

文章编号: 1000-6281(2025)01-0124-12

AI 时代中的电子显微学研究: 严峻挑战、无穷机遇与壮阔前景

刘 铮^{1*}, 沈庆涛², 隋森芳^{1,2,3}

(1. 南方科技大学冷冻电镜中心, 广东 深圳 518055; 2. 南方科技大学生命科学院, 广东 深圳 518055; 3. 清华大学生命科学学院, 北京 100084)

摘要 近十年来结构生物学的发展突飞猛进, 取得的标志性突破有两个, 一是以冷冻电镜为代表的结构解析技术方向上的突破, 二是以 AlphaFold 算法为代表的结构预测模型上的突破。2024 年 5 月 Google DeepMind 更新了其生物结构预测工具, 最新版本的 AlphaFold 3 拥有可以预测几乎所有分子类型的蛋白质复合体结构的能力, 在药物互作预测方面也实现了很高的准确性。AlphaFold 3 的发布为结构生物学研究带来巨大的变革, 展现了 AI 技术的巨大潜力, 也点燃了大众对生命科学和医学研究的热情与想象。与此同时, 电子显微学的研究并未停下脚步, 新技术、新方法层出不穷, 在解析全新蛋白结构、超大超复杂复合体结构、动态结构、原位结构, 以及更大尺度的细胞、组织、器官样品的研究中, 电子显微学依旧有着不可替代的优势。当前有观点认为结构预测模型甚至可以替代以 X 射线晶体学和电子显微学为代表的传统实验科学, 通过计算便能完成生物结构解析, 但是这种观点是片面的。事实上, 未来的结构生物学研究, 必将是一个整合实验科学与 AI 技术, 从单个蛋白或复合体的结构全面拓展到多蛋白复杂体系、细胞内原位、以及超越微观尺度进入到介观和宏观尺度等方面的研究。

关键词 电子显微学; 高分辨结构; AlphaFold; 结构模型预测

中图分类号: Q6; Q71; Q31; Q51; **文献标识码**: A

Doi: 10.3969/j.issn.1000-6281.2025.01.14

精准的生物大分子(包括蛋白质、核酸、脂肪、糖类等)、生物大分子复合体(例如:核糖体、蛋白酶体、光合反应中心等)以及细胞器(例如:线粒体、叶绿体等)的精准结构对于我们深入理解细胞的正常生理功能至关重要。阐明由于基因突变、病毒与细菌感染、环境变化等引起的结构改变,进而导致细胞的功能异常,对我们深入理解疾病的发生发展机制,从而合理设计治疗方法,并开发精准靶向药物进行干预,更是有着非常重要的意义。电子显微学是一种重要的结构生物学研究方法,它与另外两种技术,X射线晶体学和核磁共振技术一起构成了高分辨率结构生物学研究的基础,是解析生物大分子结构并揭示其功能的重要研究手段之一。

1 X 射线晶体学主导的结构生物学时代

21 世纪之前的结构生物学研究以 X 射线晶体学为主导。1895 年德国物理学家 Wilhelm Röntgen

发现了 X 射线,1901 年因此项伟大发现被授予第一个诺贝尔物理学奖。随后,另一位德国物理学家 Max von Laue 发现了晶体中 X 射线的衍射现象,并由此获得了 1914 年的诺贝尔物理学奖。之后不久,William Bragg 和 Lawrence Bragg 父子推算出衍射的数学规则,提供了一种解释 X 射线产生的衍射图案的方法,从而揭示了晶体中原子的排列,他们还设计出 X 射线光谱仪,父子共同获得 1915 年诺贝尔物理学奖。起初,Bragg 的方法主要被应用于简单无机物和有机物,如盐、苯和糖类分子,解析出它们结构。Dorothy Hodgkin 因解析青霉素、维生素 B12 等结构获得了 1964 年诺贝尔化学奖。

1953 年,X 射线晶体学获得了巨大突破,它被用于解析一个极其重要的生物大分子—DNA 的双螺旋结构,James Watson、Francis Crick 和 Maurice Wilkins 为此获得了 1962 年诺贝尔生理学或医学奖。几乎是相同时间里,John Kendrew 和 Max Perutz 成功地培育出肌红蛋白的晶体并解析出第一个蛋

收稿日期:2024-10-23;修订日期:2024-11-08

基金项目:国家自然科学基金专项资助项目(No. 32241028)。

* 通讯作者:刘铮(1970—),男(汉族),天津人,教授。E-mail: liuz3@sustech.edu.cn

白质的三维模型,两位科学家由此获得了1962年的诺贝尔化学奖。Johann Deisenhofer、Robert Huber 和 Hartmut Michel 成功地解析了细菌光合作用反应中心的三维结构,阐明了光合作用的分子机制而共同荣获了1988年诺贝尔化学奖,这是首次解析出的膜蛋白高分辨三维结构。Roderick MacKinnon 解析了细胞膜钾离子通道的结构而获得了2003年诺贝尔化学奖,这是首个获得的离子通道蛋白高分辨三维结构。Roger Kornberg 解析了RNA聚合酶,一个由12种蛋白质组成的复合体结构,由此获得了2006年诺贝尔化学奖。Venkatraman Ramakrishnan、Thomas Steitz 和 Ada Yonath 应用X射线晶体学研究核糖体这一蛋白质合成机器的结构和功能,三位科学家共同获得2009年诺贝尔化学奖。X射线晶体学解析了众多生物大分子的高分辨率结构,建立了结构与功能之间的有机联系,促进了生命科学的迅猛发展。

2 结构生物学进入到电子显微学时代

1930年代,德国科学家 Ernst Ruska 和 Max Knoll 研制出世界上第一台透射式电子显微镜,半个世纪后,1986年 Ruska 获得诺贝尔物理学奖。由于电子波长远远小于可见光波长,依据 Ernst Abbe 在1873年提出的分辨率定理^[1],电镜的分辨率可以达到亚埃米水平,显著优于光学显微镜。这之后,越来越多的科学家将电镜显微技术引入各自的研究领域。1942年,比利时裔美国生物学家 Albert Claude 将电镜应用到细胞研究上,他用电镜观察动物细胞,首次在细胞内看到了一些微细结构,发现了线粒体和内质网;他还和同事一起研究了线粒体的功能,并把线粒体定位为细胞的动力工厂。因为对细胞构造的研究,Albert Claude 与 Christian de Duve、George Palade 共同获得1974年诺贝尔生理学或医学奖。英国的分子生物学家 Aaron Klug 首次将X射线晶体学与电子显微技术相结合创立了电子晶体学,电镜照片的放大倍数和分辨率虽然远远超过光学显微镜,但所得图像是二维的照片,Klug 推测可以通过对晶体的二维照片进行傅里叶变换,类似于对晶体进行衍射分析,从而获得图像在倒易空间的晶格信息,通过从许多不同角度捕捉生物大分子的电镜图像,继而对每一个图像进行衍射分析,最终得到这一生物大分子的精确结构^[2]。1982年,Klug 因“发展出电子晶体学方法和解析重要的核酸-蛋白质复合体结构”获得了当年的诺贝尔化学奖。

冷冻电镜单颗粒技术是继电子晶体学之后出现的又一重要电子显微学方法。冷冻电镜单颗粒技术不需要对生物大分子进行结晶,而是将处于合适生理条件下的含水生物样品迅速冷冻至非晶态冰中,再用透射电镜观察处于冷冻状态下的样品,结合计算机图像处理解析生物大分子结构。近年来,伴随着冷冻电镜硬件设备的日趋完善、直接检测电子相机的问世、以及新的重构算法的飞速进步,越来越多的蛋白质、大分子复合物的结构通过单颗粒技术获得了近原子分辨率的结构,自2013年开始,爆发了一场冷冻电镜结构解析分辨率提升的革命。由于冷冻电镜单颗粒技术具有需要样品量少、无需结晶以及无需额外提供相位信息、可解析生物大分子复合物等优点,传统X射线晶体学长期无法解决的许多重要大型复合物及膜蛋白的(近)原子分辨率结构被迅速解析。应用冷冻电镜单颗粒技术,能获得的最高分辨率已经达到1.2 Å,足以清晰展示蛋白质组成中一个氨基酸的每一个原子的精确三维空间定位,甚至还可以分辨出蛋白质周围水分子中的两个氢原子和一个氧原子^[3-4]。瑞士科学家 Jacques Dubochet、美国科学家 Joachim Frank 和英国科学家 Richard Henderson 因“开发用于溶液中生物分子高分辨率结构测定的冷冻电镜技术”而获得2017年诺贝尔化学奖。David Julius 利用辣椒素来识别皮肤神经末梢上对热做出反应的感受器,克隆出一个能够使细胞对温度敏感的基因,该基因编码一种新的离子通道蛋白 TRPV1,应用冷冻电镜单颗粒技术解析了 TRPV1 的高分辨三维结构;Ardem Patapoutian 识别出对压力敏感的离子通道 Piezo,同样是应用冷冻电镜单颗粒技术解析了 Piezo 的高分辨三维结构,Julius 和 Patapoutian 共同获得了2022年诺贝尔生理学或医学奖。

冷冻电镜单颗粒技术是目前在结构生物学领域应用最广泛的技术方法。近十年来,电子显微学,尤其是冷冻电镜技术已经成为了结构生物学的主流。在2017年 Science^[5] 和2020年 Nature^[6] 的年度十大科学突破评选中,冷冻电镜因把生命科学推进到原子分辨率水平而连续当选。冷冻电镜技术的跨越式发展不仅极大促进了生命科学的基础研究,而且直接改变了生物医药领域的研究范式,成为生命科学与医药领域中最精确和最有成效的前沿领域。在药物研发方面,多个跨国公司已经将冷冻电镜技术用于药物发现。目前已经有利用冷冻电镜基于结构研发的药物进入临床试验阶段,如葛

兰素史克公司针对寄生虫关键酶开发的黑热病新型抑制剂^[7]。相信冷冻电镜技术将持续推动药物研发进程,为重大疾病的新药开发和人类健康改善作出更大贡献。

3 AI, 结构生物学下一个时代?

蛋白质一级结构(primary structure)是指其氨基酸残基组成的线性序列,蛋白质的一级结构由 20 种标准氨基酸以不同的顺序组成,是理解蛋白质完整结构和功能的必要基础。自 1953 年英国化学家 Frederick Sanger 首次测定出胰岛素氨基酸序列后,数以万计的蛋白质的一级结构不断被测序,对研究各种蛋白质的生物、生理功能奠定了基础。Sanger 因建立了测定蛋白质和核酸的序列测定技术做出重大贡献,分别于 1958 年和 1980 年两度获得诺贝尔化学奖。蛋白质二级结构(secondary structure)是指氨基酸链接形成肽链,肽链不同区段通过自身相互作用,形成局部有规则的构象,包括 α -螺旋、 β -折叠等,这些二级结构的形成由氢键和范德华力等相互作用力所驱动,是蛋白质结构的构象单元。蛋白质三级结构(tertiary structure)指多肽链在二级结构基础上,通过侧链基团相互作用进一步卷曲折叠而形成整体立体构象,蛋白质的三级结构决定了其特定的生物学功能,例如:酶的催化活性、激素的识别和信号传导等。蛋白质的四级结构(quaternary structure)是指多个蛋白质亚基组合形成复合体的结构。

20 世纪 60 年代,科学家在研究核糖核酸酶时发现蛋白质的功能与其三级结构密切相关,而特定三级结构是以氨基酸顺序为基础的,即以一级结构为基础。当用尿素和 β -巯基乙醇处理核糖核酸酶溶液,分别破坏了次级键和二硫键,使其二、三级结构遭到破坏,但一级结构不受影响,此时核糖核酸酶呈非折叠的、松散状态,失去了酶活性。分子中的 4 对二硫键被 β -巯基乙醇还原成-SH 后,若要再形成 4 对二硫键,从理论上推算,有 105 种不同配对方式,唯有与天然核糖核酸酶完全相同的配对方式才能重新呈现酶活性。实验发现,当用透析方法去除尿素和 β -巯基乙醇后,松散的多肽链按照其特定的氨基酸序列重新卷曲折叠成天然酶的空间构象,4 对二硫键也全部正确配对,此时酶活性又逐渐恢复至原来水平。1957 年发表的这一经典实验证明空间结构遭破坏的核糖核酸酶,只要其一级结构未被破坏就可能恢复到原来的三级结构,且催化活性依

然存在^[8]。这是蛋白质的一级结构决定空间结构的一个无可争辩的例证。1972 年,美国科学家 Christian Anfinsen 因“对核糖核酸酶的研究,特别是对其氨基酸序列与生物活性构象之间的联系的研究”获得诺贝尔化学奖。

既然蛋白质的一级序列决定了它们的三维结构,那么科学家推测可以不依赖结构生物学的解析便能计算出蛋白质的结构。1965 年 Joseph Kraut 在 Annual Review of Biochemistry 发表综述文章,提出如果知道了蛋白质肽链的氨基酸序列,那么通过遵循原子和分子相互作用的简单规则,就可以计算出蛋白质肽链将如何折叠形成特定的空间结构^[9]。Anfinsen 在 1972 年的诺贝尔奖演讲中重复了这一主张:“我们对序列和三维结构之间相关性的大量数据积累,加上多肽链折叠的能量学理论的日益成熟,预测蛋白质构象的想法正逐渐变为现实”。这在当时是一个非常具有远见的推测,如果可以用蛋白质折叠的规则对计算机进行编程,输入氨基酸序列,那么结构可能在几天而不是几年内得到解决,为昂贵和耗时的结构生物学实验提供一个替代方案。

然而在当时条件下,这一超前的观点是无法实现的。其后近半个世纪的时间里,结构生物学家通过 X 射线晶体学以及后来的电子显微学方法解析出一个又一个蛋白质结构,为结构预测的大厦不断添砖加瓦。2021 年 Google DeepMind 公司的 AlphaFold 2 横空出世^[10],在蛋白质结构预测领域引发了一场革命,短短几年间,AlphaFold 2 在蛋白质结模型预测、蛋白分子相互作用和药物设计领域得到了广泛应用^[11-13]。

2024 年 5 月,Google DeepMind 推出了最新换代产品 AlphaFold 3,该模型采用了大幅更新的基于扩散的架构,能够预测包括蛋白质、核酸、小分子、离子和修饰残基在内的复合体的结构^[14]。新的 AlphaFold 3 模型相比 AlphaFold 2 和其他的结构预测工具大幅提高了预测的准确度:与最先进的对接工具相比,蛋白质-配体相互作用的准确性大大提高;与核酸特异性预测器相比,蛋白质-核酸相互作用的预测准确性显著提高;与 AlphaFold-Multimer v. 2.3 相比,抗体-抗原预测准确性也有大幅提升。这些结果表明通过深度学习框架,完全可以预测生物分子相互作用的高精确度结构,人们惊呼结构生物学跑步进入了 AI 时代。

2024 年 10 月 9 日,瑞典皇家科学院宣布,将

2024 年诺贝尔化学奖的一半授予给美国西雅图华盛顿大学教授 David Baker,以表彰他用计算方法来设计蛋白质结构做出的杰出贡献;另一半则共同授予 Google DeepMind 的 Demis Hassabis 和 John M. Jumper 以表彰他们在蛋白质结构预测方面做出的杰出贡献。三位科学家都致力于 AI 蛋白质结构研究,至此,蛋白质结构预测可以说摘取了最耀眼的科学桂冠。

4 AI 对电子显微学的挑战

自 AlphaFold 2 问世开始就有一种极端的观点,称人工智能系统的问世让结构生物学家面临下岗窘境。的确,以往结构生物学家需要花费几个月甚至几年时间、经过无数次的反复试错、消耗大量的仪器设备机时才能完成的结构解析,到了 AI 时代,可以摆脱实验手段,通过使用计算机在很短时间就可以获得一个结构预测模型。DeepMind 团队还推出了 AlphaFold Server,用户可以免费访问,只需在线提交一个任务,就可以获得结构模型,而且准确度非常高。

如果是从单纯人工智能和机器学习角度,或者从结构生物学外行的角度看,的确会有一种 AlphaFold 似乎已经把所有的蛋白结构都解析完了的错觉。但实际上,AlphaFold 所做的工作只是一个开始,这之后还有很多的问题需要解决,这些问题的解决依赖于冷冻电镜技术,也给电子显微学研究提供了诸多的机遇和长远的发展前景。

5 AI 时代中电子显微学的机遇与前景

结构生物学一直以来是高度依赖实验为基础的学科,实验手段测定的数据因此成为结构生物学领域的金标准。而结合了人工智能和机器学习的结构模型预测却是高度依赖算法和算力,其预测出的结构准确性还有待于时间检验。此外,在诸多领域的结构生物学研究中,结构模型预测仍然存在很大局限,尚不能完全替代以 X 射线晶体学和电子显微学为代表的实验科学。

5.1 低同源序列蛋白质结构

AlphaFold 可以预测蛋白质的结构,但人工智能预测是基于对已有的蛋白质序列与结构数据库比对学习,寻找与其同源较高的蛋白质结构,这些同源蛋白跟目标蛋白在序列上可能稍有差异,但结构却相对近似。AlphaFold 可以从同源蛋白中预测出

目标蛋白氨基酸之间的相互作用强度图,然后利用神经网络以及蛋白质结构固有的物理约束从氨基酸之间的相互作用强度生成目标蛋白质原子的三维坐标。因此目标蛋白与同源蛋白在序列上相似程度越高,结构预测的精度也就越高。然而,面对一个与已知结构序列同源性相比较差的蛋白质做结构预测时,AlphaFold 的预测精准度就会大打折扣。即便对于序列全同的蛋白质,由于 AlphaFold 过分依赖蛋白质序列比对,对于单点或多点突变造成的结构变化乃至功能缺失,AlphaFold 同样无法做到精准预测,无从提供合理解释。

5.2 超大、超复杂复合体、复杂病毒的高分辨结构研究

很多蛋白质发挥其生物学功能需要由多个蛋白质亚基形成复杂的四级结构,AlphaFold 2 只能预测单一蛋白质的结构,AlphaFold 3 已经具备预测蛋白复合体结构的能力,然而,面对分子量超大、结构复杂的复合体结构与复杂病毒结构,结构模型预测有时候依旧束手无策。

丙酮酸脱氢酶复合体是定位于线粒体基质的超大分子量多酶复合物,负责催化糖酵解终产物丙酮酸的氧化脱羧,生成三羧酸循环的初始底物乙酰辅酶 A,承担着连接两大代谢途径的枢纽功能,多年来一直是能量代谢领域内的研究热点之一。丙酮酸脱氢酶复合体主要由三种蛋白酶单元多聚构成:E1(丙酮酸/酮戊二酸脱氢酶)、E2(转乙酰基转移酶/转琥珀酰基转移酶)、E3(二氢硫辛酰胺/二氢硫辛酰胺脱氢酶),这些蛋白酶单元结构已经被分别解析。然而整个丙酮酸脱氢酶复合体分子量近 10 Mda,由于其庞大的尺寸、组分复杂、构象异质和高度柔性,复合体的四级结构一直难以获得,酶复合体的核心亚基构成比例和外周亚基分布规律等具体结构信息也始终存在争议,这些对 AlphaFold 而言,也是难以解决的问题。科学家通过提取一种高稳定性的嗜热梭菌的丙酮酸脱氢酶复合体,应用冷冻电镜解析了整个复合体的结构,包括 1 个 E2 的 60 聚体,4 个 E3 的三聚体、20 个 E1 的四聚体和 12 个 E3 的二聚体^[15]。对这样一种超大复合体,AlphaFold3 也很难预测出其精准组装信息。

叶绿体是光合作用的主要场所,其正常运作依赖约 2 000~3 000 种蛋白质,而大多数蛋白质由细胞核基因编码,在细胞质中的核糖体上合成出非活性前体蛋白,前体蛋白转运进入叶绿体基质内,进一步剪切加工后送达其发挥特定生物学功能的目

的微区。叶绿体作为植物及藻类细胞中含有内外双层被膜的细胞器,前体蛋白在转运肽的引导下,通过跨越叶绿体外被膜转运子(translocon in the outer envelope membrane of chloroplast, TOC)和内被膜转运子(translocon in the inner envelope membrane of chloroplast, TIC),在二者的协同作用下前体蛋白得以跨越双层被膜进入叶绿体。TOC 复合体包含有 Toc120、Toc75 和 Toc34 多种蛋白亚基,而 TIC 复合体包含有 Tic20、Tic12、Tic13、Tic35 和 YlmG 多种组分,TOC-TIC 再进一步形成稳定的超复合体共同发挥转运作用。研究人员应用冷冻电镜解析了莱茵衣藻 TOC-TIC 超复合体的高分辨率结构,总共 14 个不同的蛋白亚基组成,除了以上的保守组成蛋白之外,还发现了复合体中跨越内外两层被膜和膜间隙的 Tic214,位于膜间隙区的 Tic100 和 Tic56 蛋白,还有一些仅在莱茵衣藻中发现的特有组分,如 Toc52、Toc39 和 Toc10^[16-17]。对这样一种超级复杂复合体,AlphaFold 3 也是很难预测出精准的结构。

光能吸收是光合作用的第一步。不同的光合

生物为适应不同的光环境,进化出了各自独特的捕光系统。藻胆体是蓝藻和红藻中的主要捕光复合物,是一个超大蛋白质机器,由藻胆蛋白和连接蛋白构成,在这些蛋白上共价结合着色素分子。光能被色素分子吸收,并进一步在藻胆体内传递,最后到达光系统中心进行光-化学能量转化。应用单颗粒冷冻电镜技术,科学家解析了盐泽红藻藻胆体的三维结构,分子量高达 14.7 Mda,分辨率达到了 2.8 Å^[18]。通过结构解析,研究人员发现盐泽红藻藻胆体含有 706 个蛋白质亚基,包括 528 个藻红蛋白、72 个藻蓝蛋白、46 个别藻蓝蛋白和 60 个接头蛋白。此外,还解析得到了 1 598 个色素分子的结构,包括 1 430 个藻红胆素分子、48 个藻尿胆素分子和 120 个藻蓝胆素分子(图 1)。预测如此超级复杂的蛋白复合体结构、以及复杂的蛋白色素分子相互作用对 AlphaFold 3 也是很难完成的。

病毒作为一种非细胞型生物,其个体微小、结构简单,通常只含有一种核酸(DNA 或 RNA),编码数种至数百种蛋白质;完整的病毒粒子包括内部的

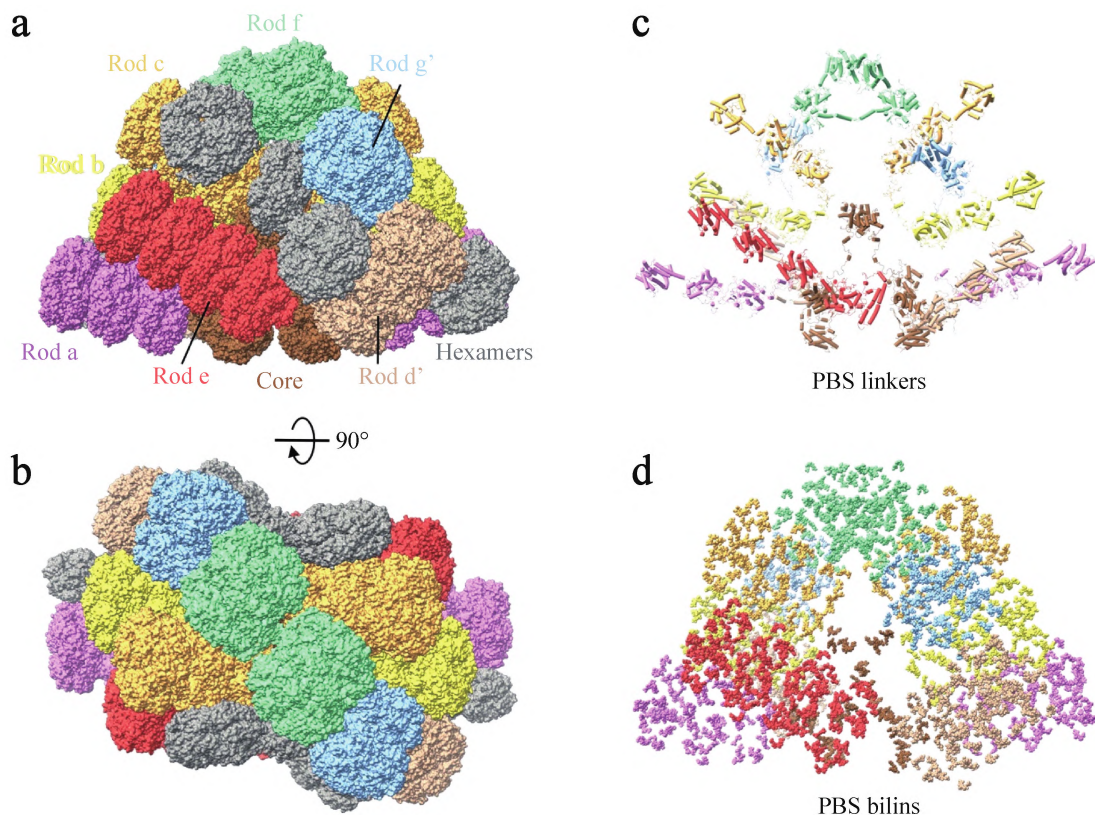


图 1 红藻 *P. purpureum* 藻胆体的整体结构。a, b. 藻胆体结构的表面展示图。其中杆以不同颜色展示,核心以棕色展示;c. 藻胆体中连接蛋白的结构;d. 藻胆体中的色素分布。

Fig. 1 Overall architecture of the PBS from *P. purpureum*. a, b. Overall structure of the PBS shown in surface representation. The rods are shown in different colours, and the core is coloured brown; c. Structure of the PBS linker proteins; d. Arrangement of the PBS bilins.

基因组和衣壳蛋白 (capsid),一些复杂的病毒外边还有由脂质和糖蛋白构成被膜 (envelop)。在一项2024年9月发表的最新研究中,科学家应用AlphaFold预测了源自4 463种感染真核生物病毒的67 715种蛋白质的3D结构,结果表明62%的病毒蛋白质在3D结构上是完全不同的,而且在AlphaFold数据库中缺乏同源蛋白,只有38%可与之前已知结构的蛋白相匹配^[19],显示出在AI时代,结构生物学依旧有很多的未知结构蛋白需要解析。这一研究还只是预测病毒蛋白的单一结构,尚未涉及完整病毒粒子的结构预测。很多种类的病毒衣壳带有二十面体对称性,这是最早应用电子显微学解析结构的病毒粒子之一^[20],也是在直接电子探测器应用于冷冻电镜数据收集之前,使用传统照相底片可以解析到近原子分辨率的极少数研究对象之一^[21]。即便是对这样有高度对称性的病毒结构预测,AlphaFold依旧没有很好的解决方案,对完整病毒粒子的结构研究仍然依赖于电子显微学,比如2019年爆发的猪瘟疫情,其病原体猪瘟病毒粒子直径达到280~300 nm,基因组为双链DNA,编码235种蛋白,是目前已知的最大最复杂的病毒。应用冷冻电镜,科学家成功解析了猪瘟病毒的精细3D结构^[22],包含有17 280个蛋白分子,共分5层,从外到内第一层外膜 (outer membrane)、第二层衣壳、第三层内膜 (inner membrane)、第四层壳核 (core shell)以及最里面的DNA,如此复杂的病毒粒子结构,用AlphaFold是很难预测其完整的结构。

某些病毒可以依赖宿主细胞对蛋白进行翻译后修饰,包括糖基化、甲基化、酰基化、磷酸化等,以加强病毒的免疫逃逸、感染复制、结构稳定、调控宿主细胞代谢,例如:新冠病毒依靠刺突蛋白的糖基化来逃避抗体的结合^[23]。对这些蛋白共价修饰的预测,AlphaFold的准确性还有待进一步提高,在发布AlphaFold 3的论文中,研究团队对蛋白质糖基化修饰进行了分析,其单个氨基酸残基糖基的成功率46.1%,多残基糖基的成功预测百分比为42.1%^[14],对其他翻译后修饰的预测有待进一步开发。而冷冻电镜对病毒蛋白的翻译后修饰确可以做到精准解析,例如:对人畜共患病毒Gatah病毒的解析,分辨率达到了2.8 Å^[24],不仅在病毒衣壳蛋白上解析出8个糖基化位点,揭示其参与病毒的免疫逃逸和入侵宿主的机制,还首次在衣壳蛋白跨膜螺旋膜内侧首次解析出5个酰基化位点,证实了酰基化可以极大程度地稳定单螺旋膜蛋白的跨膜组装。

5.3 蛋白质、复合体的动态结构

AlphaFold 3主要预测静态结构,而无法捕捉生物分子系统在溶液中的动态行为。蛋白质结构和功能的关系,是动态的;在执行生物学功能的过程中,蛋白质都要发生一些局部或全局的构象变化,这些局部或全局的构象变化是蛋白活性的决定因素。

有研究团队对AlphaFold预测蛋白构象变化的能力进行了大规模基准测试^[25],他们选取了92对已知的构象变化蛋白,每对蛋白质具有相同或高度相似的氨基酸序列,但可以形成两种截然不同的二级/三级结构。研究团队使用多个版本的AlphaFold模型,包括AF2.3.1、AF3以及两种增强采样方法,共生成超过282 000个结构预测,对AlphaFold的性能进行了全面评估。研究发现,即使对于可能已经包含在训练集中的92对构象变化蛋白,AlphaFold也只能成功预测其中35%的双重构象,这表明着AlphaFold在预测蛋白质多构象特性方面还存在不足。对于7个训练集外的构象变化蛋白,AlphaFold的预测成功率仅为14%,这表明AlphaFold对未见过的构象变化蛋白的预测能力有限。

最近几年引发的新冠肺炎大流行的新冠病毒,应用冷冻电镜,科学家解析了病毒表面的刺突蛋白spike的结构。spike是一个同源三聚体,三聚体存在有融合前 (pre-fusion)和融合后 (post-fusion)两种构象;此外在融合前构象中每个单体又独立存在预激活状态 (active-down)和完全激活状态 (active-up)两种构象^[26-27]。这些不同的构象变化不仅在刺突蛋白与宿主细胞ACE2受体结合中发挥重要作用,在开发抗体药物的进程中,不同构象的刺突蛋白与抗体结合的方式也是关键因素之一^[28-29]。

单颗粒冷冻电镜技术被广泛应用于离子通道蛋白结构研究,不仅仅是对常见的钾通道、钙通道和钠通道^[30-32],也包括一些特殊的离子通道。机械力感知决定我们的日常行为,如握手、拥抱、行走、血压等。力是无形的,那我们人体如何感知?直到2010年,Patapoutian教授团队报道了一类介导人体机械力感知的分子受体PIEZO,2012年证实了PIEZO蛋白是在哺乳动物中鉴定发现的首类机械门控阳离子通道,为我们认识了解这一基本的生物学问题带来了突破。2017-2020年多个弯曲状态PIEZO蛋白结构获得了解析,但未能解析出PIEZO受力开放的结构。2022年,科学家利用不同直径的脂质体,将PIEZO蛋白重组进磷脂膜中,通过蛋白

与脂质体之间的曲率差异来引入不同的膜张力,应用冷冻电镜技术,研究者们最终得到 PIEZO1 在膜上契合状态和受力展平的两种结构,分别命名为弯曲和平展构象,对应于离子通道关闭和打开两种构象,佐证了 PIEZO1 蛋白具备可逆形变和感知脂膜曲率变化的特殊能力^[33]。

蛋白质降解调控是极其重要的基本生物化学过程,在细胞周期、信号转导、免疫响应、基因调控、新陈代谢以及蛋白毒性响应等主要细胞分子过程中发挥着关键的调控作用。在真核细胞中,绝大部分胞内蛋白都是通过泛素蛋白酶体途径,经过泛素化标记被蛋白酶体全酶降解的。蛋白酶体全酶,又称为 26S Proteasome,是由中间一个圆柱形 20S 核心颗粒和两端覆盖的一个或两个 19S 调节颗粒组成。19S 包含一个环形异源六聚体马达——AAA-ATPase,通过多个协同 ATP 水解模式调控蛋白酶体降解泛素化底物,去泛素化酶 USP14 是最主要的蛋白酶体调控分子。USP14 通过结合 26S 而被激活,然后以毫秒的时间尺度剪切底物上的泛素链;但其如何被蛋白酶体激活并调控蛋白酶体功能,一直是全球研究机构和生物制药领域期待解决的关键科学问题。科学家通过采集海量的冷冻电镜数据,通过三维结构分类,将这些超过 300 万个 USP14-26S-泛素底物复合体的颗粒图像进行分类,捕获了 USP14-26S 复合体降解多泛素化底物过程的 13 种不同功能中间状态的高分辨 (3.0~3.6 Å) 构象,通过时间分辨冷冻电镜分析,重建了受控蛋白酶体的完整动力学工作周期,并结合分子生物学功能和基因突变研究,阐明了 USP14 和 26S 相互调控活性的原子结构基础和非平衡动力学机制^[34]。

此外,研究人员设计出毫秒量级时间分辨的冷冻电镜制样设备,提高了蛋白质动态构象研究的可行性。在常规的冷冻制样设备基础上,开发了喷雾-混合系统 Spotiton,可以在 10 ms 内,两个喷头把约 50 mL 的样品液滴分别洒在电镜样品载网上,两种样品开始混合,大约 100 ms 后,网格被浸入液体乙烷,样品的混合也随之停止。研究人员利用这一时间分辨的冷冻制样装置,研究了核糖体亚基之间的结合、RNA 聚合酶与启动子的结合、钙离子与钾离子通道的结合、动力蛋白 dynamin 对 GTP 的水解这些对时间分辨有着极高要求的样品^[35]。

5.4 原位结构

与其他常规的结构生物学手段一样,X 射线晶体学和单颗粒冷冻电镜制样依旧需要对蛋白质或

生物大分子复合物进行分离纯化,这一过程不可避免地要将其从天然的细胞环境中分离出来。对于很多蛋白质来说其正常生物学功能的执行需要周边其他蛋白或生物大分子与之相互配合、协调工作,很多膜蛋白的功能也依赖于周边特殊膜环境。这为单颗粒冷冻电镜提出一个巨大的疑问,即其所解析出的结构能否真实反映细胞中原位的构象?

冷冻电子断层成像技术 (Cryo-electron tomography, Cryo-ET) 是近年来快速发展起来的另一项冷冻电镜技术。Cryo-ET 可以直接对结构不均一的大分子复合体、细胞器、细胞、甚至是组织样本进行纳米尺度的三维高分辨率结构研究^[36]。Cryo-ET 可以在天然状态下获得细胞、组织、甚至是器官原位的三维精细结构、生物大分子的原位高分辨结构信息、以及大分子复合体原位相互作用结构,可以很好地与单颗粒技术互补,被认为是沟通分子生物学和细胞生物学的纽带^[37-38]。Cryo-ET 的优点包括:不需要化学固定,不需要对组织或细胞脱水,也不需要重金属染色等后期处理。Cryo-ET 直接将组织或细胞在近生理条件下进行高压冷冻或投入式快速冷冻,使含水的组织或细胞快速冷冻至非晶的玻璃态冰中,其结构的解析完全近乎于组织细胞内原位、自然状态。当然,Cryo-ET 与单颗粒技术相比较还存在分辨率较低的缺点,但最新的亚断层成像平均技术 (sub-tomogram averaging) 已经可以将分辨率提升至近原子分辨率,例如肺炎支原体细胞质中核糖体的原位结构^[39]。

在最新的一项研究中,科学家设计出一种结合亚断层成像平均技术与单颗粒技术的新方法,“子断层平均引导的单颗粒分析 (Subtomograms averaging guided single particle analysis, STAgSPA)”这一原位结构解析新策略。该策略中,研究者们首先使用冷冻离子束减薄 Cryo-FIB 技术制备了大量细胞切片,而后使用 Cryo-ET 技术分析了盐泽红藻藻胆体-光系统 II-光系统 I-捕光复合物超大复合体 (PBS-PSII-PSI-LHC magacomplex) 在类囊体膜上的分布特征,并使用子断层平均方法获取了中等分辨率的结构。借助这些信息,研究者们进一步通过高分辨模板匹配方法^[40-41],从原位的高剂量无倾角单颗粒照片中充分地挑选出潜在的复合体颗粒并进行原位单颗粒重构,解析了 PBS-PSII-PSI-LHC 超大复合体两种构象的原位结构,分辨率分别达到 3.3 和 4.3 Å^[42]。在这些结构中,研究人员发现了一些在分离纯化至体外的样品中未曾观察过的蛋

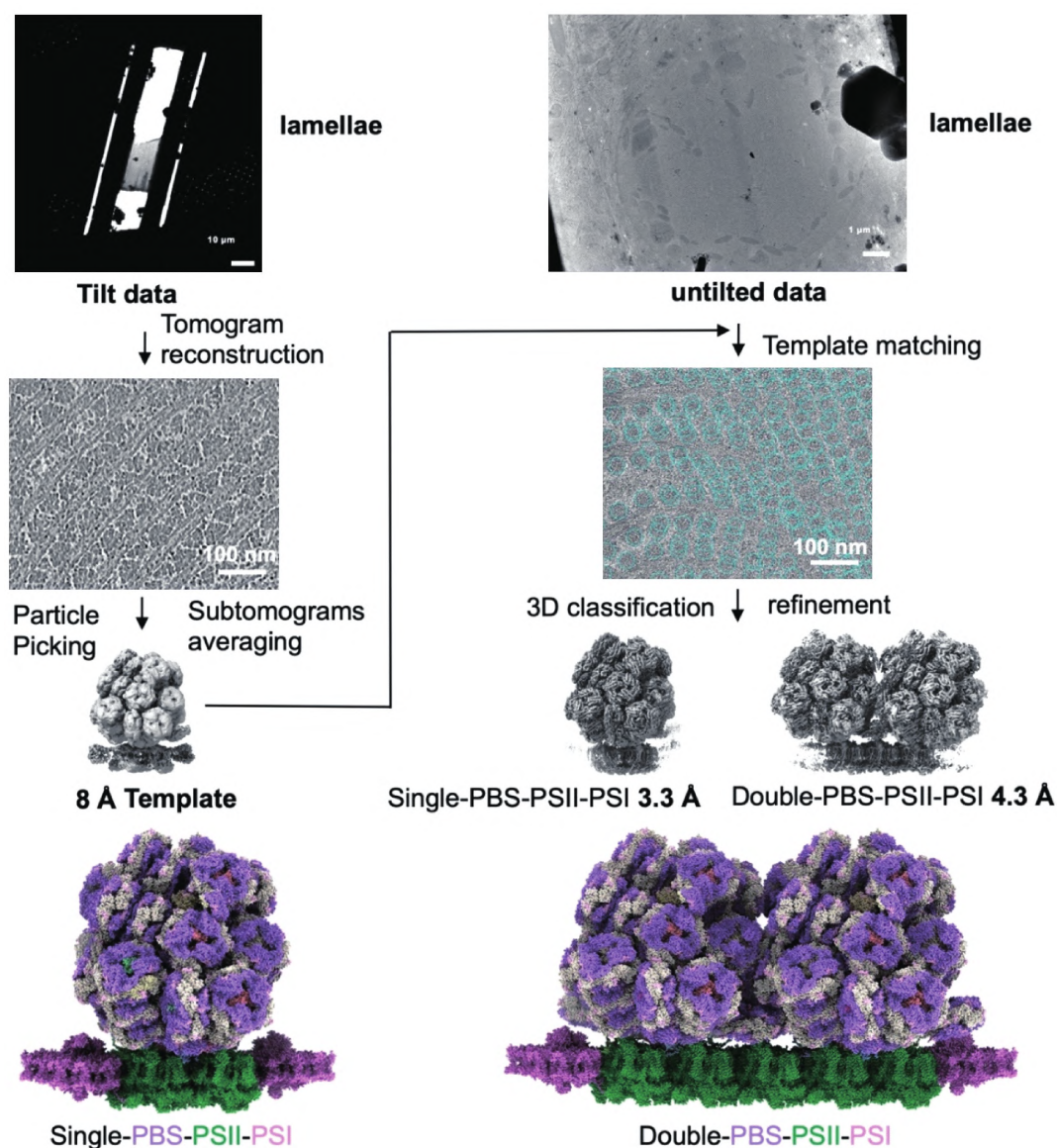


图2 红藻 *P. purpureum* 藻胆体-光系统 II-光系统 I 复合体的原位结构解析及整体结构。藻胆体-光系统 II-光系统 I 复合体的原位结构解析过程,以及两种构象的结构展示。

Fig. 2 *In situ* structural analysis and overall architecture of the PBS-PSII-PSI megacomplex from *P. purpureum*. Work flow of structural analysis for the PBS-PSII-PSI megacomplex, and the structural presentation of two type of complexes.

白分子在 PBS-PSII-PSI-LHC 超大复合体的组装中发挥至关重要的作用(图2)。应用相同的策略,科学家还解析了蓝藻藻胆体-光系统 II (PBS-PSII) 在类囊体膜上的原位结构,分辨率达到 $3.5 \text{ \AA}$ ^[43]。目前,在绝大多数原位研究中获取膜蛋白的近原子分辨率结构依然是一大挑战,该策略有效结合了子断层平均技术与单颗粒技术的优势,在获取原位蛋白三维分布信息的同时达到了高分辨结构解析的目的,为未来更多的原位结构分析工作提供了有价值的参考。

5.5 细胞、组织样本的大尺度超微结构研究

体电子显微电镜技术 (volume electron microscopy, VEM) 是近年来又一蓬勃发展的电镜技术,该技术主要对树脂包埋样品通过连续超薄切片及序列成像、扫描电镜原位块表面序列成像、扫描电镜原位聚焦电子束切割成像等技术手段最终获得大体积高分辨显微结构数据的方法^[44]。

电镜的三维重构技术通常要在分辨率和视场 (field of view, FOV) 之间有所取舍。体电镜的视场 (X-Y 平面上) 可以达到几十微米甚至数毫米尺度,

样品的厚度可以从数微米到数毫米,但它的分辨率不及冷冻电镜,受化学染色的限制,分辨率通常在 5~10 nm,而冷冻样品的分辨率甚至能达到 0.1 nm。然而冷冻电镜的视场通常只有数微米,且对样品的厚度也有所限制(通常厚度 < 200 nm)。对于生物样品而言,应根据样品的实际情况和目标来选择合适的三维重构技术。在研究大尺度生物样品,例如细胞器、真核细胞结构、组织结构和器官构造时,体电镜技术无疑是最佳选择。正因为体电镜在生物学和医学研究中的极大潜在价值,2023 年 1 月 Nature 杂志将体电镜与单分子蛋白质测序技术、詹姆斯韦伯太空望远镜、基因编辑 CRISPR 技术、高精度放射性碳测量技术、单细胞代谢组学技术和体外胚胎模型评选为 2023 年最值得关注的七项新技术^[45],2024 年 5 月谷歌与哈佛大学 Lichtman 实验室合作,将体电镜应用于人类大脑皮层 nm 级分辨率的超微结构研究^[46]。这项研究中,研究者获得了一位癫痫患者在进行脑部手术过程中切除的左前颞叶的脑组织,样品经由固定、染色、包埋后,使用超微切片机对其进行切片,共收集 5 019 个组织切片,平均切片厚度为 33.9 nm,样品总厚度为 0.170 mm。随后,这些切片在扫描电镜下以 4 nm×4 nm 的分辨率成像,总成像体积约为 1 mm³;每个切片的原始数据大小高达 350 千兆,原始数据集大小高达 1.4 PB,这一立方毫米大脑组织样本中分割出了 57 000 个细胞、约 230 mm 的血管,以及大约 1.5 亿个突触。类似这样几乎宏观尺度(macroscopic scale)和介观尺度(mesosopic scale)对超级复杂的组织样本进行电子显微学研究会是结构生物学研究的下一个热点,这也是目前 AlphaFold 3 结构预测望尘莫及的。

5.6 高精度药物分子设计

结构生物学的终极目标之一便是通过解析高分辨率结构来实现靶向药物分子的设计。AlphaFold 不仅可以预测高精度的蛋白质结构,也能够对药物分子(例如:与蛋白质结合的配体和抗体)进行预测,进而加快药物设计流程,为药物设计赋能。氨基酸侧链是蛋白折叠驱动力与药物设计的关键,然而,通过 AlphaFold“预测”所得出的模型,在侧链准确度上仍不足够,这会影响药物设计的准确度;另一方面,AlphaFold 的预测模型仅与实验解析

的“结构模型”高度相似,这些模型可能继续甚至放大了实验数据中的误差,进而造成靶向药物设计的准确。

还需强调的是,目前基于结构的药物发现过分强调了单一静态结构的重要性,却忽略了生物大分子往往以不同的能量分布形式即多种不同构象存在。这些不同构象间的分布概率形成了能量地形图,有利于解释单一结构无法解释的药物学问题,也将是药物开发的新方向。更为关键的问题是,体外研发的药物分子往往在体内被其他生物大分子截获,这也会影响药物分子设计。显而易见,药物分子设计,包括 AlphaFold 辅助的药物分子设计,依旧任重道远。

6 结语

基于结构生物学长期积累的信息,蛋白结构预测技术 AlphaFold 等已经实现了对单个蛋白的较为精准的结构预测。然而在很多情况下,单个蛋白的高分辨率结构不足以全面揭示蛋白质功能和作用机制,未来的结构生物学是以冷冻电镜单颗粒技术为核心、结合冷冻电子断层成像技术和体电镜技术的电子显微学,辅以传统的光学显微镜、X 射线晶体学、核磁共振以及基于人工智能的蛋白质预测技术 AlphaFold 等,将从单个蛋白或复合体的结构,全面拓展到更为复杂的“复合物”、“多蛋白体系”、“体内原位”、“微观、介观和宏观尺度”等研究中。在多个尺度、不同模态上实现对细胞器、细胞的原位高分辨率动态结构解析。这些实验手段解析的高分辨率结构是目前人工智能所无法实现和替代的,为人工智能提供了新挑战。

所有这些复杂结构将会为人工智能技术继续提供更多的信息,也为人工智能在结构生物学方向提供了新方向与新机遇。伴随着算力的爆炸式增长以及算法的持续更新迭代,人工智能也将在更复杂体系的高分辨率动态结构预测方面取得突破性的进展。在某种意义上,冷冻电镜技术将持续引领人工智能技术在结构生物学方向的发展;而人工智能技术将发挥加速器与倍增器的作用,促进以冷冻电镜为代表的结构(细胞)生物学的高速发展(图 3)。

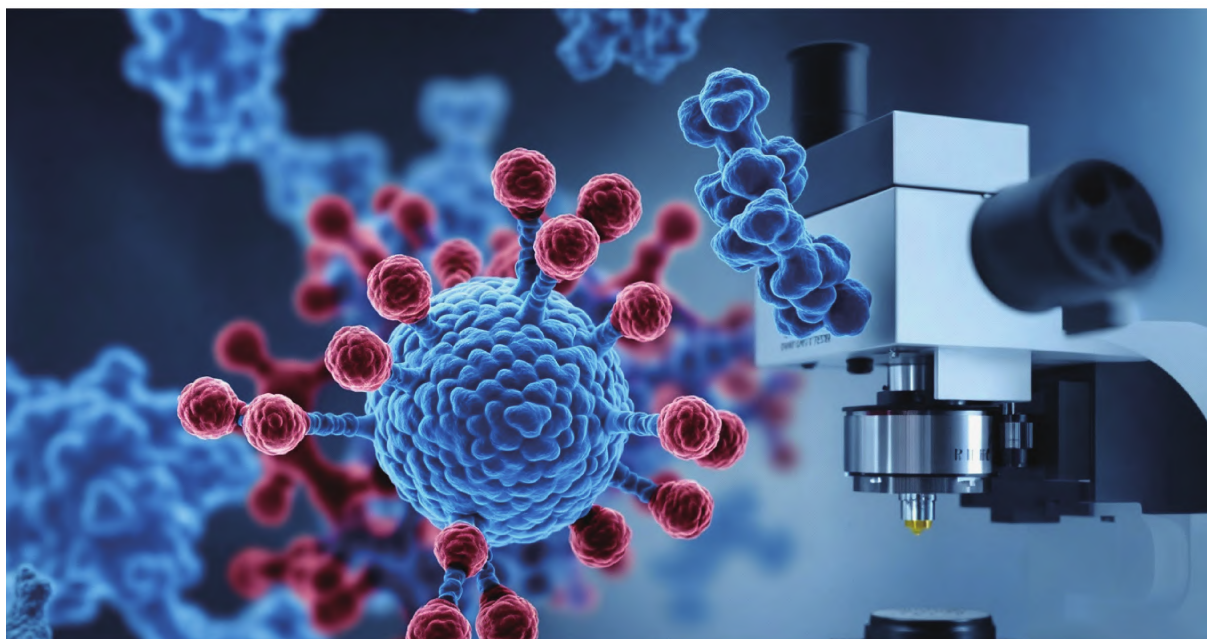


图3 AI生成的电子显微镜与结构预测关系的图片 <https://www.vivago.ai>。AI智能快速生成了此图片,具有很好的艺术效果。然而,由于受到训练集及训练指令的影响,AI智能无法正确生成电子显微镜,无法很完美地呈现电子显微镜与结构预测的关系,这充分反映了基于AI智能的结构预测遇到的困境。图片由 <https://www.vivago.ai> 生成。

Fig. 3 AI-generated fake image on the relationship between cryo-electron microscopy and structural prediction. Limited by the training data, AI can only generate light microscope instead of electron microscope, and cannot build the right relationship between EM and structural prediction. The image was produced in <https://www.vivago.ai>.

参考文献:

- [1] ABBE E. Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung [J]. Archiv für Mikroskopische Anatomie, 1873,9: 413-418.
- [2] DE ROSIER D J, KLUG A. Reconstruction of three dimensional structures from electron micrographs [J]. Nature, 1968,217 (5124): 130-134.
- [3] NAKANE T, KOTECHE A, SENTE A, et al. Single-particle Cryo-EM at atomic resolution [J]. Nature, 2020,587 (7832): 152-156.
- [4] YIP K M, FISCHER N, PAKNIA R E, et al. Atomic-resolution protein structure determination by Cryo-EM [J]. Nature, 2020,587 (7832): 157-161.
- [5] STOKSTAD E, PENNISI E, KAISER J, et al. The runners up [J]. Science, 2017,358 (6370): 1522-1527.
- [6] Viruses, microscopy and fast radio bursts: 10 remarkable discoveries from 2020 [J]. Nature, 2020,588 (7839): 596-598.
- [7] WYLLIE S, BRAND S, THOMAS M, et al. Preclinical candidate for the treatment of visceral leishmaniasis that acts through proteasome inhibition [J]. Proc Natl Acad Sci USA, 2019,116 (19): 9318-9323.
- [8] SELA M, WHITE F H, ANFINSEN C B. Reductive cleavage of disulfide bridges in ribonuclease [J]. Science, 1957,125 (3250): 691-692.
- [9] KRAUT J. Structural studies with X-rays [J]. Annu Rev Biochem, 1965,34: 247-268.
- [10] JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold [J]. Nature, 2021,596 (7873): 583-589.
- [11] KREITZ J, FRIEDRICH M J, GURU A, et al. Programmable protein delivery with a bacterial contractile injection system [J]. Nature, 2023,616 (7956): 357-364.
- [12] LIM Y, TAMAYO-ORREGO L, SCHMID E, et al. In silico protein interaction screening uncovers DONSON's role in replication initiation [J]. Science, 2023,381 (6664): eadi3448.
- [13] MOSALAGANTI S, OBARSKA-KOSINSKA A, SIGGEL M, et al. AI-based structure prediction empowers integrative structural analysis of human nuclear pores [J]. Science, 2022,376 (6598): eabm9506.
- [14] ABRAMSON J, ADLER J, DUNGER J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3 [J]. Nature, 2024,630 (8016): 493-500.

- [15] KYRILIS F L, SEMCHONOK D A, SKALIDIS I, et al. Integrative structure of a 10 – megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts [J]. *Cell Rep*, 2021, 34 (6): 108727.
- [16] JIN Z, WAN L, ZHANG Y, et al. Structure of a TOC-TIC supercomplex spanning two chloroplast envelope membranes [J]. *Cell*, 2022, 185 (25): 4788–4800.
- [17] LIU H, LI A, ROCHAIX J D, et al. Architecture of chloroplast TOC-TIC translocon supercomplex [J]. *Nature*, 2023, 615 (7951): 349–357.
- [18] MA J, YOU X, SUN S, et al. Structural basis of energy transfer in *Porphyridium purpureum* phycobilisome [J]. *Nature*, 2020, 579 (7797): 146–151.
- [19] NOMBURG J, DOHERTY E E, PRICE N, et al. Birth of protein folds and functions in the virome [J]. *Nature*, 2024, 633 (8030): 710–717.
- [20] CROWTHER R A, AMOS L A, FINCH J T, et al. Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs [J]. *Nature*, 1970, 226 (5244): 421–425.
- [21] ZHANG X, JIN L, FANG Q, et al. 3.3 Å Cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry [J]. *Cell*, 2010, 141 (3): 472–482.
- [22] WANG N, ZHAO D, WANG J, et al. Architecture of African swine fever virus and implications for viral assembly [J]. *Science*, 2019, 366 (6465): 640–644.
- [23] TIAN W, LI D, ZHANG N, et al. O-glycosylation pattern of the SARS-CoV-2 spike protein reveals an “O-Follow-N” rule [J]. *Cell Res*, 2021, 31 (10): 1123–1125.
- [24] WANG A, ZHOU F, LIU C, et al. Structure of infective Getah virus at 2.8 Å resolution determined by Cryo-electron microscopy [J]. *Cell Discov*, 2022, 8 (1): 12.
- [25] CHAKRAVARTY D, SCHAFER J W, CHEN E A, et al. AlphaFold predictions of fold-switched conformations are driven by structure memorization [J]. *Nat Commun*, 2024, 15 (1): 7296.
- [26] YAO H, SONG Y, CHEN Y, et al. Molecular Architecture of the SARS-CoV-2 Virus [J]. *Cell*, 2020, 183 (3): 730–738.
- [27] MA X, WANG Y, GAO Y, et al. Structural characteristics of the SARS-CoV-2 Omicron lineages BA.1 and BA.2 virions [J]. *Signal Transduct Target Ther*, 2023, 8 (1): 131.
- [28] WANG Y, YAN A, SONG D, et al. Biparatopic antibody BA7208/7125 effectively neutralizes SARS-CoV-2 variants including Omicron BA.1–BA.5 [J]. *Cell Discov*, 2023, 9 (1): 3.
- [29] YAN R, ZHANG Y, LI Y, et al. Structural basis for the different states of the spike protein of SARS-CoV-2 in complex with ACE2 [J]. *Cell Res*, 2021, 31 (6): 717–719.
- [30] SHEN H, L Z I, JIANG Y, et al. Structural basis for the modulation of voltage-gated sodium channels by animal toxins [J]. *Science*, 2018, 362 (6412): eaau2596.
- [31] GAO S, YAO X, CHEN J, et al. Structural basis for human Ca_v1.2 inhibition by multiple drugs and the neurotoxin calciseptine [J]. *Cell*, 2023, 186 (24): 5363–5374.
- [32] HITE R K, TAO X, MACKINNON R. Structural basis for gating the high-conductance Ca²⁺-activated K⁺ channel [J]. *Nature*, 2017, 541 (7635): 52–57.
- [33] YANG X, LIN C, CHEN X, et al. Structure deformation and curvature sensing of PIEZO1 in lipid membranes [J]. *Nature*, 2022, 604 (7905): 377–383.
- [34] ZHANG S, ZOU S, YIN D, et al. USP14–regulated allostery of the human proteasome by time-resolved Cryo-EM [J]. *Nature*, 2022, 605 (7910): 567–574.
- [35] DANDEY V P, BUDELL W C, WEI H, et al. Time-resolved Cryo-EM using Spotiton [J]. *Nat Methods*, 2020, 17 (9): 897–900.
- [36] LUČIĆ V, RIGORT A, BAUMEISTER W. Cryo-electron tomography: The challenge of doing structural biology *in situ* [J]. *J Cell Biol*, 2013, 202 (3): 407–419.
- [37] KONING R I, KOSTER A J, SHARP T H. Advances in Cryo-electron tomography for biology and medicine [J]. *Ann Anat*, 2018, 217: 82–96.
- [38] TURK, M, BAUMEISTER W. The promise and the challenges of Cryo-electron tomography [J]. *FEBS Lett*, 2020, 594 (20): 3243–3261.
- [39] XUE L, LENZ S, ZIMMERMANN-KOGADEEVA M, et al. Visualizing translation dynamics at atomic detail inside a bacterial cell [J]. *Nature*, 2022, 610 (7930): 205–211.
- [40] CHENG J, LI B, SI L, et al. Determining structures in a native environment using single-particle cryoelectron microscopy images [J]. *Innovation (Camb)*, 2021, 2 (4): 100166.
- [41] CHENG J, LIU T, YOU X, et al. Determining protein structures in cellular lamella at pseudo-atomic resolution by GisSPA [J]. *Nat Commun*, 2023, 14 (1): 1282.
- [42] YOU X, ZHANG X, CHENG J, et al. *In situ* structure of the red algal phycobilisome-PSII-PSI-LHC megacomplex [J]. *Nature*, 2023, 616 (7955): 199–206.

- [43] ZHANG X, XIAO Y, YOU X, et al. *In situ* structural determination of cyanobacterial phycobilisome-PSII supercomplex by STAgSPA strategy [J]. *Nat Commun*, 2024,15 (1): 7201.
- [44] PEDDIE C J, GENOUD C, KRESHUK A, et al. Volume electron microscopy [J]. *Nature Reviews Methods Primers*, 2022,2 (1): 51.
- [45] EISENSTEIN M. Seven technologies to watch in 2023 [J]. *Nature*, 2023,613 (7945): 794–797.
- [46] SHAPSON-COE A, JANUSZEWSKI M, BERGER D R, et al. A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution [J]. *Science*, 2024,384 (6696): eadk4858.

Electron microscopy research in the age of AI: Severe challenges, Infinite opportunities, and Magnificent prospects

LIU Zheng^{1*}, SHEN Qingtao², SUI Senfang^{1,2,3}

(1. Cryo-Electron Microscopy Center, Southern University of Science and Technology, Shenzhen Guangdong 518055;

2. School of Life Sciences, Southern University of Science and Technology, Shenzhen Guangdong 518055;

3. School of Life Science, Tsinghua University, Beijing 100084, China)

Abstract Over the past decade, structural biology has made significant achievements, there are two representational breakthroughs: First, in structural determination techniques, represented by Cryo-electron microscopy, and second, in structural prediction models, exemplified by the AlphaFold algorithm. In May 2024, Google DeepMind updated its biological structure prediction tool. The latest version, AlphaFold 3, possesses the ability to predict the structures of nearly all types of protein complexes and has also achieved high accuracy in predicting drug interactions. The release of AlphaFold 3 has brought about transformative changes structural biology, showcasing the immense potential of AI technology and fueling public enthusiasm and imagination for life science and biomedical research. Meanwhile, research in electron microscopy continues to advance with continuous emergence of new technologies and methods, electron microscopy science retains irreplaceable advantages in analyzing novel protein structures, extremely large and complicate complexes, dynamic structures, *in-situ* structures, and studies of larger scale samples including cells, tissues, and organs. Some propose that structural prediction models could even replace traditional experimental approaches such as X-ray crystallography and electron microscopy, allowing biomolecular structures analysis through computation alone. However, such opinions are biased, future research in structural biology will inevitably integrate experimental science and AI technology, expanding from the structures of individual proteins or complexes to comprehensive studies of multi-protein systems, intracellular *in-situ* locations, and beyond microscopic to mesoscopic and macroscopic scales.

Keywords electron microscopy science; high-resolution structure; AlphaFold; structure prediction

* Corresponding author