

蛋白质三维结构解析的方法

徐甲强¹, 崔倩倩^{1,2}

(1.上海大学 理学院化学系;新能源与传感技术实验室,上海 200444;2.中国科学院 上海药物研究所,上海 201203)

摘要:三维结构是生物大分子功能分析的基础.对生物大分子的原子框架深入研究可以确定其结构、功能和进化的根源,有助于指导疾病的治疗以及靶向药物的开发.蛋白质是生命活动的物质承担者和体现者,对蛋白质结构的解析是阐明其发挥功能的分子机制的前提.从原理、发展历程及优缺点等方面对现有的 4 种蛋白质结构解析方法进行了简要概述,即 X 射线衍射法、核磁共振法、冷冻电子显微镜法以及基于大数据进行蛋白质结构预测的 AlphaFold 人工智能系统,旨在回顾结构生物学的发展历程,系统梳理目前蛋白质结构解析的发展现状.

关键词:蛋白质三维结构; X 射线衍射;核磁共振;冷冻电子显微镜;AlphaFold

中图分类号:Q334

文献标志码:A

蛋白质是执行生命活动过程中重要的生物大分子,参与肌肉收缩、能量转换、氧气输送以及抵御微生物入侵等各种生理过程.虽然蛋白质的功能各不相同,但是其基本组成是类似的.其通过 20 种氨基酸排列组合形成不同的肽链,再进一步地组装和折叠,成为具有独特且复杂三维结构的生物大分子^[1].结构的正确解析是理解蛋白质生物学功能、认识各蛋白质之间相互作用的关键^[2],可以为药物的研发提供重要的靶点信息^[3].

目前蛋白质结构解析的方法分为实验方法和预测方法两类,前者主要包括 X 射线晶体衍射(X-ray Crystalline Diffraction)法、核磁共振(Nuclear Magnetic Resonance, NMR)法和冷冻电子显微镜(Cryogenic Electron Microscopy, Cryo-EM)法,这 3 种方法逐渐发展成熟,已成为蛋白质结构实验分析的三大技术平台.后者主要以 AlphaFold 人工智能系统的建立为标志,基于大数据的计算对蛋白质结构进行预测.各种结构解析方法占比情况见图 1.

1 X 射线晶体衍射法

生物大分子结构测定技术的历史可以追溯到 19 世纪末.1895 年,德国物理学家伦琴发现了 X 射线.X 射线是一种高能量、短波长的电磁波(本质上属于光子束),当 X 射线击打在分子晶体颗粒上时,因与晶体中的电子发生作用而产生衍射效应.通过探测器收集这些衍射信号,可以确定晶体中电子密度的分布,从而获得粒子的位置信息,为解析生物大分子的晶体结构提供依据.X 射线衍射法凭借其优异的性能在接下来的一个多世纪获得了极大发展,被广泛用于生物医学和工业领域.只要是能够形成结晶的蛋白质,通过 X 射线衍射后就能看清楚其三维结构,因此 X 射线晶体衍射法成为解析蛋白质大分子晶体结构的主要手段,并逐渐发展为一门学科即 X 射线晶体学(X-ray Crystallography).X 射线晶体学成为最早且最主要研究测定蛋白质结构的学科^[4].

1934 年,BEMAL 和 CROWFOOT 获得了第一张蛋白质晶体的 X 射线衍射照片,但是由于理论的局限性,并未解析出蛋白质的三维结构^[5].1958 年,英国科学家 MAX PERUTZ 和 JOHN KENDREW 将重金属原子引入蛋白质晶体中,通过运用多对同晶置换法(Multiple Isomorphous Replacement, MIR)成功解决蛋白

收稿日期:2022-05-28;修回日期:2022-06-13.

基金项目:国家自然科学基金(31670743).

作者简介(通信作者):徐甲强(1963—),男,河南卫辉人,上海大学教授,博士生导师,研究方向为纳米材料化学,E-mail:

xujiaqiang@shu.edu.cn.

质大分子晶体衍射的相位问题,并得到了高分辨率的肌红蛋白(Myoglobin)的三维结构^[6](如图2),以及更复杂的血红蛋白(Hemoglobin)的晶体结构^[7]。自此,X射线衍射技术在蛋白质结构测定中得到了实际的应用,PERUTZ与KENDREW也获得了1962年的诺贝尔化学奖。1964年,AARON KLUG将X射线晶体学与电子显微镜结合起来,发明了一种可以解析更大的蛋白质或蛋白质-核酸复合物的新的分析仪器——电子晶体学显微镜(Crystallographic Electron Microscopy)^[8]。由于KLUG在结构生物学领域的杰出贡献,被授予了1982年的诺贝尔化学奖。1965年,BLAKE和PHILLIPS用X射线衍射技术测定了溶菌酶的三维结构并首次解释了蛋白质的分子机制^[9-10]。但由于此时产生的X射线存在相应的缺陷,如没有相干性、亮度不高、偏振性较差等,在很大程度上限制X射线的应用,直到同步辐射技术的出现才使得部分难题得到解决。1947年,学者在美国通用电气公司的一台70 MeV的同步加速器中首次观察到“同步光”现象,并将该现象命名为同步辐射。同步辐射是指速度接近光速的带电粒子在做曲线运动时,沿切线方向发出的电磁辐射。该辐射拥有高强度、宽波谱、高准直性及方便可调节的入射X射线等优点,弥补了之前X射线在亮度和偏振性方面的不足,使得X射线晶体学得到了极大的发展^[11]。1976年,KEITH HODGSON首次使用同步辐射装置对单个蛋白质晶体进行照射,观察到生物大分子的电子和结构环境^[12]。然而,蛋白质结晶过程并不简单,有时需要改变样品的结构才能得到有序的晶体,这使得人们可能会错过蛋白天然状态下的结构信息,因此科学家们尝试在非结晶的状态下解析蛋白质结构。同步辐射的发展为小角X射线散射(Small Angle X-ray Scattering, SAXS)提供了高强度的入射光源,实现对样品的快速曝光,因此,与同步辐射只能解析静态或微动态的蛋白质结构不同,应用SAXS技术可以获得溶液状态下蛋白质结构的动态变化。然而,科学家们对X射线晶体学的探索不止于此。21世纪初,德国率先建造了X射线自由电子激光器(X-ray Free-Electron Laser, XFEL),实现了对生物大分子皮秒(ps)量级甚至飞秒(fs)量级的动态结构的观察^[11]。大分子结构的预测不再局限于静止状态,已经进入了时间分辨晶体学的新时代。

自20世纪90年代以来,X射线晶体学迎来了飞速发展的时期。直至今日,X射线晶体衍射技术仍是最常用且最准确的测定蛋白质三维结构的技术^[13]。目前PDB(Protein Data Bank)数据库中已经收录了超过18万个生物大分子结构,如图1所示,其中约86.78%是通过X射线衍射技术获得的。X射线晶体衍射技术虽然得到了长足的发展,但仍存在一定的缺点。比如,X射线本身带有极大的能量,容易对晶体造成辐射损伤,即使是液体状态下的蛋白质也会受到影响。此外,X射线衍射方法不能用来解析分子量较大的蛋白质,且对于不能结晶的蛋白质的解析也存在一定的局限性。随后核磁共振波谱学的发展在一定程度上弥补了X射线晶体衍射技术的不足,逐渐发展成为蛋白质结构解析的新方法。

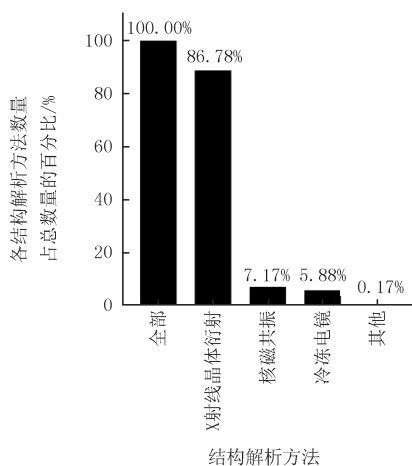


图1 Protein Data Bank数据库中通过X射线晶体衍射、核磁共振波谱、冷冻电镜及其他技术解析的结构数量占总结构数量的百分比(统计至2022年6月)

Fig.1 The number of structures resolved by X-ray crystallography, nuclear magnetic resonance spectroscopy, Cryo-electron microscopy and other techniques in the Protein Data Bank database as the percentage of the total (data released in Protein Data Bank until June 2022)

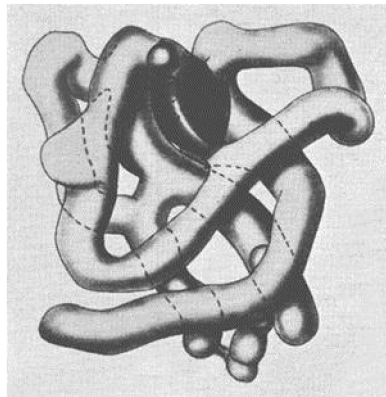


图2 X射线衍射解析得到的肌红蛋白的三维结构图,分辨率为0.6 nm^[6]
Fig.2 The three-dimensional structure diagram of myoglobin obtained by X-ray diffraction analysis, with a resolution of 0.6 nm^[6]

2 核磁共振波谱法

利用核磁共振波谱进行的结构研究可以观察到蛋白质分子的天然结构和动力学特征^[14].核磁共振的基本理论是带有孤对电子的原子核在外界磁场的影响下发生自旋,并通过塞曼分裂发生能级跃迁,从而吸收并释放电磁辐射,产生共振频谱.这种共振电磁辐射的频率与所处磁场强度成一定比例.由于不同分子中原子核所处的化学环境不同,将会产生不同的共振频率,从而产生不同的共振谱.通过记录波谱的变化可以判断该原子在分子中所处的位置及相对数目,用以进行定量分析及分子量的测定,并对分子进行结构分析.

1938年,美国科学家 ISIDOR ISAAC RABI 首次通过分子束实验发现在磁场中的原子核会沿磁场方向呈正向或反向有序且平行的排列,在施加外磁场后,原子核的自旋方向发生翻转^[15].这是人类关于原子核与磁场以及外加射频场相互作用的最早认识.1946年,PURCELL 和 BLOCH 都独立观察到了固体(石蜡)和液体(水)状态下的核磁共振吸收^[16-17],两位科学家也因此获得了 1952 年的诺贝尔物理学奖.1948年,核磁弛豫理论建立.1950年,化学位移和自旋耦合现象被相继发现.之后,RICHARD R ERNST 团队于 1966 年建立了傅立叶变换核磁共振分光法,并于 1976 年建立二维核磁共振技术.这两项技术极大地提高了 NMR 测量的灵敏度和分辨率,并将核磁共振成功应用到化学、结构生物学和日常的医学诊断中^[18].随后,WUTH-RICH 将该方法成功应用于生物大分子结构与动力学研究^[19].这些理论学说的建立使得 NMR 快速发展.1978年,核磁共振技术首次被用于蛋白质结构的解析^[20].经过半个多世纪的发展,尤其是高场核磁共振谱仪的出现、各种新的核磁实验方法的应用、分子生物学的进步以及各种标记方法(如同位素标记技术)的产生等使得核磁共振技术在结构生物学中得到广泛应用^[18].直至今日,核磁共振波谱技术仍然是能够在原子分辨率下测定溶液中生物大分子三维结构的最可靠的方法之一.如图 1 所示,PDB 数据库中收录的生物大分子三维结构中约有 7.17%是通过 NMR 技术获得的.

核磁共振法解析蛋白质结构的一般流程如下:1)对蛋白质进行同位素标记,通常用¹⁵NH₄Cl 为氮源、¹³C-葡萄糖为碳源的 M9 培养基培养大肠杆菌并表达蛋白^[21](现在逐渐发展成为对氨基酸的选择性标记);2)将蛋白放置到强磁场中,用射频信号进行探测,进行¹H、¹³C、¹⁵N 三共振 NMR 实验,观测到的共振信号可以反映相邻原子核之间的相互作用和成键原子之间的局部构象,即对蛋白质的主、侧链进行归属^[22-23];3)对采集完的 NMR 谱图进行氨基酸的比对,结合 NOE(核欧沃豪斯效应增强)约束、二面角约束(通过耦合常数所得)以及空间距离约束(由三维¹⁵N-¹³C-NOESY-HSQC 实验获得),来构建蛋白质原子模型;4)结合计算机不断演算,得到最终的蛋白质结构.图 3 为文献[24]中利用核磁共振光谱测定溶液状态下的锌指核酸结合基序的三维结构图.

NMR 结构解析大多是对溶液状态下的蛋白质结构进行非破坏的检测,因此一般认为比起 X 射线晶体衍射,NMR 能更准确地描述生物大分子的天然结构.核磁共振波谱技术适合研究瞬时存在和不稳定的复合物.如采用半滤波的核欧沃豪斯效应谱(Half-Filrered NOESY)可以测定存在弱相互作用复合物的三维结构^[25-26].同时,核磁共振技术也适合研究蛋白质动力学,其时间尺度可以囊括皮秒到秒的范围^[18],可精确得到目的基团的动态变化信息.此外,利用液体或固体核磁共振的方法,逐步突破了分子大小以及长程约束的限制,使得 NMR 还可以研究膜蛋白的三维结构,特别是一些通过 X 射线晶体衍射技术和 Cryo-EM 技术不能确认的蛋白质的柔性区域的结构信息.当然,NMR 也存在一些劣势,如蛋白质在溶液中存在结构不稳定的情况,这种状态下很难获取稳定的 NMR 信号.此外,由于大分子量的蛋白质在 NMR 谱中存在重叠峰的问题,因此 NMR 技术在处理大型生物分子时的能力有限^[14].而冷冻电子显微镜技术突破了蛋白质分子量的限制,成功打破 X 射线晶体衍射技术和 NMR 技术的壁垒,成为当前结构生物学领域最前沿的成像技术之一,极大地推动了膜蛋白及大分子复合物的结构解析进程.

3 冷冻电子显微镜法

使用冷冻电子显微镜对液体样品进行玻璃化处理,集中了 X 射线晶体衍射和核磁共振波谱两种方法的优点,成为结构研究最合适的替代方法^[14].冷冻电子显微镜技术的原理是将生物大分子在毫秒的时间尺度

内快速冷冻并包埋在玻璃态的冰中,应用低温透射电子显微镜收集样品信息,即用高度相干的电子做光源,采用 80~300 kV 加速电压将电子束加速,透过样品和附近冰层,使样品受到散射,再利用探测器和透镜系统把散射信号记录下来,收集生物大分子不同方向上的二维投影,再经图像处理并利用三维重构的方法计算得到大分子三维精细结构。

电子显微镜最早出现在 1931 年,其设计之初的目的就是为了获得高分辨率的病毒图像。但由于电子显微镜必须在高真空下工作,而生物样品大多在真空状态下不能保持稳定甚至失活,因此制备高分辨率的生物样品是当时制约电子显微镜技术发展的主要瓶颈,导致其在结构生物学方面的应用要远远落后于材料科学。在冷冻电镜技术发展之前,一般使用重金属盐覆盖生物样品,用常温负染透射电镜进行观察,但是获得的生物大分子分辨率较低^[27]。

1968 年,DE ROSIER 和 KLUG 成功地用负染的电子显微镜照片重建了噬菌体尾部(一种螺旋对称的粒子)的三维结构,开创了电子显微镜三维重构技术,取得了蛋白质结构解析的重大突破^[28](图 4)。1974 年,ROBERT GLAESER 和 KEN TAYLOR 首次提出对含水生物样品冷冻后再进行电子显微镜成像,通过测试发现该方法可以有效降低辐射损伤对样品高分辨率结构的破坏,开创了维持高真空状态下高分辨率成像的新思路,这就是冷冻电子显微镜的雏形^[29]。1975 年,HENDERSON 等人开创了二维电子晶体学三维重构技术,即应用电子显微镜从二维晶体中获得近原子分辨率的三维蛋白质结构,并使用该方法解析了第一个膜蛋白的三维结构——视紫红质蛋白^[30]。FRANK 等人建立了基于冷冻电镜单颗粒重构技术的图像分析方法,能够从 2D 电子显微图像构建 3D 结构^[31]。DUBOCHET 教授创建了样品制备的低温玻璃化法,该方法通过使用液氮冷却的液态乙烷实现了毫秒之内的生物样品的快速冷冻,成功制备了不形成冰晶的玻璃态冰包埋的生物样品,能够使生物分子在真空中保持其自然状态,并有效抵御电子的辐照损伤^[32]。上述开拓性的科研成果共同奠定了冷冻电子显微镜技术发展的基础,并获得了科学界的认可。HENDERSON、FRANK 和 DUBOCHET 3 位科学家获得了 2017 年的诺贝尔化学奖。

起初,使用冷冻电子显微镜技术只能解析高对称性且大分子量的病毒颗粒的近原子分辨率的三维结构^[33],而后这项技术取得了革命性的进步,其主要的的原因之一是电子直接探测器(Direct Electron-Detector Device,DDD)的发展。DDD 允许完全相同的图像区域的多个图片通过帧对齐和取平均的方法校正光束引起的背景运动,显著提高图像的信噪比(Signal-to-Noise Ratio,SNR)^[14]。除了图像处理硬件的突破之外,冷冻电子显微镜技术的进步还得益于图像处理软件的发展^[34]。2012 年 SJORS SCHERES 开发的 RELION 算法能更有效地处理低信噪比的图像,这成为单颗粒结构解析的利器。基于前人的研究基础,2013 年,新一代 DDD 相机(Gatan K2 camera)拍摄了近九万张单颗粒图像,并解析得到了约 300 kDa 的膜蛋白——瞬时受体

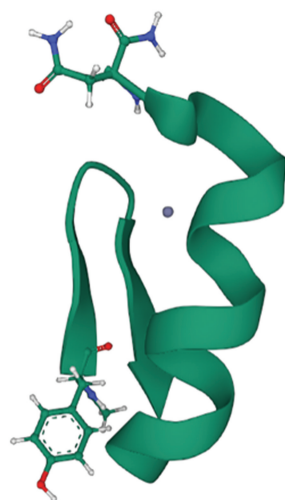


图3 通过核磁共振光谱测定溶液状态下的锌指核酸结合基序的三维结构^[24]

Fig.3 The three-dimensional structure of zinc finger nucleic acid binding motif in solution determined by nuclear magnetic resonance spectroscopy^[24]

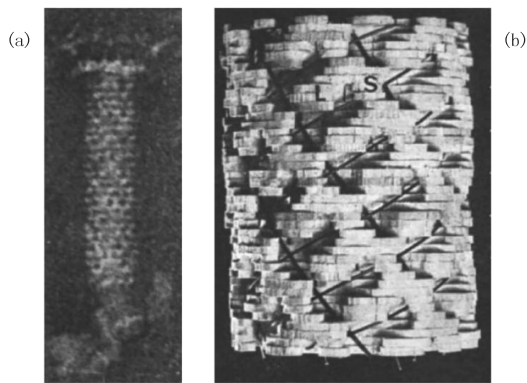


图4 T4噬菌体经甲酸铀酰负染后的电子显微镜图(a);伸展状态下的噬菌体尾部(b)^[28]

Fig.4 Electron micrograph of a tail of bacteriophage T4, negatively stained with uranyl formate(a); Model of the phage tail in the extended state(b)^[28]

电位通道蛋白(TRPV1)四聚体的冷冻电镜结构,分辨率高达近原子级别的 $0.34\text{ nm}^{[35]}$,标志着冷冻电子显微镜技术进入“原子分辨率”时代,并掀起了一场“分辨率革命”。

近年来,通过冷冻电子显微镜测定的大分子复合物的高分辨率结构的数量呈现指数级的增长趋势,同时结构解析可达到的最高分辨率也在不断提高。2014年,英国SCHERES团队通过改进电子显微镜技术,成功获得了酵母菌的线粒体核糖体大亚基的图像,分辨率达到 $0.34\text{ nm}^{[36]}$ 。2015年,清华大学施一公教授团队首次在上世界上揭示了分辨率高达 0.43 nm 的人体 γ -分泌酶的电镜结构^[37],这为理解 γ -分泌酶的工作机制及阿尔茨海默病的发病机理提供了重要基础。2016年,美国国家癌症研究所科学家发布的谷氨酸脱氢酶(分子量 334 kDa)结构的分辨率已经达到了 $0.18\text{ nm}^{[38]}$ 。2020年,德国HOLGER STAR课题组解析了脱铁蛋白(Apoferritin)的三维结构,并达到了原子级别的分辨率(0.125 nm),直接可视化蛋白质中的所有原子(包括氢原子)^[39],打破了之前日本KEIICHI NAMBA团队保持的 0.154 nm 的记录,实现了世界首次可视化蛋白质中的单个原子。相信随着冷冻电子显微镜技术的发展,这个纪录将会被不断打破。

综上所述,突破分辨率限制的生物分子一般有着分子量大、化学性质稳定、结构对称性好等特点,而对于小分子量蛋白质(一般为小于 200 kDa)颗粒由于其在冷冻样品中衬度不足等原因,其高分辨解析工作对目前的技术手段而言仍然是很大的挑战。但是这项技术壁垒随着冷冻电子显微镜单颗粒分析技术(Single Particle Analysis, SPA)的发展逐渐被打破。2019年,FAN等^[40]解析的分子量为 52 kDa 的链霉亲和素蛋白的分辨率达到 0.32 nm 。该研究成果一度刷新利用冷冻电子显微镜单颗粒方法解析至近原子分辨率生物大分子分子量最小记录。2022年3月,斯坦福大学医学院ZHANG等^[41]通过设计蛋白质双壳来克服尺寸限制,将环磷酸腺苷效应元件结合蛋白(CAMP-response element binding protein, CREB)中结合蛋白(CREB binding protein, CBP)的KIX结构域(11 kDa)夹在Apoferritin内壳和麦芽糖结合蛋白外壳的中间,构成分子笼,并获得了分辨率为 0.26 nm 的三维结构。这项研究表明,将小分子量蛋白质融合为笼状结构是打破冷冻电子显微镜结构测定的分辨率极限的可行方案,但是该方法的适用性仍需要通过实验进一步验证。

随着时间的推移,冷冻电子显微镜技术已经向4个方向上扩展,即电子二维晶体学(Electron 2D Crystallography)^[42]、单颗粒分析(SPA)^[43]、冷冻电子断层扫描(Cryo-Electron Tomography, Cryo-ET)^[44]以及最新的微晶电子衍射(Microcrystal Electron Diffraction, MicroED)^[45]。目前占据主导位置的是单颗粒分析技术,也是推动“分辨率革命”的主要方法。通过使用重组颗粒方法,例如洗涤剂胶束^[46]或纳米圆盘^[47],单颗粒冷冻电镜已经能够克服旧方法在结构异质性方面的困难^[48]。而随着近些年聚焦离子束(Focused Ion Beam, FIB)减薄技术和相位板^[49]等技术的发展,以及光镜-电镜结合实现样品定位策略的应用,使得更大尺度上的冷冻电子断层成像技术逐渐发展成熟。Cryo-ET的适用尺度非常广泛,包括从分子水平的蛋白质到亚细胞水平的细胞器,以至细胞水平的组织结构,都可以被观察到,这项技术也已经被成功应用于直接观察病理条件下细胞及组织内部的目标大分子复合物的定位、互作及引发疾病的构象变化机制。Cryo-ET有效填补了X射线晶体学、核磁共振以及SPA等方法得到的高精度结构和光学显微镜技术得到的低分辨率的细胞整体图像之间的空白,已经成为连接结构生物学和细胞生物学的重要桥梁^[27]。

冷冻电子显微镜解析蛋白质三维结构方法的优点是可直接获得生物大分子在接近生理条件下的结构及构象变化、无需结晶、所需样品量相对较少、研究对象广泛,适合解析大分子复合物(分子量大于 80 kDa 的蛋白质、病毒及复合物)的三维精细结构,且样品分子量越大、对称性越高,越容易区分异质性,从而越容易获得高分辨率结构^[33]。此外,作为当前结构生物学领域最前沿的成像技术之一,冷冻电镜技术已经将生物大分子复合物的结构解析能力拓展至原子分辨率水平。“分辨率革命”使Cryo-EM成为结构引导药物设计领域的未来工具^[50-52],在膜蛋白结构解析和小分子药物及疫苗的开发中显露出巨大的潜力^[27]。

尽管仍然面临一些挑战,如冷冻电子显微镜对于样品制备的要求较高,特别是分子量较小、均一性差的样品,获得的分辨率较低,设备及实验成本昂贵等,但是科技的突破还在不断进行,冷冻电子显微镜技术未来可期^[14]。

4 AlphaFold 人工智能预测蛋白质结构

由于上述方法实验测定过程十分复杂,且成本高、费时,对某些不易结晶的蛋白质还存在局限,因此,随

着计算机技术的发展,以生物信息学为基础发展起来的蛋白质结构预测方法逐渐受到重视。

蛋白质结构预测的方法主要分为两类,第一类为理论分析方法,即通过理论计算进行结构预测,该类方法假设折叠后的蛋白质取能量最低的构象,理论上是可行的,但是从实际出发,该方法存在缺陷:1)天然状态与未折叠状态下的蛋白质结构之间的能垒很小,结构变化细微;2)蛋白质可能的空间构象数量庞大,针对蛋白质折叠方式的计算量巨大;3)计算模型中参数的不确定性,这些缺点限制了理论分析方法的发展。第二类为统计方法,实质上是根据蛋白质的编码基因推测出可能的氨基酸序列,之后对已知蛋白质结构进行统计分析,建立序列到结构的映射模型,根据已掌握的蛋白质的结构特点推测未知蛋白质的可能结构,并以此结构为基础,预测该蛋白质的生物学功能^[2]。目前蛋白质结构预测的方法主要以第二类为主。

20世纪70年代,美国科学家 ANFINSEN 提出蛋白质的氨基酸序列决定其三维结构的概念,成为开展蛋白质结构预测的基本信念^[53]。也正是由此开始,计算科学家开始进入蛋白质结构预测领域,并建立计算机模型,以预测给定的蛋白质如何折叠。起初建立的模型只适用于小分子量蛋白质或大分子量蛋白质的短片段,此外,由于前期实验方法解析蛋白质结构的进度缓慢,无法提供足够多的可供参考的蛋白质三维结构模型,导致基于算法的蛋白质结构预测也是进展缓慢。直到20世纪90年代,X射线晶体学、冷冻电子显微镜等技术的突飞猛进使得 PDB 数据库中的蛋白质三维结构数量与日俱增,极大推动了计算生物学的发展。1994年,美国科学家 MOULT 和他的同事共同发起了两年一次的国际蛋白质结构预测竞赛(CASP)。组织方提供大约一百种已经通过实验方法解析出三维结构但未向公众发布的蛋白质的氨基酸序列,参赛者通过算法预测蛋白质结构,之后将预测结果与实验结果进行比较并评分(Global Distance Test,GDT),只有大于90分才被认为该算法与实验方法相比有竞争力^[54]。此后的二十多年间,竞赛中出现的最高的计算机测试成绩也只达到了40分,直到2018年,谷歌公司旗下的 DeepMind 公司研发出的 AlphaFold 首次参加 CASP,并获得了历史最高的近80分的好成绩,但是预测准确度仍有一定差距。时隔两年,在2020年的第14届 CASP 中,DeepMind 团队拿出升级版的 AlphaFold2^[55]系统,竞赛评分达到了92.4分,在蛋白质结构预测方面表现出了无与伦比的准确性,甚至可以分析X射线晶体学难以解决的存在跨膜结构域的蛋白质结构。AlphaFold2 系统优越的表现引起了科学界的热切关注。

2021年7月,DeepMind 团队在 Nature 杂志上连续发文,首先公布了 AlphaFold2 的源代码,描述 AlphaFold2 是基于神经网络的全新设计的 AlphaFold 版本,其预测的蛋白质结构能达到原子水平的准确度^[56](图5)。随后,该团队用 AlphaFold 完成了98.5%的人类蛋白质的结构预测,并且对其中35.7%的蛋白质结构预测达到了极高准确度^[57]。AlphaFold2 的成功可归因于其神经网络的架构和参考了实验方法解析的蛋白质三维结构的训练程序^[58],即基于深度学习的蛋白质结构预测方法,这在结构生物学界引发了巨大的冲击^[58]。同年7月,DAVID BAKER 领导一支计算生物学家团队,基于 AlphaFold 深度学习的基础,成功开发一款名为 RoseTTAFold 的工具^[59],能够根据有限的信息快速准确地预测出目标蛋白质的结构,达到与 AlphaFold2 不相上下的准确度。RoseTTAFold 所需的计算耗能与计算时间均比 AlphaFold2 低,仅用一台游戏计算机,在短短十分钟内就可以可靠地计算出蛋白质结构。2021年11月,Science 杂志公布了2021年的年度科学突破榜单,AlphaFold 和 RoseTTAFold 两种基于人工智能预测蛋白质结构的技术位列榜首^[60],这充分肯定了蛋白质结构预测的光明前景。

虽然人工智能预测蛋白质三维结构有着传统实验方法无法企及的优势,如耗时短、无需对蛋白质进行结

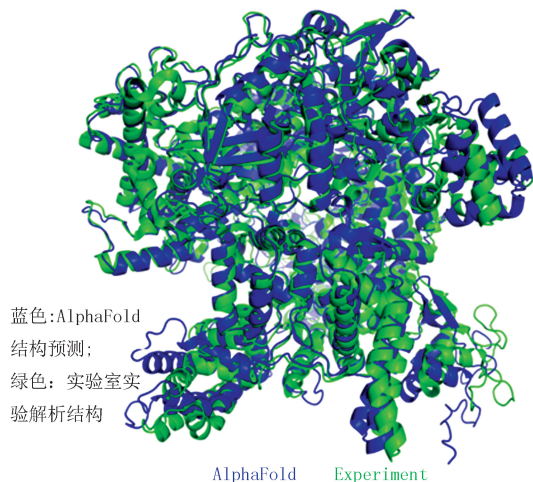


图5 CASP比赛中T1044(PDB 6VR4)——一种有2180个氨基酸残基的蛋白的AlphaFold结构预测^[56]

Fig. 5 T1044(PDB 6VR4) in CASP competition-AlphaFold structure prediction of a protein with 2180 amino acid residues^[56]

晶、无需使用昂贵的实验仪器及材料进行研究;以及源代码的免费开放意味着它是真正适合所有人的蛋白质预测模型,但仍不能否定实验方法的重要性。因为首先需要通过实验方法积累足够多的蛋白质三维结构,才能使得基于算法的 AI 结构预测有充分的前提条件进行学习模拟;其次,AlphaFold 预测蛋白质结构也并非全能的,对于数据库中未能解析的蛋白质柔性部位,使用软件也不能获得准确的结构。AI 所做的工作还处于科学实验的辅助阶段,它的可信度还需要实验来证实。

蛋白质的结构预测将有助于为基础生物学提供新的见解,特别对理论假说的提出具有启示和借鉴意义^[15],此外还有助于揭示具有临床意义的新药靶点^[1],它会与 X 射线衍射、核磁共振和冷冻电子显微镜等实验方法一起,成为解析蛋白质结构的强大工具,在科学的道路上推进对人类自身的认识和对生命的理解。

5 展 望

人们对于蛋白质结构的深入研究对揭示生命体生化反应的发生机制、研发靶向性药物有着重要意义。X 射线衍射晶体学、核磁共振、冷冻电子显微镜三大实验分析技术对于蛋白质结构解析的重要性无需赘言,并且随着科学技术的不断创新,各项技术的壁垒逐渐被打破,彼此之间的界限也变得越来越模糊,这无疑令人振奋的现象,预示着蛋白质结构解析方法高速发展时期的到来。三大实验技术与 AlphaFold 人工智能系统相辅相成,已经成为蛋白质结构解析领域最强有力的工具,在未来也必将引领结构生物学的发展,向人们揭示生命的奥秘。

参 考 文 献

- [1] HUMPHREYS I R, PEI J M, BAEK M, et al. Computed structures of core eukaryotic protein complexes[J]. *Science*, 2021, 374(6573): eabm4805.
- [2] 常爱英, 王明义. 生物信息学在预测未知蛋白结构与功能中的应用[J]. *医学检验与临床*, 2013, 24(2): 53-55.
CHANG A Y, WANG M Y. Application of bioinformatics in predicting the structure and function of unknown proteins[J]. *Medical Laboratory Science and Clinics*, 2013, 24(2): 53-55.
- [3] LUCHINAT E, BIANCI L. In-cell NMR: from target structure and dynamics to drug screening[J]. *Current Opinion in Structural Biology*, 2022, 74: 102374.
- [4] 林亚静, 刘志杰, 龚为民. 蛋白质结构研究[J]. *生命科学*, 2007, 19(3): 289-293.
LIN Y J, LIU Z J, GONG W M. The research of protein structure[J]. *Chinese Bulletin of Life Sciences*, 2007, 19(3): 289-293.
- [5] 熊强, 丁立新, 姜晓燕, 等. X 射线测定蛋白质结构的技术进展与研究现状[J]. *癌变·畸变·突变*, 2019, 31(1): 82-85.
XIONG Q, DING L X, JIANG X Y, et al. Technical progress and research status of X-ray measurement of protein structure[J]. *Carcinogenesis, Teratogenesis & Mutagenesis*, 2019, 31(1): 82-85.
- [6] KENDREW J C, DICKERSON R E, STRANDBERG B E, et al. Structure of myoglobin: a three-dimensional Fourier synthesis at 2 Å resolution[J]. *Nature*, 1960, 185(4711): 422-427.
- [7] PERUTZ M F, ROSSMANN M G, CULLIS A F, et al. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis[J]. *Nature*, 1960, 185(4711): 416-422.
- [8] CASPAR D L, KLUG A. Physical principles in the construction of regular viruses[J]. *Cold Spring Harbor Symposia on Quantitative Biology*, 1962, 27: 1-24.
- [9] BLAKE C C F, KOENIG D F, MAIR G A, et al. Structure of hen egg-white lysozyme: a three-dimensional Fourier synthesis at 2 Å resolution[J]. *Nature*, 1965, 206(4986): 757-761.
- [10] PHILLIPS D C. The three-dimensional structure of an enzyme molecule[J]. *Scientific American*, 1966, 215(5): 78-90.
- [11] 赵璇, 张文凯. X 射线自由电子激光: 原理、现状及应用[J]. *现代物理知识*, 2019, 31(2): 47-52.
- [12] ZHANG H H, FILIPPONI A, DI CICCO A, et al. Multiple-edge XAS studies of synthetic iron-copper bridged molecular assemblies relevant to cytochrome c oxidase. structure determination using multiple-scattering analysis with statistical evaluation of errors[J]. *Inorganic Chemistry*, 1996, 35(17): 4819-4828.
- [13] SAYERS Z, AVŞAR B, CHOLAK E, et al. Application of advanced X-ray methods in life sciences[J]. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 2017, 1861(1): 3671-3685.
- [14] AKBAR S, MOZUMDER S, SENGUPTA J. Retrospect and prospect of single particle cryo-electron microscopy: the class of integral membrane proteins as an example[J]. *Journal of Chemical Information and Modeling*, 2020, 60(5): 2448-2457.
- [15] 赵云波. AI 预测可以代替科学实验吗?: 以 AlphaFold 破解蛋白质折叠难题为中心[J]. *医学与哲学*, 2021, 42(6): 17-21.

- ZHAO Y B.Can AI prediction replace scientific experiments?:focusing on AlphaFold cracking the protein folding puzzle[J].*Medicine and Philosophy*,2021,42(6):17-21.
- [16] PURCELL E M,TORREY H C,POUND R V.Resonance absorption by nuclear magnetic moments in a solid[J].*Physical Review*,1946,69(1/2):37-38.
- [17] RIGDEN J S.Quantum states and precession;the two discoveries of NMR[J].*Reviews of Modern Physics*,1986,58(2):433-448.
- [18] 施蕴渝,吴季辉.核磁共振波谱应用于结构生物学的研究进展[J].*生物物理学报*,2007,23(4):240-245.
- SHI Y Y,WU J H.Progress in nuclear magnetic resonance spectroscopy used to structural biology[J].*Acta Biophysica Sinica*,2007,23(4):240-245.
- [19] PELLECCIA M,SEM D S,WÜTHRICH K.Nmr in drug discovery[J].*Nature Reviews Drug Discovery*,2002,1(3):211-219.
- [20] ALDERSON T R,KAY L E.Unveiling invisible protein states with NMR spectroscopy[J].*Current Opinion in Structural Biology*,2020,60:39-49.
- [21] 蔡聪.基于核磁共振波谱的蛋白质 CGL2373 与 TGIF1 的结构与功能研究[D].武汉:华中农业大学,2020.
- CAI C.Study on the structure and function of protein CGL2373 and TGIF1 based on NMR spectroscopy[D].Wuhan:Huazhong Agricultural University,2020.
- [22] FESIK S W,ZUIDERWEG E R.Heteronuclear three-dimensional NMR spectroscopy of isotopically labelled biological macromolecules [J].*Quarterly Reviews of Biophysics*,1990,23(2):97-131.
- [23] PERVUSHIN K,BRAUN D,FERNÁNDEZ C,et al.[¹⁵N,¹H]/[¹³C,¹H]-TROSY for simultaneous detection of backbone ¹⁵N-¹H,aromatic ¹³C-¹H and side-chain ¹⁵N-¹H₂ correlations in large proteins[J].*Journal of Biomolecular NMR*,2000,17(3):195-202.
- [24] LEE M S,GIPPERT G P,SOMAN K V,et al.Three-dimensional solution structure of a single zinc finger DNA-binding domain[J].*Science*,1989,245(4918):635-637.
- [25] TAKEUCHI K,WAGNER G.NMR studies of protein interactions[J].*Current Opinion in Structural Biology*,2006,16(1):109-117.
- [26] VAYNBERG J,QIN J.Weak protein-protein interactions as probed by NMR spectroscopy[J].*Trends in Biotechnology*,2006,24(1):22-27.
- [27] 许文青,丛尧,王权.冷冻电镜:新时代蛋白质科学和药物研发的利器[J].*张江科技评论*,2020(3):43-45.
- [28] DE ROSIER D J,KLUG A.Reconstruction of three dimensional structures from electron micrographs[J].*Nature*,1968,217(5124):130-134.
- [29] 冷冻电镜:一个发给了物理学家的诺贝尔化学奖,奖励他们帮助了生物学家|“知识分子”特别报道[EB/OL].(2017-10-04)[2022-05-20].<https://www.zhihu.com/question/66202533>.
- [30] HENDERSON R,BALDWIN J M,CESKA T A,et al.Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy[J].*Journal of Molecular Biology*,1990,213(4):899-929.
- [31] FRANK J,RADERMACHER M,PENCZEK P,et al.SPIDER and WEB:processing and visualization of images in 3D electron microscopy and related fields[J].*Journal of Structural Biology*,1996,116(1):190-199.
- [32] DUBOCHET J,ADRIAN M,CHANG J J,et al.Cryo-electron microscopy of vitrified specimens[J].*Quarterly Reviews of Biophysics*,1988,21(2):129-228.
- [33] 张世超,欧阳燕,刘善辉,等.冷冻电镜单颗粒技术解析生物大分子结构综述[J].*生物学杂志*,2017,34(3):74-77.
- ZHANG S C,OUYANG Y,LIU S H,et al.A brief introduction to single-particle cryo-EM of biological macromolecules[J].*Journal of Biology*,2017,34(3):74-77.
- [34] MILNE J L S,BORGNIA M J,BARTESAGHI A,et al.Cryo-electron microscopy:a primer for the non-microscopist[J].*The FEBS Journal*,2013,280(1):28-45.
- [35] LIAO M F,CAO E H,JULIUS D,et al.Structure of the TRPV1 ion channel determined by electron cryo-microscopy[J].*Nature*,2013,504(7478):107-112.
- [36] VOORHEES R M,FERNÁNDEZ I S,SCHERES S H W,et al.Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution [J].*Cell*,2014,157(7):1632-1643.
- [37] SUN L F,ZHAO L Y,YANG G H,et al.Structural basis of human γ -secretase assembly[J].*Proceedings of the National Academy of Sciences of the United States of America*,2015,112(19):6003-6008.
- [38] BORGNIA M J,BANERJEE S,MERK A,et al.Using cryo-EM to map small ligands on dynamic metabolic enzymes:studies with glutamate dehydrogenase[J].*Molecular Pharmacology*,2016,89(6):645-651.
- [39] YIP K M,FISCHER N,PAKNIA E,et al.Atomic-resolution protein structure determination by cryo-EM[J].*Nature*,2020,587(7832):157-161.
- [40] FAN X,WANG J,ZHANG X,et al.Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution[J].*Nature Communications*,2019,10(1):1-11.
- [41] ZHANG K M,HORIKOSHI N,LI S S,et al.Cryo-EM,protein engineering,and simulation enable the development of peptide therapeutics against acute myeloid leukemia[J].*ACS Central Science*,2022,8(2):214-222.
- [42] WALZ T,GRIGORIEFF N.Electron crystallography of two-dimensional crystals of membrane proteins[J].*Journal of Structural Biology*,

- 1998,121(2):142-161.
- [43] FRANK J.Single-particle reconstruction of biological macromolecules in electron microscopy:30 years[J].Quarterly Reviews of Biophysics,2009,42(3):139-158.
- [44] LUCIĆV,FÖRSTER F,BAUMEISTER W.Structural studies by electron tomography:from cells to molecules[J].Annual Review of Biochemistry,2005,74:833-865.
- [45] NANNENGA B L,BU G H,SHI D.The evolution and the advantages of MicroED[J].Frontiers in Molecular Biosciences,2018,5:114.
- [46] GARCÍA-NAFRÍA J,LEE Y,BAI X C,et al.Cryo-EM structure of the adenosine a(2A)receptor coupled to an engineered heterotrimeric G protein[J].eLife,2018,7:e35946.
- [47] ALAM A,KOWAL J,BROUDE E,et al.Structural insight into substrate and inhibitor discrimination by human P-glycoprotein[J].Science,2019,363(6428):753-756.
- [48] WHITE H E,IGNATIOU A,CLARE D K,et al.Structural study of heterogeneous biological samples by cryoelectron microscopy and image processing[J].BioMed Research International,2017,2017:1032432.
- [49] VON LOEFFELHOLZ O,PAPAI G,DANEV R,et al.Volta phase plate data collection facilitates image processing and cryo-EM structure determination[J].Journal of Structural Biology,2018,202(3):191-199.
- [50] CESKA T,CHUNG C W,COOKE R,et al.Cryo-EM in drug discovery[J].Biochemical Society Transactions,2019,47(1):281-293.
- [51] RENAUD J P,CHARI A,CIFERRI C,et al.Cryo-EM in drug discovery:achievements,limitations and prospects[J].Nature Reviews Drug Discovery,2018,17(7):471-492.
- [52] SUBRAMANIAM S,EARL L A,FALCONIERI V,et al.Resolution advances in cryo-EM enable application to drug discovery[J].Current Opinion in Structural Biology,2016,41:194-202.
- [53] ANFINSEN C B.Principles that govern the folding of protein chains[J].Science,1973,181(4096):223-230.
- [54] SERVICE R F.The game has changed:AI triumphs at solving protein structures[EB/OL].(2020-11-30)[2022-05-20].<https://www.science.org/news/2020/11/game-has-changed-ai-triumphs-solving-protein-structures>.
- [55] CRAMER P.AlphaFold2 and the future of structural biology[J].Nature Structural & Molecular Biology,2021,28(9):704-705.
- [56] JUMPER J,EVANS R,PRITZEL A,et al.Highly accurate protein structure prediction with AlphaFold[J].Nature,2021,596(7873):583-589.
- [57] TUNYASUVUNAKOOL K,ADLER J,WU Z,et al.Highly accurate protein structure prediction for the human proteome[J].Nature,2021,596(7873):590-596.
- [58] Method of the Year 2021:protein structure prediction[J].Nature Methods,2022,19(1):1.
- [59] BAEK M,DIMAIO F,ANISHCHENKO I,et al.Accurate prediction of protein structures and interactions using a three-track neural network[J].Science,2021,373(6557):871-876.
- [60] THORP H H.Proteins,proteins everywhere[J].Science,2021,374(6574):1415.

Methods for the analysis of the three-dimensional structure of proteins

Xu Jiaqiang¹, Cui Qianqian^{1,2}

(1. Department of Chemistry, College of Science; Novel Energy and Sensing Technology Lab, Shanghai University, Shanghai 200444, China; 2. Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China)

Abstract: Three-dimensional structure is the basis of functional analysis of biological macromolecules. In-depth study of the atomic framework, and determination of structure, function and evolutionary roots of biological macromolecules are helpful to guide the treatment of diseases and the development of target drugs. Protein is the material undertaker and embodiment of life activities. The structure analysis is the premise to clarify the function and action mechanism of protein. This paper briefly summarizes four existing methods of protein structure from the aspects of principle, development process, advantages and disadvantages, for example X-ray diffraction, nuclear magnetic resonance, Cryo-electron microscope and AlphaFold, which is an artificial intelligence system for protein structure prediction based on big data, in order to review the development history of structural biology and systematically sort out the current development status of protein structure analysis.

Keywords: three-dimensional structure of protein; X-ray diffraction; nuclear magnetic resonance; Cryo-electron microscope; AlphaFold