

报告标题：AI 赋能生物医药：从蛋白质折叠到药物研发的革命

资料来源：[硅谷101](#)

1. 引言：AI 入侵生物医药领域

人工智能（AI）正以前所未有的速度和深度重塑生物医药领域，如同“入侵”一般，带来颠覆性的变革。2024 年诺贝尔化学奖授予 DeepMind 的 Demis Hassabis 和 John Jumper，以及华盛顿大学的 David Baker，表彰他们在蛋白质结构预测和计算蛋白质设计方面的开创性工作，这标志着 AI 与生物医药的跨界融合得到了高度认可，也预示着一个新时代的到来。正如《硅谷101》节目中提到的，生物学研究的复杂性（从单个蛋白质到整个生物体）以及传统实验方法的局限性（耗时、昂贵），为 AI，特别是深度学习技术，提供了巨大的机遇和挑战。

2. 蛋白质折叠：生物学的核心挑战

蛋白质是生命的基础，执行着细胞和身体内几乎所有至关重要的功能。蛋白质的功能与其三维结构密切相关。蛋白质折叠问题，即如何根据氨基酸序列预测其三维结构，长期以来一直是生物学领域的重大挑战。正如《硅谷101》视频中提到的，这就像要从一串字母预测出复杂折纸作品的最终形状。传统的实验方法（如 X 射线晶体学、NMR 和冷冻电镜）虽然精确，但耗时、昂贵，且并非适用于所有蛋白质。

3. AlphaFold：颠覆蛋白质结构预测

AlphaFold，Google 旗下 DeepMind 的 AI 程序，是解决蛋白质折叠问题的关键。

- 突破性进展：
 - 2020 年 CASP 大赛中，AlphaFold 2 展现出前所未有的准确性（精度超 90%，误差仅 0.16 纳米，达原子级别），远超其他竞争对手，被《自然》杂志誉为“改变一切”。
 - AlphaFold 2 预测速度快，几天内即可完成，而传统方法需数年 and 数十万美元。
 - 首次与实验技术的精度相匹配
- 知识共享：
 - DeepMind 与 EMBL-EBI 合作建立 AlphaFold 数据库，包含数百万个预测结构。

- 到 2023 年，该数据库已包含几乎所有已知（超 2 亿个）蛋白质的预测结构，研究人员可像使用搜索引擎一样查找。
- 持续进化：
 - 2024 年，DeepMind 发布 AlphaFold 3，采用 Transformer + Diffusion 架构。
 - AlphaFold 3 不仅预测蛋白质，还能预测 DNA、RNA、小分子及其相互作用，模拟细胞过程。
 - 预测分子相互作用的精度显著提升，超越传统方法。
 - 谷歌推出免费研究平台 AlphaFold Server。
 - 2024 年 11 月，AlphaFold 3 开源，标志着技术共享的新里程碑。

4. AI 驱动的生物医药研究机构概览

除了 DeepMind，还有多个机构在 AI 驱动的生物医药研究中做出了重要贡献：

- **DeepMind**：AlphaFold 系列模型的开发者，蛋白质结构预测领域的领导者。
- **Meta AI**：开发的 ESMFold 以其预测速度快（比 AlphaFold 2 快 60 倍）和庞大的数据库（6 亿多种蛋白质结构）而闻名，但准确性通常低于 AlphaFold。
- **华盛顿大学**：David Baker 团队开发的 RosettaFold，融入了更深入的生物学知识，计算速度更快，所需算力更少。
- **欧洲生物信息学研究所（EBI）**：在数据共享和应用方面发挥重要作用。
- **加州大学伯克利分校（UC Berkeley）**：诺贝尔生理学或医学奖得主 Randy Schekman 认为 AI 将持续颠覆传统科学。
- **其他参与者**：生物技术公司（如 Insilico Medicine, imig therapeutics）和科技巨头（如英伟达、亚马逊、谷歌）积极参与，推动领域发展。

5. AI 赋能生物医药的技术发展三阶段

AI 在生物医药领域的应用经历了三个主要阶段：

1. 机器学习的简单应用（20 世纪 60 年代 - 21 世纪初）：

- 早期尝试：计算机和定量数学方法解读化学结构与药效的关系（如 QSAR）。
- 分子对接模型：模拟小分子与大分子相互作用。
- 数据库兴起：通过数据库学习结构特征。

2. 深度学习的出现（21 世纪 10 年代初）：深度学习算法极大提升了数据处理和分析能力，第一代 AI 药物发现公司出现。

3. 真正的 AI 时代（2020 年至今）：

- 端到端学习：从化学式直接预测功能，减少人为干预。
- AlphaFold 的分水岭意义：标志着 AI 技术在该领域的成熟应用。
- AI 模型快速迭代：推动生物医药研究加速发展。
- 正如受访嘉宾所说，现在已经进入了 AI for Life Science for drug Discovery 的新时代。

6. AI 赋能新药研发：合作、挑战与未来

科技巨头与生物医药公司的合作模式：投资与合作研发、平台与工具支持、数据共享与分析。AI 加速药物研发的机制：缩短研发周期并降低成本、减少实验依赖并提高效率、精准预测分子结构与相互作用、助力监管审批。

AI 面临的挑战：对实验需求的理解、科学问题的复杂性、数据的质量与可及性、伦理与监管。正如采访中提到的，AI 在制药领域的最大障碍是对实验需求的理解和对科学问题的理解。

AI 的未来趋势：端到端学习深化、跨学科合作加强、新应用场景（生物材料、农业、太空生物技术）、千亿级美元产业潜力。Fusion Fund 创始合伙人张路甚至提到，随着 Starship 发射成功，未来太空环境可能为生物医药研发提供新机遇。

7. 结论：AI 与生物医药的未来

AI 正在以前所未有的力量改变生物医药领域，尤其是在蛋白质结构预测方面取得了里程碑式的成就。以 AlphaFold 为代表的 AI 模型正在加速药物研发，降低成本，并为生命科学研究开辟新的道路。随着技术的不断进步和跨界合作的加强，AI 在生物医药领域的潜力将得到进一步释放，为人类健康带来更大的福祉。正如多位专家和行业人士所强调的，AI 正在影响生物行业的方方面面，从药物研发到临床试验，再到监管审批。医疗健康行业拥有大量高质量数据，这为 AI 的应用提供了“金矿”，而 AI 技术的进步也将反过来推动生物医药行业的创新。

8. 附录

详细时间线

- **20 世纪 60 年代:** 人工智能的早期研究，尝试用计算机和定量数学方法解读化学结构及其与药效的关系，出现“定量构效关系 (QSAR)”概念。
- **20 世纪 70-80 年代:** 化学结构数据库的建立逐渐成为研究重点。

- **20 世纪 80-90 年代:** 全球化学家和药物学家开始将化学结构及其生物活性的信息汇总到数据库中, 通过数据库学习结构特征成为主流。
- **20 世纪 90 年代:** 与分子对接 (Docking) 相关的模型建立, 如 UCSF 和牛津大学的研究。分子对接技术模拟小分子与大分子间的相互作用, 预测结合方式。
- **21 世纪初:** 机器学习和深度学习技术逐步应用于药物发现领域, 第一代 AI 药物发现公司开始出现, 利用机器学习分析药物分子结构和药效, 尝试设计新药分子。高通量筛选技术普及, 为机器学习提供大量数据。
- **2010 年代初:** 深度学习兴起, 为生物医药领域的数据处理和分析能力带来前所未有的变革。
- **2018 年 12 月:** 第 13 届 CASP 大会, DeepMind 首次推出 AlphaFold (AlphaFold v1), 在蛋白质结构预测中表现出色, 在 98 个参赛队伍中名列前茅。但其准确性尚未达到彻底改变整个领域的水平。
- **2020 年底 (11 月底至 12 月初):** 由于新冠疫情, CASP 大会首次在线上举行。DeepMind 的 John Jumper 展示了革命性的工具 AlphaFold v2 (AlphaFold R)。
- **2020 年 11 月 30 日:** John Jumper 通过 Zoom 展示 AlphaFold R 的成果, 在 3D 蛋白质结构预测方面展现出前所未有的准确性, 精度超过 90%, 远超其他竞争对手。被认为几乎一举解决了困扰科学界近 50 年的蛋白质折叠问题。《自然》杂志评价其将“改变一切”。
- **2021 年:** David Baker 教授及其团队开发出新型蛋白质结构预测工具 RoseTTAFold。
- **2021 年:** DeepMind 与欧洲生物信息学研究所 (EMBL-EBI) 合作启动 AlphaFold 数据库, 纳入 35 万个蛋白质预测结构。
- **2022 年:** Meta AI 研究团队推出 ESMFold, 一个强大的蛋白质结构预测模型, 并公布了 6 亿多种蛋白质结构预测结果, 覆盖地球环境样本中未被充分研究的蛋白质。ESMFold 在计算效率方面表现突出。
- **2023 年春季:** 作为 Meta 公司大范围裁员的一部分, ESMFold 的部门被解散, 引发学术界对数据维护的担忧。
- **2023 年 5 月至 11 月:** 英伟达频繁投资 AI 药物研发领域, 投资了 9 家 AI 治疗公司。
- **2023 年:** 谷歌云宣布与生物制药上市公司 Insitro 合作, 利用 AI 技术提高效率, 缩短新药开发和交付时间。
- **2023 年:** 亚马逊云科技宣布与生命科学行业商业服务提供商 Evensana 合作, 共同推广 AI 驱动制药等应用。
- **2023 年:** DeepMind 公布了包含几乎所有已知两亿多个蛋白质可能结构的 AlphaFold 数据库更新。

- **2024 年 5 月:** 谷歌在《自然》杂志上发表关于 AlphaFold v3 的突破性研究。AlphaFold 3 由 DeepMind 和 Isomorphic Labs (DeepMind 的拆分公司) 开发, 采用 Transformer 加 Diffusion 的架构, 能够预测蛋白质、核酸 (DNA、RNA) 和小分子的 3D 结构, 并揭示它们如何组合在一起。
- **2024 年 10 月:** 瑞典皇家科学院宣布 2024 年诺贝尔化学奖授予 DeepMind 的 Demis Hassabis 和 John Jumper, 以及华盛顿大学蛋白质设计研究所所长 David Baker, 以表彰他们在蛋白质结构预测和计算蛋白质设计方面的研究。前一天, Jeffrey Hinton 和 Yann LeCun 刚获得诺贝尔物理学奖。
- **2024 年 11 月 11 日:** 谷歌悄悄开源了 AlphaFold v3 的模型代码和权重, 供学术用途, 以帮助前沿研究。

人物列表

- **Demis Hassabis (戴密斯·哈萨比斯):** Google 旗下 DeepMind 的联合创始人兼首席执行官。因在蛋白质结构预测研究方面的贡献, 与 John Jumper 共同获得 2024 年诺贝尔化学奖。
- **John Jumper:** Google 旗下 DeepMind 的研究主管。因主导开发 AlphaFold 系列模型, 并在蛋白质结构预测研究方面做出卓越贡献, 与 Demis Hassabis 共同获得 2024 年诺贝尔化学奖。
- **David Baker (大卫·贝克):** 华盛顿大学蛋白质设计研究所所长。因在计算蛋白质设计研究方面的贡献, 获得 2024 年诺贝尔化学奖, 其团队开发的 RoseTTAFold 也是生物医药领域重要的蛋白质分析工具。
- **Jeffrey Hinton (杰弗里·辛顿):** 人工智能领域的先驱, 尤其在深度学习方面贡献卓著, 被称为“深度学习之父”。
- **John Hopfield (约翰·霍普菲尔德):** 著名物理学家和神经网络研究的先驱。
- 陈庆: “硅谷 101”节目的主持人。
- **John Moult:** CASP 大会的组织者。
- **Randy Schekman (兰迪·谢克曼):** 诺贝尔生理学或医学奖得主。
- 张路: Fusion Fund 的创始合伙人。
- **Hansch (汉斯):** 早期定量构效关系 (QSAR) 领域的代表人物。

术语表

英文	中文	含义
Artificial Intelligence (AI)	人工智能	计算机系统理论与发展的领域, 旨在使计算机系统能够执行通常需要人类智能才能完成的任务, 例如视觉感知、语音识别、决策制定以及语言之间的翻译。

英文	中文	含义
Biomedicine	生物医药	医学的一个分支，涉及将生物学和生物化学原理应用于医学研究或实践。
Protein Structure Prediction	蛋白质结构预测	根据蛋白质的氨基酸序列确定其三维结构。
Protein Folding Problem	蛋白质折叠问题	根据蛋白质的一维氨基酸序列预测其三维结构的挑战。
AlphaFold	AlphaFold	由 DeepMind 开发的用于预测蛋白质结构的人工智能程序。
CASP (Critical Assessment of protein Structure Prediction)	CASP	每两年举行一次的蛋白质结构预测领域的社区性盲测实验。
Transformer Model	Transformer 模型	一种神经网络架构，完全依赖自注意力机制来计算其输入和输出的表示，从而执行序列到序列的转换。
Diffusion Model	Diffusion 模型	一类概率生成模型，通过迭代地将随机噪声信号细化为结构化数据样本来工作。
AlphaFold Database	AlphaFold Database	由 DeepMind 和欧洲生物信息学研究所 (EMBL-EBI) 开发的，包含 AlphaFold 预测的蛋白质结构的综合数据库。
AlphaFold 3	AlphaFold 3	AlphaFold 的最新版本，能够预测各种生物分子（包括蛋白质、DNA、RNA 和小分子）的结构和相互作用。
ESMFold	ESMFold	由 Meta AI 开发的蛋白质结构预测模型，以其速度和大规模预测而闻名。
RoseTTAFold	RoseTTAFold	由 Baker 实验室开发的蛋白质结构预测计算方法，将深度学习与物理原理相结合。
QSAR (Quantitative Structure-Activity Relationship)	定量构效关系	一种用于根据化合物的结构特性来定量预测其生物活性的技术。
Docking (Molecular Docking)	分子对接	一种计算方法，用于预测小分子（例如候选药物）如何与蛋白质受体或其他大分子结合。
High-throughput Screening	高通量筛选	一种自动化方法，用于快速测试大量化合物的生物或生化活性。
End-to-end Learning	端到端学习	一种机器学习方法，其中单个神经网络学习将原始输入直接映射到输出，无需明确的中间步骤或人为设计的特征。

