

News & Highlights

## AlphaFold 获得诺贝尔奖后功能增强, 却终止开放获取

Chris Palmer

Senior Technology Writer

2024年10月9日, 瑞典皇家科学院宣布将本年度诺贝尔化学奖授予谷歌DeepMind公司(英国伦敦)的联合创始人兼首席执行官Demis Hassabis及总监John M. Jumper, 以表彰两人开发人工智能(AI)驱动的蛋白质结构预测模型AlphaFold2(AF2)的突破性贡献[1]。另一位获奖者是华盛顿大学(美国西雅图)生物化学教授David Baker。该奖项由Hassabis、Jumper与Baker共享——前两人共得一半奖金, Baker独得另一半。Baker的贡献集中在计算蛋白质设计领域: 自20世纪90年代中期起, 他便致力于研究该方向, 主导开发了Rosetta软件套件。该套件基于物理原理构建蛋白质结构模型[2], 历经迭代升级后现已整合AI技术[3]。

自2021年7月以开放获取的方式发布以来[4], AF2推动了结构生物学领域的跨越式发展, 并为药理学开辟了新路径。美国范德比尔特大学(Vanderbilt University, 位于田纳西州纳什维尔市)分子生理学与生物物理学助理教授Stephanie Wankowicz表示: “学界普遍认为, AlphaFold迟早会获得诺贝尔奖。这一技术彻底改变了我们理解蛋白质结构的研究方式。”

此外, 2024年5月8日, DeepMind公司在《自然》期刊上发布了新一代蛋白质结构预测模型AlphaFold3(AF3)及其研究成果[5]。该模型取得重大突破, 能够预测蛋白质与非蛋白质分子(如DNA、RNA)相互作用时的复合结构, 这一能力对阐明蛋白质在细胞中的特定功能至关重要(图1)[5]。除深化对细胞动力学的认知外, AF3还将助

力科学家设计能更有效阻断或增强疾病相关蛋白功能的药物[6]。



图1. DeepMind公司最新研发的AI蛋白质结构预测模型AF3能够预测蛋白质与非蛋白质分子(如DNA、RNA)的结构互动。图中展示了DNA分子(底部)与信号蛋白Clr-3'、5'-环磷酸腺苷(cAMP, 顶部)形成的复合物结构。图片来源: 蛋白质数据库(公共领域)。

在AF2问世前的数十年间, 解析蛋白质三维结构的唯一途径是依靠实验技术手段, 如核磁共振成像、X射线晶体学和冷冻电子显微镜等。但这些方法不仅耗时费力, 且因所需设备极其昂贵, 多数科研人员难以接触使用。此外, 许多蛋白质并不适合借助这些工具进行研究。

到20世纪90年代中期, 科学家已研发出多种蛋白质结构预测计算方法。为系统评估各类预测模型的效能, 国际学界于1994年创立了两年一度的“蛋白质结构预测关键评估竞赛”(CASP)[7]。自首届竞赛以来, 相关模型

的预测精度虽持续提升但进展缓慢，直至2018年DeepMind携初代AlphaFold参赛并以显著优势获胜[8]。在2020年举办的下一届CASP竞赛中，DeepMind公司推出的全面升级版AF2更以压倒性优势再次击败所有竞争对手[7,9]。

初代AlphaFold模型采用分步方式预测蛋白质结构，而AF2则通过神经网络实现了全流程一体化预测——该架构的性能可随着训练数据的增加持续优化。2021年7月，DeepMind公司在《自然》期刊上发表了详细介绍AF2的报告[4,9]，同时于GitHub平台共享了该模型的全部源代码和权重参数，供全球科研人员免费获取与使用[10]。

AF2迅速展现了其重要价值。2021年7月，在《自然》期刊发表相关报告的同期，DeepMind公司与欧洲分子生物学实验室合作建立了AlphaFold蛋白质结构数据库，在其初始阶段收录了21种模式生物的蛋白质结构数据[11]。至2022年1月，数据库新增27种模式生物的蛋白质结构数据[9]。到2022年年中，该数据库已涵盖约100万个物种，预测结构总数突破2.2亿个[12-13]。

得益于这一开放策略，全球研究人员能够基于该平台持续开展创新研究。例如，AF2早期的一项新增功能允许用户预测多种蛋白质之间的相互作用[14]。2023年3月16日，DeepMind公司在社交媒体平台X（原Twitter）上宣布，已将该功能整合至AF2的更新版本中。

美国威斯康星大学麦迪逊分校生物统计学与医学信息学副教授Anthony Gitter表示：“AlphaFold2在2020年CASP中惊艳亮相后，随即在计算科学领域及多学科（包括遗传学、结构生物学、化学等）湿实验研究群体中引发广泛关注。这一技术革新不仅推动了相关领域的理论发展，更加速了实验研究的范式转变。尤为重要的是，科研人员依托这一核心算法平台，通过创新性的二次开发，实现了诸多超越原始设计框架的应用突破——这充分彰显了科学软件开源共享的独特价值。”

英国伦敦大学学院（University College London）生物信息学教授David Jones是首篇AlphaFold论文的合著者之一[8]。他指出，与AF2相比，AF3在蛋白质折叠预测方面的改进有限。他表示，两者的核心差异在于结构生成机制：“AlphaFold2采用专为蛋白质设计的手动编码方法，而AlphaFold3则借助扩散过程生成结构，这使得蛋白质和其他分子能够以完全相同的方式进行建模，将整个系统视为不同类型原子的集合体。”但Jones特别强调：“由于扩散过程难以有效维持手性等关键特征，且无法完全避免原子重叠，因此可能导致AlphaFold3在某些蛋白质结构预测中的表现逊于AlphaFold2。”

相较于学界对AF2的广泛赞誉[13]，AF3的发布引发了科研界的显著质疑[15]。不同于AF2的完全开源，AF3最初仅允许通过DeepMind公司网站服务器进行非商业性访问[15]。此外，DeepMind公司决定不公开AF3的代码。DeepMind公司在其发表于《自然》期刊的AF3相关报告中并未对此举作出解释，仅简单注明“未提供代码”——这一做法似乎违反该期刊对科研成果开放性的政策要求[15]。Gitter表示：“学术界的强烈反应表明这一决定令人始料未及，毕竟前代模型是完全开放的。”

DeepMind公司还对AF3服务器的访问权限施加了严格限制：用户每日仅可进行20次预测[15]（初始限额为10次），且可分析的分子类型也受到限制[6]。例如，用户无法使用AF3服务器预测蛋白质与新型药物之间的相互作用，此举或许是为了避免与由Hassabis创立的DeepMind衍生公司——位于英国伦敦的Isomorphic Labs——在药物研发领域形成竞争[15]。Jones表示：“这显然是商业考量而非科学决策。AlphaFold3网络服务器所限制的功能，恰恰是模拟药物分子与蛋白质结合所必需的核心功能。”

2024年5月11日，即《自然》期刊发表AF3报告但未公开其代码后的首周内，一封公开信获得了650余名学者的签名[16]。包括Wankowicz和Gitter在内的联名学者在信中强调：“尽管企业有权从其创新中获利，但借学术出版之便却既不提供结果复现可能，更阻碍后续研究拓展，此举有悖于学术研究的宗旨。”[16]

公开信发布后不久，DeepMind公司研究团队便表示，将陆续公布更多关于AF3的技术细节。DeepMind公司研究副总裁Pushmeet Kohli于2024年5月13日在社交媒体平台X上发布动态称，DeepMind公司正“着力推进AF3模型面向学术用途的开放工作”，预计将在六个月内实现此目标。Kohli同时解释，此前设置的使用限制确实是为了避免响其旗下公司Isomorphic Labs的商业药物研发计划[6,17]。

几乎恰好六个月后，即2024年11月11日，Kohli在X平台宣布，DeepMind公司已开放AF3代码下载，并允许非商业用途使用[18]。但此次公开程度与AF2存在显著差异。Gitter表示：“代码发布后，AlphaFold3现在更容易获取了，不过仍未完全开源，其使用条款和许可协议比AlphaFold2更为严苛。”Jones教授进一步补充：“即便获得最新发布的代码，研究人员若想复现该模型开发过程，仍将面临重重阻碍。”

鉴于AF3的发布方式，Gitter预计其影响力将弱于AF2。他表示：“AlphaFold2问世后迅速成为学界主导工具，而当前呈现的却是研究格局的显著分化。”虽然部分研究人员接受每日20次的服务器限额，但“更多团队正全力重构

AF3 代码”，他提到，全球多个研究组都在致力于开发 AF3 的开源替代版本[18]。此外，部分研究人员已转向采用替代工具，如 Baker 团队开发的 RoseTTAFold All-Atom [19]。

尽管 DeepMind 公司对 AF3 采取了更为严格的限制措施，但 Jones 仍认为该公司的研究成果影响深远，其价值足以迅速获得诺贝尔奖。他指出：“AlphaFold 极大地提升了计算生物学和机器学习的关注度，尤其是在更广泛的生物学界。虽然 AlphaFold 并未真正消除结构生物学对于实验工作的需求，但它至少引发了‘未来是否可能实现’的讨论——这本身就标志着该领域思维方式的重大转变。”

## References

- [1] Heikkilä M. Google DeepMind leaders share Nobel Prize in Chemistry for protein prediction AI [Internet]. Cambridge: MIT Technology Review; 2024 Oct 9 [cited 2024 Nov 2]. Available from: <https://www.technologyreview.com/2024/10/09/1105335/google-deepmind-wins-joint-nobel-prize-in-chemistry-for-protein-prediction-ai/>.
- [2] Leman JK, Weitzner BD, Lewis SM, Adolf-Bryfogle J, Alam N, Rebecca FA, et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods* 2020;17:665–80.
- [3] Baek M, Dimairo F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 2021;373(6557):871–6.
- [4] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–9.
- [5] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;630:493–500.
- [6] Callaway E. Major AlphaFold upgrade offers boost for drug discovery [Internet]. London: Nature; 2024 May 8 [cited 2024 Oct 19]. Available from: <https://www.nature.com/articles/d41586-024-01383-z>.
- [7] O’Neill S. Artificial intelligence cracks a 50-year-old grand challenge in biology. *Engineering* 2021;7(6):706–8.
- [8] Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. *Nature* 2020; 577:706–10.
- [9] O’Neill S. Machine learning turbocharges structural biology. *Engineering* 2021; 7(6):706–8.
- [10] Hassabis D. Putting the power of AlphaFold into the world’s hands [Internet]. London: DeepMind; 2022 Jul 22 [cited 2024 Nov 2]. Available from: <https://deepmind.google/discover/blog/putting-the-power-of-alphafold-into-the-worlds-hands/>.
- [11] Hatch V. DeepMind and EMBL release the most complete database of predicted 3D structures of human proteins [Internet]. Heidelberg: European Molecular Biology Laboratory’s European Bioinformatics Institute; 2021 Jul 22 [cited 2024 Nov 2]. Available from: <https://www.ebi.ac.uk/about/news/announcements/alphafold-database-launch/>.
- [12] Callaway E. “The entire protein universe”: AI predicts shape of nearly every known protein [Internet]. London: Nature; 2022 Jul 28 [cited 2024 Oct 19]. Available from: <https://www.nature.com/articles/d41586-022-02083-2>.
- [13] Leslie M. Artificial intelligence could revolutionize science—if we can trust it. *Engineering* 2024;35:4–6.
- [14] Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* 2022;13:1265.
- [15] Offord C. Limits on access to DeepMind’s new protein program trigger backlash [Internet]. Washington, DC: Science; 2024 May 15 [cited 2024 Oct 19]. Available from: <https://www.science.org/content/article/limits-access-deepmind-s-new-protein-program-trigger-backlash>.
- [16] Wankowicz S, Beltrao P, Cravatt B, Dunbrack R, Gitter A, Lindorff-Larsen K, et al. AlphaFold3 transparency and reproducibility [Internet]. Meyrin: Zenodo; 2024 May 13 [cited 2024 Oct 19]. Available from: <https://zenodo.org/records/11391920>.
- [17] Buntz B. The latest on Isomorphic Labs’ plans to use AI models to revamp drug discovery [Internet]. Cleveland: Drug Discovery & Development; 2024 Aug 30 [cited 2024 Oct 19]. Available from: <https://www.drugdiscoverytrends.com/the-latest-on-isomorphic-labs-plans-to-use-ai-models-to-revamp-drug-discovery/>.
- [18] Callaway E. AI protein-prediction tool AlphaFold3 is now open source [Internet]. London: Nature; 2024 Nov 12 [cited 2024 Nov 13]. Available from: <https://www.nature.com/articles/d41586-024-03708-4>.
- [19] Krishna R, Wang J, Ahern W, Sturmfels P, Venkatesh P, Kalvet I, et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* 2024;384(6693):ead12528.