻的结构,拍摄高分辨的分子电影,展现分子在实现功能过程中的精细动力学,为准确理解分子机理奠定了基础。

关键词 分子电影, X射线自由电子激光, 串行晶体衍射, 时间分辨晶体学

Abstract The biological functions of proteins are highly related to their structures and dynamics. The determination of protein structures at atomic resolution has advanced our understanding of the underlying molecular mechanisms, thanks to the development of X-ray diffraction and electron microscopy techniques, especially facilitated by computational algorithms and software. However, most studies are based on static structures, as it is challenging to probe the dynamics of proteins, i.e., the dynamical structures of “molecules-in-action”. The X-ray free electron lasers generate ultrabright femtosecond pulses, enabling time-resolved crystallography in “pump-probe” approaches. By determining a series of structures at specific time points after activation, molecular movies can be produced. Only by revealing the detailed dynamics through molecular movies can the mechanisms be fully understood.

Keywords molecular movie, X-ray free electron laser, serial femtosecond crystallography, time-resolved crystallography

* 国家自然科学基金(批准号: U1930402, 3197110508)资助项目

1 X射线自由电子激光特性与“损伤前探测”

X射线是波长为0.01—10 nm的电磁波,由伦琴发现以后很快被用于研究物质的内部结构。随着人造X射线光源亮度和稳定性的提高(图1),X射线被越来越广泛地应用于物理、化学、生物、材料、医学等诸多领域。根据麦克斯韦方程,对高能电子进行加速(改变速度或/和运动方向)时会辐射电磁波,调整电子能量和加速度就能够获得所需特定波长的X射线。常用的X射线源包括电子射线管,通过高能电子轰击金属靶获得X射线。亮度更高的X射线可以在同步辐射装置产生。电子在同步辐射装置的电子储存环(包括直线部分和装备了磁场的转弯部分)中高速运动达到准光速,再利用强磁场对电子进行调制,从而辐射X射线(辐射的波长由于相对论效应被缩短)。同步辐射设施能够产生稳定的并且波长可调的X射线用于大量实验研究。最近涌现的自由电子激光设施(XFEL)的原理也是利用周期性磁场调制通过其中的准光速电子而获得X射线辐射^[1, 2]。电子枪发射的电子首先被压缩为飞秒脉冲,然后通过直线加速器获得动能,最后接近光速的电子进入波荡器阵列(undulator array)产生的周期性横向磁场,电子被磁场调制从而发射电磁波,这些电磁波又对行进中的电子产生反馈,最终把单个脉冲包含的全部电子调制成为同相位震荡的电子,发射全相干的X射线,也就是X射线自由电子激光。

XFEL激光脉冲有三个特点,即全相干、极高的亮度以及飞秒宽度的脉冲(1 fs=10⁻¹⁵ s)。这三个特点使得XFEL能够对微米尺寸的蛋白质分子晶体以“损伤前探测”的模式进行衍射测量^[4]。“损伤前探测”实验模式可以通俗地理解为,将几万亿个X射线光子压缩为飞秒脉冲,对样品进行“饱和攻击”。由于X射线带来的辐射损伤对晶体的破坏效应需要一定的时间才展现出来(动力学模拟的数据显示大约10 fs以后,单个蛋白分子才会因为库仑排斥力导致显著结构形变^[5]),而XFEL衍射测量在样品破坏之前就已经完成。当然,每

个晶体被XFEL脉冲照射一次之后就因受到“饱和攻击”而损坏,所以需要大量晶体依次进行衍射,也就是串行晶体衍射(serial crystallography)。制备高品质的蛋白质单晶是限制蛋白质晶体学的一个主要瓶颈。很多有重要生物学、医学意义的膜蛋白很难形成足够大的晶体用于同步辐射的衍射实验,但是在一种被称为脂立方相(lipidic cubic phase, LCP)的特殊磷脂分子环境中,可以逐渐排列成为小晶体^[6]。自从2013年首次利用XFEL解析出原子分辨率的GPCR蛋白分子结构以来^[7],串行晶体衍射方法在这类感知外界信号的受体蛋白质研究中发挥了重要作用^[8]。超短的XFEL脉冲还促进了对蛋白质分子结构变化过程(即动力学)的研究^[9]。录制“分子电影”不再是一个梦想,如何录制“分子电影”正是本文重点介绍的内容。

2 录制“分子电影”的难点和解决方案

静止的三维结构可以告诉我们蛋白质的某些特点,比如活性位点、表面特性等,并推测潜在

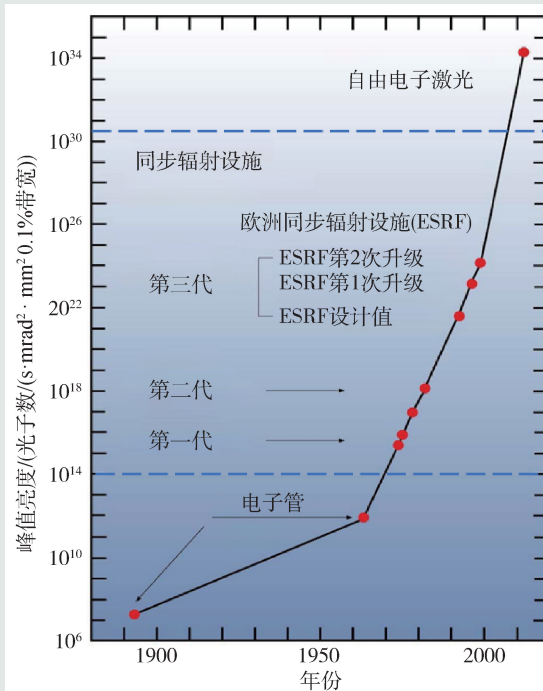


图1 X射线光源的峰值亮度比较。XFEL的峰值亮度比同步辐射设施提升了将近10个数量级^[3]

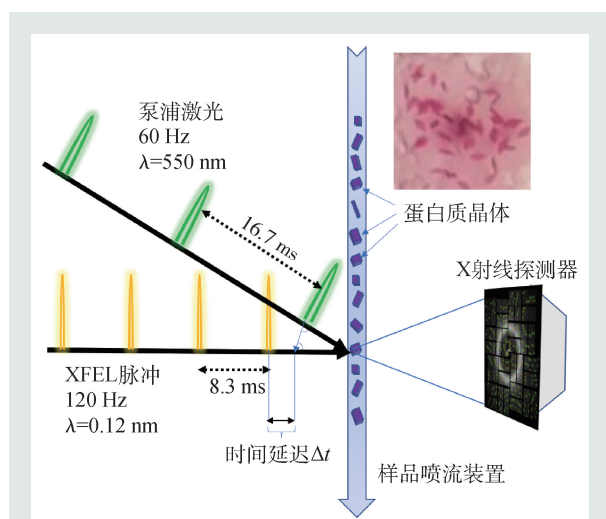


图2 时间分辨串行晶体衍射实验示意图。微米晶体依次输送到X射线焦点进行衍射，在测量之前的 Δt 时刻被“泵浦”激光(或者其他激发装置)启动动力学过程，因此测量到的衍射信号对应激发之后时间点的结构信息。通过调整 Δt ，可以获得不同时间点的信息进行模型重建

的作用机制。这类似于看到了一个陌生人的雕像(或者蜡像)，根据外形、身材、长相等，具有丰富经验的人力资源经理(或者侦探)可以推测其从事的工作类型、有何特殊技能等。基于分子的静态结构(通常对应的是最稳定的状态，即基态)理解生物分子工作机制，难免有偏差、甚至错误。“分子电影”可以在三维空间结构的基础上增加时间维度，把多个静态结构按照时间顺序串联起来，观看“工作状态”的蛋白分子动态过程，加深和提高对分子机制的理解。拍摄“分子电影”需要通过“泵浦—等待—拍摄”三部曲(简称“泵浦—探测”)来实现，通过调整“等待”时间，测量不同时刻的状态(图2)。

蛋白质分子晶体作为“演员”就位之后，录制“分子电影”还需要三个要素：摄像器械，能够指挥演员的导演，以及专业的剪辑。

(1)摄像器械，包括能够探测原子分辨率结构的光源、探测器等。XFEL飞秒脉冲和配套的二维像素阵列探测器正好满足这些硬件需求，其中XFEL的超高亮度可以探测微米尺寸的晶体产生的高分辨率衍射信号。XFEL的飞秒脉冲能够将“曝光时间”缩短到飞秒尺度，是研究超快结构变化的基石。

(2)能够指挥演员的导演，指的是有效的“泵浦”方法。比如，对于感光蛋白质分子结构变化的激发，脉冲激光就是最佳的导演。利用激光对蛋白质分子进行“泵浦”，可以迅速、准确地启动蛋白质分子的结构演化。除此以外，温度、酸碱度、溶液中反应物的混合等都可以作为“泵浦”方法，启动分子结构演化。

(3)专业的剪辑，指的是数据分析方法和软件。“剪辑”是从拍摄到的大量数据中，提取出真正的信号，并且转化为与特定的时间点一一对应的高分辨率结构。有了这些关键时间点的结构，最终才能获得“分子电影”。

先介绍一下摄像器械的细节。XFEL装置的原理以及XFEL激光脉冲的物理参数不是本文的重点，有兴趣的读者可以从《物理》关于XFEL的专题中获得更多的信息(2018年第7期)。XFEL的飞秒脉冲赋予其超高的峰值亮度，是其成为“分子电影”摄像器械的关键属性。XFEL脉冲到底有多亮？如果地球上感受到的太阳亮度为1，第三代同步辐射装置的X射线亮度大约是100亿，XFEL照射的亮度是100亿 \times 100亿(具体指的是峰值亮度。XFEL单发飞秒脉冲包含的光子数约等于同步辐射设施一秒钟累计发射的光子总数)。正因如此，用同步辐射X射线只能测量到低分辨率衍射信号的微米尺寸晶体，通过XFEL脉冲的曝光，经常能够得到原子分辨率的信号(即，在电子密度分布图中可以区分相邻原子)。另一个值得考虑的因素是，生物分子需要水环境维持其生理状态，所以单个生物分子发出的X射线信号会淹没在水溶液产生的背景强度中。在蛋白质分子晶体中，周期性排列的分子能够利用X射线干涉效应将信号集中并放大到晶体对应的布拉格点(Bragg spots)。假想一个蛋白质分子占据边长为10 nm的立方体空间(晶体的基本单元，也被称为晶胞)，将其排列为边长1000 nm(即1 μ m)的立方晶体，则能够包含100 \times 100 \times 100个蛋白质分子。我们知道，X射线衍射信号的强度与晶体中包含重复单元个数 N (对于前面的示例 $N=10^6$)的平方成正比，因此1 μ m的晶体理论上能够把10 nm的单个分子信号放大了1万亿倍($10^6 \times 10^6$)，

令其衍射信号从水溶液的背景信号中凸显出来。

再说导演。导演，主要责任是对分子进行“泵浦”。蛋白质分子在平衡态倾向于保持其最低能量结构。“泵浦”的有效性(导演是否合格)需要至少从两个方面考虑：对蛋白分子激发的成功率和对所测量的分子样品激发的同时性。激发成功率包括收到激发信号的分子所占比例和被激发之后启动结构演化的比例。比如，用脉冲激光作为“泵浦”方法，需要调控泵浦脉冲的光子密度(即光脉冲能量)，尽可能地增加光子数，又要保证分子晶体样品不会因为光照太强而受到损伤或者单个蛋白吸收多个光子产生副产品。吸收光子之后，蛋白质分子从基态跃迁到激发态，既可以朝着预期的结构演化，也可以回到基态。为了能够增加激发成功率，还需要选择适宜的激光波长(即单个光子的能量)。激发就类似跑步比赛中的发令枪，告诉蛋白质分子开始结构演化。为了让蛋白质分子“同时”行动，就要求激发信号尽可能“同时”到达所有的蛋白质分子。从这个角度考虑，激光脉冲是最佳的“泵浦”方法。然而，激光脉冲只适用于“泵浦”感光蛋白质分子，很多重要的蛋白质分子需要其他方式激发，比如具有催化功能的各种蛋白酶需要与底物结合之后才启动结构变化。微流混合的方法能够促进蛋白酶与反应底物进行结合，启动催化反应。在启动结构变化的同时性和有效性方面，微晶比更大尺寸的晶体有显著的优势：无论是光脉冲激发，还是微流混合激发，小尺寸的晶体都便于信号深入晶体内部。

最后是专业的剪辑。如前文所述，任何“泵浦”的方法都不能保证激发100%的生物分子。这给数据分析带来了巨大的挑战，而这就是“专业的剪辑”大显身手的地方。传统晶体学数据分析已经获得长足的发展，使用成熟的软件，通过标准的流程，就可以从实验数据重构实空间的电子密度，并结合蛋白质的氨基酸序列构造原子模型。在“泵浦—探测”晶体衍射实验中，“泵浦”之后任一时刻测量的数据都是多种结构叠加形成的信号，已有的软件还不能够直接用于分析这种多态混合的数据。因此，在数据分析方面需要开

发新的方法分离出单个结构对应的信号，并整合其他实验方法获得的信息进行交叉验证。下文将结合具体的蛋白质分子体系，详细介绍“分子电影”制作的关键步骤，即如何确定主要的结构类型和所占比例，并且从混合的信号中分离出该时刻每种单一结构对应的结构信息。

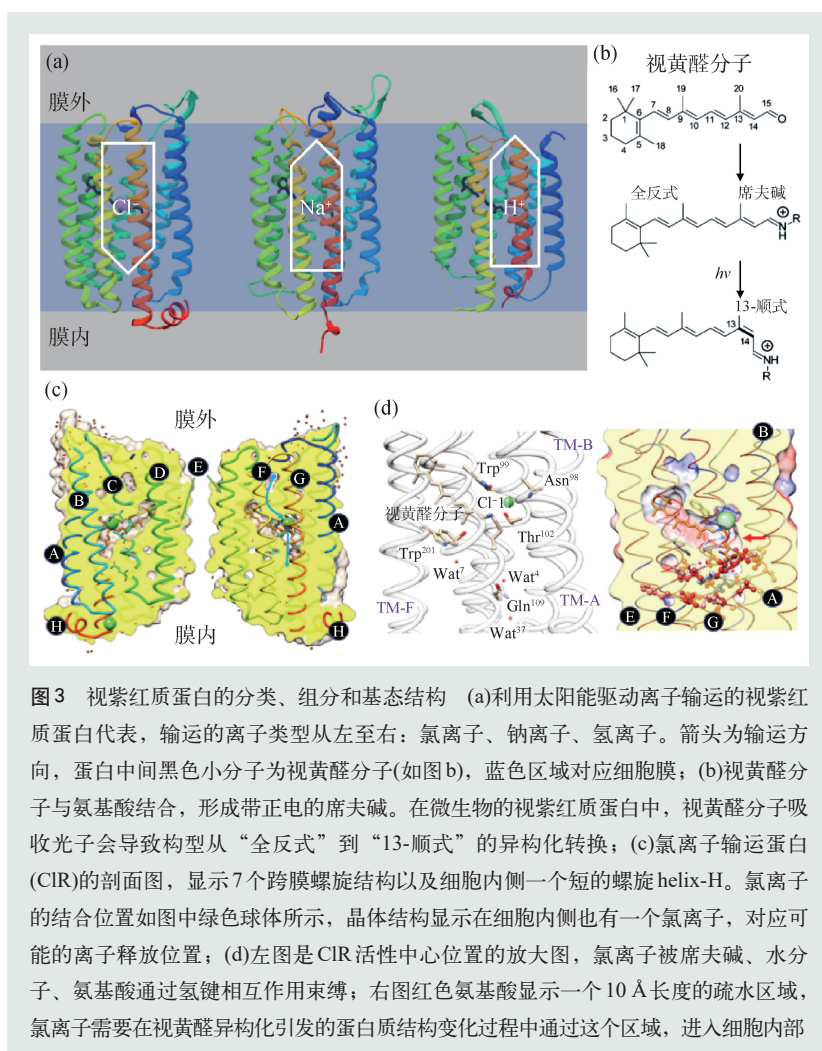
具备“摄像—导演—剪辑”三要素之后，就可以开展时间分辨的晶体衍射实验。接下来，我们基于对光驱动的氯离子输运蛋白质的动力学研究成果，介绍分子电影的拍摄和制作过程。

3 视紫红质蛋白如何利用光子能量输运离子

3.1 视紫红质蛋白是细胞的感光分子

自然界中有大量的生物能够利用太阳能驱动细胞功能，比如光合作用、昼夜节律调节、维持细胞内外离子浓度等。针对太阳能在细胞内转化的研究在能源、医疗、健康等多个领域都有重要意义。在众多的感光蛋白体系中，视紫红质蛋白(rhodopsin)是一个“宠儿”，作为模式系统得到了广泛的关注和研究(图3)^[10]。

视紫红质蛋白属于跨膜蛋白，具有7重跨膜螺旋结构，和哺乳动物细胞膜上著名的G蛋白耦合受体蛋白(GPCR)在结构上高度相似，也与视网膜上感光的视蛋白(opsin)紧密相关。视紫红质蛋白结合了一个感光的视黄醛分子(retinal)，并将其包裹在蛋白内部(图3)。视黄醛分子能够吸收光子，并且利用获得的能量完成异构化(isomerization)。由于视黄醛与视紫红质蛋白的紧密结合，视黄醛的异构化会引发蛋白质结构的变化，改变蛋白在构象空间的势能面：暗状态的结构对应的能量被抬高，推动结构朝着新的能量低点演化。这些视紫红质蛋白利用光能可以输运多种离子，包括氢离子、钠离子、氯离子等(图3(a))。其中，氢离子输运蛋白最早被发现，得到了最丰富的研究。因为是在细菌中发现的，所以被称为菌视紫红质(bacteriorhodopsin, bR)。综合多种方法的研究成果，人们已经初步勾勒出光子吸收导致的视



黄醛异构化过程, 以及在bR分子中引发的结构演化。最近, 利用XFEL开展的“泵浦—探测”实验中, 科学家揭示了光子驱动bR分子形成的多个中间态的原子分辨率结构^[11-14], 为理解氢离子运输动力学过程和分子机理提供了丰富的信息。然而, 氢离子与X射线的相互作用非常弱(氢离子周围没有电子能够与X射线电磁波相互作用), 作为被运输的“主角”, 氢离子的位置不能直接测量, 必须根据氨基酸、视黄醛、以及周围的水分子的化学状态做出推测。氯离子则具有得天独厚的优势, 因为其带有18个电子, 对X射线有很强的散射效应, 所以能够通过X射线的信号直接测定氯离子的位置信息, 这也是氯离子运输蛋白(chloride ion pumping rhodopsin, CIR)作为研究体系的优势。

CIR蛋白的结构如图3(c)所示, 视黄醛分子通

过共价键结合在第7个跨膜螺旋中间部位的赖氨酸(Lysine-235)上, 结合部位形成了一个席夫碱原子集团, 带有一个正电荷(图3(b))。在黑暗环境中, 氯离子与席夫碱、周围的氨基酸以及附近的水分子形成多个氢键, 被束缚在一个稳定的结合状态。通过时间分辨光谱的研究, 我们知道, 光照之后, 视黄醛分子吸收一个光子被激发, 全反式的碳链中C13=C14的双键翻转大约180°, 导致其构型变为“13-顺式”(图3(b))。视黄醛分子的异构化过程在百飞秒至1 ps时间内完成, 而视紫红质蛋白受此驱动的结构演化通常需要几十毫秒^[15]。在这几十毫秒的时间内, 视紫红质蛋白将氯离子运送到细胞内部, 包括跨越蛋白质中间一段长度约10 Å的疏水氨基酸区域的能量势垒(图3(d))。在这个过程中, 蛋白质的结构是如何演化的, 需要通过“泵浦—探测”实验研究才能揭晓, 其中视黄醛分子异构化之后

很短时间内发生的超快动力学变化必须用XFEL的飞秒脉冲才能捕捉。

3.2 用LCLS的X射线激光拍摄分子电影

2017—2018年, 我们与韩国延世大学 Weontae Lee 研究组合作, 利用美国斯坦福线性加速器中心(SLAC)国家实验室的XFEL设施开展了一系列的时间分辨晶体学研究。在美国自然科学基金BioXFEL科技中心研究人员和北京师范大学张文凯教授团队的帮助下, 解析出了CIR蛋白质分子在黑暗环境以及光激发之后100 ps之内四个时刻的原子分辨率结构^[16]。这是世界上首次观看到光驱动氯离子运输的“分子电影”, 详细揭示了视黄醛异构化对氯离子、蛋白质结构的影响。在这个电影中, 氯离子初期的运动与预期方向相反, 先

朝着细胞外的方向短暂移动，再向细胞内扩散。这种出乎意料的运动模式乍看违背常理，其实可以通过简单的受力分析获得合理的解释。

如前文所述，为了捕捉超快的结构变化，先进的“摄像装置”必不可少。直线加速器相干光源(Linac Coherent Light Source, LCLS)是世界上第一个硬X射线自由电子激光设施，是美国能源部资助SLAC国家实验室在其原有的直线加速器的基础上建设的，耗资大约4亿美元。2009年首次成功发射X射线激光之后，取得了一系列的科研突破，而飞秒串行晶体学也被《科学》选为2012年“十大突破”之一^[17]。LCLS每秒最多发射120个XFEL脉冲，可以将 10^{12} 个光子压缩在几十飞秒的单发脉冲之内——这是用微晶进行高时空分辨衍射的硬件基础。

3.3 飞秒激光指挥蛋白质的统一行动

飞秒激光器作为“泵浦”装置，在这个实验中充当“导演”职务。为了确定“泵浦”激光的最佳波长和功率，张文凯实验室系统地测量了CIR蛋白的吸收光谱和激发后1纳秒之内多个时刻的瞬态光谱。结合LCLS的激光仪器参数，研究团队最终将泵浦激光的波长确定在550 nm，并且估算出泵浦激光器的单个脉冲能量需要达到 mJ/mm^2 的量级。由于“泵浦”和“探测”都采用飞秒激光脉冲，保证了激发的高度同时性和测量的时间分辨率。XFEL脉冲是由2 km外的电子枪发射出来的，到达样品的时间有一定的不确定性(抖动范围大约在1000 fs量级)。为了进一步提高时间延迟的准确性，实验过程中使用“计时工具”(time-tool)来校正每次测量的时间延迟(XFEL脉冲和泵浦激光脉冲到达样品的时间差)^[18]。通过设定“泵浦—测量”的时间差，我们测量了多个重要时间点的结构信息。

由于单个晶体被激发后测量一次就被XFEL脉冲损伤，需要将大量CIR蛋白晶体依次输送到X射线焦点进行衍射测量。根据经验，一组完整的衍射信号需要累积几万颗晶体产生的衍射图谱，才能够达到原子分辨率。当然，所需衍射图

谱数量依赖多个因素，比如晶体的尺寸、晶胞的参数和对称性、XFEL脉冲的能量等。在对CIR的实验中，单个时间点的衍射信号包含2.4—4.8万张有效测量的衍射图(即能够进行指标化分析确定衍射图像中衍射斑对应的密勒指数，将衍射强度汇总到傅里叶空间)。考虑到泵浦激光的焦斑超过150 μm ，我们需要在连续两个泵浦脉冲的时间间隔，把被上一个脉冲激光照射的样品全部移出焦斑占据的空间，才能保证样品只被照射一次。这极大地增加了晶体样品的损耗，亟需在将来的实验中改进实验技术来降低对样品量的需求(比如缩小泵浦激光的焦斑)。

在2017、2018年两次(共7天)实验中，我们首先收集了光照之前的晶体衍射信号(需要指出的是，为了保证蛋白质分子处于基态，样品的制备是在暗室中进行的)；然后用“泵浦—探测”的方式，收集了光脉冲激发之后1 ps到25 ms之内多个时间点的信号，累积超过100 TB的原始数据。接下来就需要“专业的剪辑师”上场，对这些数据进行系统、细致的分析。

3.4 从海量数据到“分子电影”

蛋白分子被激发之后很短的时间内，所有被激发的分子的结构变化近似相同。这好比是百米赛跑，激光脉冲作为发令枪，在10秒钟左右，顶级选手们都接近百米终点位置。与之相对的，长时间的动力学演化，就类似马拉松比赛，参赛选手的差距会被拉开。因此，在短时间的超快动力学过程中，被激发的分子近似认为具有相同结构。尽管如此，“剪辑师”也需要确定激发的分子所占的比例。信号提取和结构确定的过程如图4所示。

首先，确定分子的基态结构。通过测量照光之前的晶体衍射信号，利用常规的晶体衍射数据分析和相位恢复算法获得“暗状态”的分子结构。得益于高品质的晶体和高亮度XFEL脉冲，CIR分子基态结构的衍射信号达到了1.65 Å，精确解析出原子分辨率的蛋白结构以及氯离子、视黄醛分子、重要水分子的位置信息。至此，我们拥

有了基态结构对应的X射线衍射强度和相位信息。

第二步, 考虑到激发之后的结构在短时间内的结构变化很小, 可以先通过基态的衍射相位信息和实验测量的激发前后衍射信号强度的差异, 计算激发前后的电子密度差异图, 确定原子位置的移动方向, 初步构造对应该时刻的中间态分子结构。一旦激光照射引发某些原子位置的变化, 这些原子附近的电子密度差异会出现成对的正负密度。比如, 在氯离子附近就观察到成对的正负电子密度, 而且负密度信号强度达到14个标准差(标准差根据电子密度差异数值计算得出), 显示

光照前后氯离子位置有显著变化(从负密度位置向正密度位置移动)。

第三步, 确定被激发分子在全部分子中所占比例。由于第二步计算实验测量的衍射强度信号差异是一个相对的数值, 光照之后还处于基态结构的分子信号被抵消, 所以不能直接得出被激发分子的比例。给定时刻的中间态的电子密度图可以按照信号线性叠加的原则, 近似表达为第二步得到的电子密度差异图与基态的电子密度图的叠加。通过调整比例系数, 观察有显著位置变化的原子附近区域叠加得到的电子密度的变化, 从而

估计出被激发的分子所占比例。比如, 在CIR蛋白中, 氯离子被激发后离开其基态位置, 如果没有其他原子迅速填充氯离子的基态位置, 氯离子基态位置附近的电子密度则为0。我们通过将电子密度差异图与基态的电子密度图进行叠加, 并逐渐增大密度差异图的系数, 直到指定区域的电子总数变为0, 估计出“泵浦”激光对CIR蛋白的激发效率。被激发的分子占比与泵浦激光的脉冲能量紧密相关, 为了避免对样品产生附加的热效应或者飞秒脉冲导致的多光子吸收的副产品, 实验中我们采用相对保守的泵浦激光能量, 因此只有大约10%的蛋白分子被激发, 进入结构演化进程。这些是真正的“演员”, 光照之后任意时刻测量的数据都是这10%“活跃”蛋白分子与另外90%“懒惰”基态分子的信号叠加。另外, 在“泵浦”之后较长的时间里, 这10%的分子将会表现出马拉松赛道上的情形, 演化速率的差异将导致更多中间状态与基态的叠加。对于这种情形的数据分析, 需要用到诸如主成分分析、奇异值分解或者机器学习等先进方法(有待后续研究)。对蛋白质分子的动力学演化过程分析可以降低分析复杂度。比如通过时间分辨的瞬态吸收光谱重构反应动力学, 得出动力学时间常数(即反应速率), 进而找到亚稳态累积的时刻。这些亚稳态的位置类似赛道上的障碍物, 导

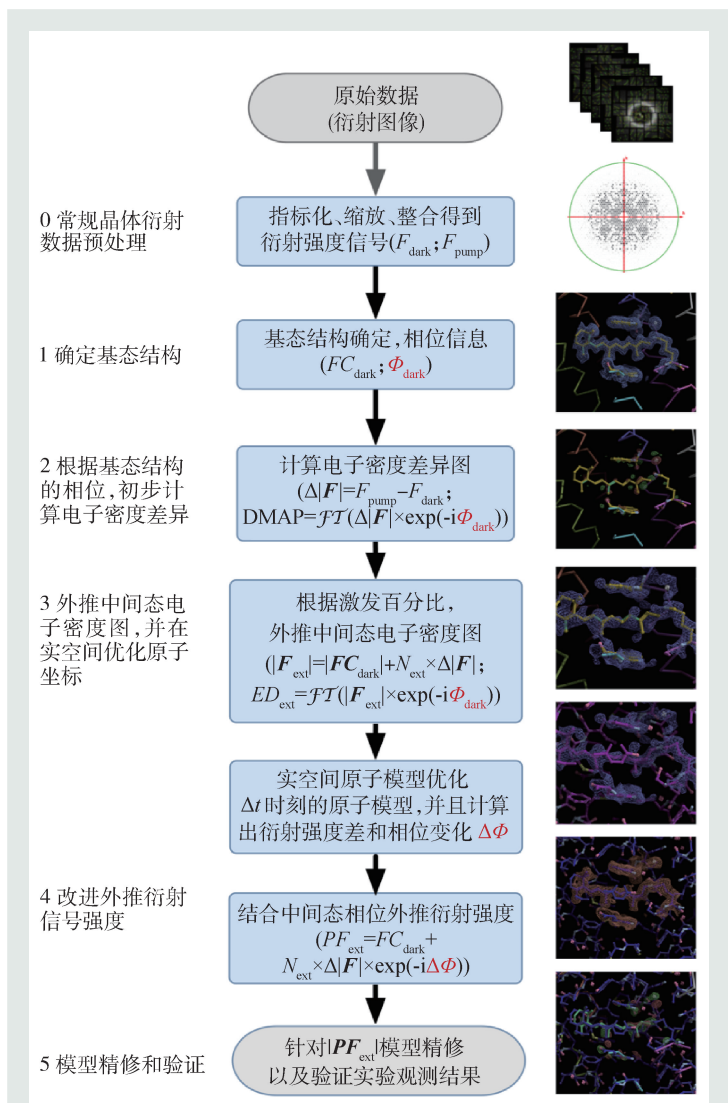


图4 从衍射数据到原子结构的分析流程。其中0、1两步是常规串行晶体衍射的分析方法, 2—5则是与“泵浦—探测”相关的结构分析方法, 需要进行迭代分析才能更准确地确定激发之后指定时刻结构的原子坐标(图片根据文献[16]修改)

致大量选手(蛋白分子)在障碍物附近聚集。用XFEL晶体学针对这些特定时刻进行测量,可以降低组分的复杂度。

第四步,根据成功激发分子的占比,计算激发后该时刻中间态的电子密度图,并以此为约束,在实空间对原子位置进行优化。至此,激发之后该时刻对应中间态结构的衍射强度和相位得到更准确描述。新的相位信息用于修正合成的衍射强度(对应该时刻中间态结构),将差异信号和基态结构的衍射信号在复平面进行向量叠加。在这个步骤之前,基态结构的相位被近似用于描述中间态。针对改进的衍射信号,再次优化该时刻的中间态结构,并最终确定原子坐标和晶体结构的精修统计数据。

第五步,根据基态和激发后给定时刻的中间态结构,计算理论电子密度的差异图,并与第二步得到的实验结果进行对比分析。如二者有显著差异,则需要再重复第四步对原子坐标进行调整。

最后,将多个结构按照时间顺序串联,就得到了具有原子分辨率的“分子电影”。

3.5 “分子电影”告诉了我们什么?

利用时间分辨的晶体衍射方法,我们获得了0—100 ps之间5个时间点的原子分辨率结构,再现了光子驱动的氯离子转运早期阶段的CIR蛋白质动力学过程(图5)。不同于转运氢离子的菌视紫红质蛋白,被转运的氯离子在这个“分子电影”中清晰地展现了光照之后独特的位置变化。按照以前的模型,视黄醛分子异构化导致带正电的席夫碱朝细胞内的方向翻转,静电相互作用应该带动氯离子朝细胞内移动。我们用XFEL拍摄的“分子电影”则展示了一个“脱离—扩散”的过程:视黄醛异构化之后,氯离子周围的氢键网络发生变化,释放了被氢键网络“束缚”的氯离子。经过几十皮秒的扩散运动,氯离子被一个高度保守的苏氨酸(Thr¹⁰²)吸引,开始朝细胞内移动。这个看似违背常理的现象,是由视黄醛分子超快的异构化过程与氯离子移动时间尺度的分离导致的。根据时间分辨吸收光谱的实验数据以及

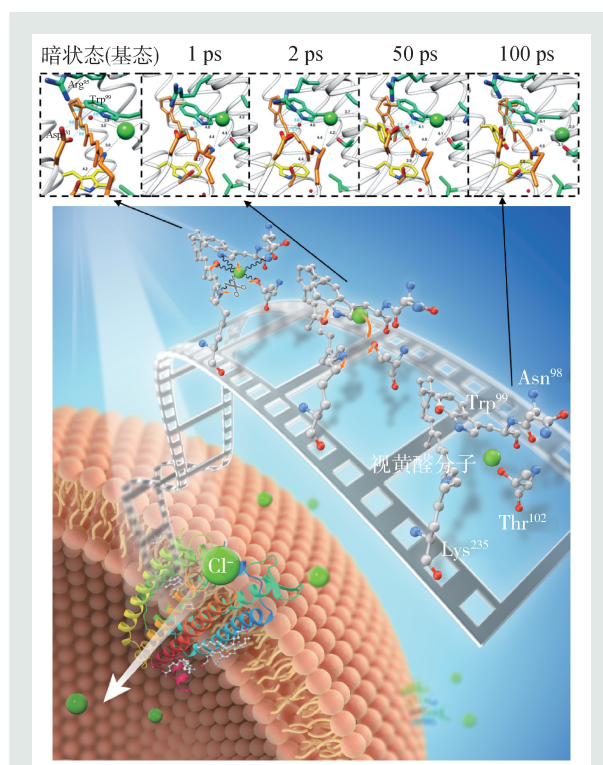


图5 光照驱动的离子转运视紫红质蛋白结构的演化。视黄醛在1 ps之内异构化,释放氯离子。周围的氨基酸和水分子重新排布,适应“13-顺式”的视黄醛构型。在100 ps左右,氯离子被苏氨酸102吸引,开始朝细胞内运动(具体运动参看正文)

对菌视紫红质蛋白(bR)的视黄醛异构化的研究结果,异构化在几百飞秒之内就完成了C13=C14双键的翻转。也就是说,带正电的席夫碱在光照之后,迅速远离氯离子,导致二者之间的库仑力减弱。与此同时,蛋白质还未做出响应,结构没有发生变化,氯离子以及附近的氨基酸和水分子还处于基态位置。这就导致氯离子受力的均衡被打破,席夫碱之外的原子产生的合力“拉动”氯离子朝远离席夫碱的方向移动(图6)。视黄醛两种构型(全反式和13-顺式)对氯离子约束力的差异在分子动力学模拟中得到了验证,为实验观看到的“分子电影”提供了支撑。

除了氯离子在100 ps之内的有趣运动之外,虽然视黄醛分子在1 ps时刻已经显示出“13-顺式”的构型,但是分子电影显示这个异构化还处于某种震荡状态,需要弛豫到100 ps才稳定在“13-顺式”的构型。氯离子和视黄醛分子周围的

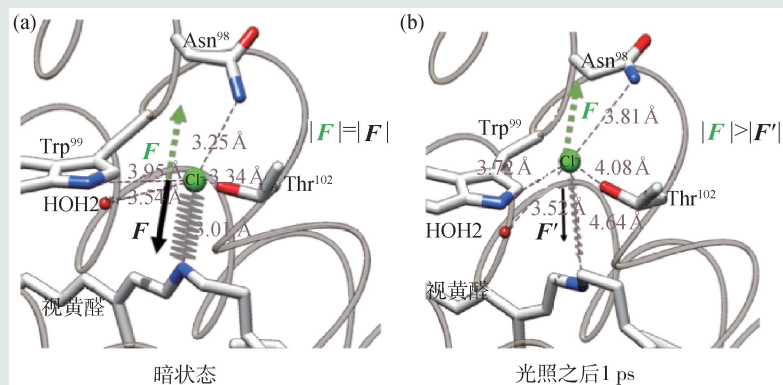


图6 氯离子的受力分析与运动方向 (a)在光照前,氯离子被氢键束缚,稳定地结合在席夫碱附近。席夫碱对氯离子的吸引力(黑色)与其他原子对氯离子的合力(绿色)达到平衡;(b)光照之后,席夫碱在1 ps之内翻转到远离氯离子的位置,导致二者之间的静电吸引力减弱,而其他原子的位置来不及发生显著变化,导致合力方向指向细胞外侧,氯离子被“拉”向胞外

氨基酸也随视黄醛异构化产生了相应的结构变化,以适应新的构型。比如第98位的天冬酰胺(Asn^{98})在基态与氯离子形成氢键,其运动模式与氯离子密切相关,光照之后1—2 ps, Asn^{98} 先向胞外方向偏移,待到50—100 ps又随着氯离子向胞内方向移动。不同于“自由”的氯离子,氨基酸还连接在蛋白质的主链上,所以在100 ps的时候, Asn^{98} 与氯离子的氢键断开。两个色氨酸(Trp^{99} 和 Trp^{201})自始至终为视黄醛分子提供支撑,将视黄醛包裹在CIR蛋白内部的同时,色氨酸侧链的位置移动展示了其如何配合视黄醛构型变化。更有意思的是,氯离子附近的一个水分子在1—50 ps时刻,与氯离子保持3.5—3.8 Å的距离,维持二者的氢键相互作用。然而,在100 ps时,氯离子被 Thr^{102} 吸引,这个水分子与氯离子的氢键断开(距离超过5.7 Å)。高度协调的原子位置的变化显示氯离子运输的动力学过程有特定的路径,很可能是蛋白质分子长期演化过程中优化的结果。

光照之后100 ps期间的原子位置移动主要集中在视黄醛和氯离子周边的局部形变,而氯离子进入细胞还需要CIR分子更大范围的结构变化。由于氯离子本身带有负电荷,在其穿越CIR分子内部极度疏水的区域时,必然遭遇较高的能量势垒。根据CIR分子结构和已经看到的部分“分子电影”,我们推测有两种可能性:(1)蛋白质的七

重跨膜螺旋有大尺度移动,导致细胞膜内侧的结构打开更大的通道,允许水分子进入,帮助氯离子扩散到细胞内;(2)氯离子与氢离子结合,形成氯化氢分子(HCl),组团进入细胞内部,然后再水解。第一种可能性可以参照菌视紫红质蛋白输运氢离子的研究,在氢离子进入bR分子的时间点,第六个跨膜螺旋向外倾斜,打开了一个通道允许水分子深入蛋白内部。第二个猜测也基于一些bR输运氢离子的模型,涉及到席夫碱质子化和去质子化的转变。这些猜想为进一步利用

XFEL研究氯离子运输过程提供了思路,激励我们继续录制“分子电影”,解密蛋白质分子如何利用太阳光输运离子,实现能量转化的分子机理。这100 ps的“分子电影”只是光子驱动蛋白动力学的预告片,完整版的“分子电影”有待更深入地分析和更多的实验研究。

4 结语

XFEL的发展推动了诸多领域的科研工作,“损伤前探测”的实验模式突破了辐射剂量对传统成像方法的局限。利用XFEL飞秒脉冲从微米甚至纳米晶体获得原子分辨率结构的方法,不仅解析了很多重要的蛋白质分子的静态结构,而且受益于小晶体易于激发(相对增大的表面积)而促进了时间分辨的晶体学研究(注意,同步辐射设施也可以开展时间分辨晶体衍射研究,但是时空分辨率受到限制)。本文以光感蛋白作为示例,介绍了“分子电影”的录制过程。这个基本流程和思路也适用于以其他方法“泵浦”的分子体系,比如电压、磁场、温度、酸碱度、化学反应的底物分子微流混合等。其中一些方法需要克服“泵浦”的均一性和同时性的难题(即需要一个好导演)才能更好地进行“分子电影”的拍摄。我们也认识到,虽然蛋白质分子在晶体中排列相对疏松,蛋白分子间也有很多的水分子维持水环境,但晶体

堆叠会影响较大尺度的结构变化,因此,分子电影的录制,需要进一步发展单分子的高分辨成像方法。冷冻电子显微镜在单分子成像领域获得了巨大的进步,如果能够解决“摄像—导演—剪辑”三要素的问题,则有望成为“分子电影”的另一种拍摄方案^[19, 20]。AlphaFold 2 在结构预测领域的突破^[21],也将推动结构生物学的研究重点从对单一静态结构的确定转向对动态过程

的研究。

最后, XFEL 是造价昂贵的大型设施,需要十亿甚至百亿元的建设经费。尽管如此,我国从长远着眼,提前进行了部署,在上海同步辐射设施(即上海光源)的基础上,于2019年开工建设上海高重频自由电子激光设施(1秒最多发射100万发 XFEL 脉冲)。在不久的将来,我们将可以使用自己的 XFEL 光源开展尖端科学研究。

参考文献

- [1] Madey J M J. *J. Appl. Phys.*, 1971, 42 (5): 1906
- [2] Emma P, Akre R, Arthur J *et al.* *Nat. Photonics*, 2010, 4 (9): 641
- [3] Marklund M, Shukla P K. *Rev. Mod. Phys.*, 2006, 78(2): 591
- [4] Chapman H N, Fromme P, Barty A *et al.* *Nature*, 2011, 470(7332): 73
- [5] Neutze R, Wouts R, van der Spoel D *et al.* *Nature*, 2000, 406(6797): 752
- [6] Caffrey M. *Acta Crystallogr. Sect. F: Structural Biol. Commun.*, 2015, 71: 3
- [7] Liu W, Wacker D, Gati C *et al.* *Science*, 2013, 342 (6165): 1521
- [8] Xiang J, Chun E, Liu C *et al.* *Trends Pharmacol. Sci.*, 2016, 37(12): 1055
- [9] Colletier J P, Schirò G, Weik M. *X-Ray Free Electron Lasers*. Springer International Publishing, 2018
- [10] Ernst O P, Lodowski D T, Elstner M *et al.* *Chem. Rev.*, 2014, 114(1): 126
- [11] Nogly P, Weinert T, James D *et al.* *Science*, 2018, 361 (6398): eaat0094
- [12] Nass Kovacs G, Colletier J P, Grünbein M L *et al.* *Nat. Commun.*, 2019, 10 (1): 3177
- [13] Weinert T, Skopintsev P, James D *et al.* *Science*, 2019, 364(6448): 61
- [14] Wickstrand C, Nogly P, Nango E *et al.* *Annu. Rev. Biochem.*, 2019, 88(1): 59
- [15] Inoue K, Koua F H M, Kato Y *et al.* *J. Phys. Chem. B*, 2014, 118(38): 11190
- [16] Yun J H, Li X, Yue J *et al.* *Proc. Natl. Acad. Sci. U. S. A.*, 2021, 118(13): e2020486118
- [17] Science/AAAS. Special Issue: Breakthrough of the Year, 2012. <http://www.sciencemag.org/site/special/btoy2012/>
- [18] Harmand M, Coffee R, Bionta M R *et al.* *Nat. Photonics*, 2013, 7(3): 215
- [19] Frank J. J. *Struct. Biol.*, 2017, 200(3): 303
- [20] Mäeots M E, Lee B, Nans A *et al.* *Nat. Commun.*, 2020, 11(1): 3465
- [21] Jumper J, Evans R, Pritzel A *et al.* *Nature*, 2021, 596: 583

更正声明

《物理学名词》(第三版)更正

由科学出版社2019年出版的《物理学名词》(第三版)面世以来,得到了广大物理学领域工作者的关心和支持。最近,我们收到河北师范大学杨大卫教授反馈:本书347页,序号09.0929的名词“介子”的注释中写“又称重电子(heavy electron)”是错误的,介子不应称为重电子。经过委员会部分委员讨论,认为将介子又称为重电子确实不妥,有误导读者之嫌,应当删去。特此更正。

此外,原物理名词审定委员会的刘寄星副主任委员请中国科学院半导体研究所姬扬研究员撰写了一篇短文(本期《物理》第639页),介绍介子与重电子的概念。

全国科学技术名词审定委员会第三届物理学名词审定委员会

2021年7月