

· 科技评述:2024年诺贝尔奖评述 ·

## 开物成务:2024年诺贝尔化学奖 “计算蛋白质设计与结构预测”

刘源<sup>1\*</sup> 王初<sup>1,2\*</sup>

1. 北京大学 化学与分子工程学院,北京 100871
2. 北京大学 前沿交叉学科研究院,北京 100871

**[摘要]** 2024年的诺贝尔化学奖颁发给美国华盛顿大学的科学家 David Baker,表彰他在“计算蛋白质设计”领域的杰出贡献;以及英国 DeepMind 公司的两位科学家 Demis Hassabis 和 John M. Jumper,表彰他们在“蛋白质结构预测”领域的突破性成果。

**[关键词]** 诺贝尔化学奖;蛋白质折叠;蛋白质结构预测;蛋白质序列设计;AlphaFold;Rosetta

2024年10月9日,瑞典皇家科学院诺贝尔奖委员会宣布将2024年诺贝尔化学奖颁发给美国华盛顿大学的科学家 David Baker,表彰他在“计算蛋白质设计”领域的杰出贡献;以及英国 DeepMind 公司的两位科学家 Demis Hassabis 和 John M. Jumper,表彰他们在“蛋白质结构预测”领域的突破性成果。蛋白质作为重要的分子机器,执行和维持多种生命功能。蛋白质折叠问题——即理解氨基酸的线性序列如何自发折叠成精确的三维结构,半个多世纪来以来一直是生物学最引人入胜的谜题之一<sup>[1]</sup>(图1)。在20世纪60年代末提出的 Levinthal's paradox<sup>[2]</sup>已经认识到,蛋白质理论上可以采取的构象数量极其庞大,不可能通过计算所有可能构象来预测蛋白质结构;而70年代提出的 Anfinsen's Dogma<sup>[3]</sup>指出,蛋白质的结构可以由其氨基酸序列所决定。这为蛋白质折叠研究提供了指导思想,而揭开真正的奥秘则需要生物学、物理学、化学和计算机科学等各个领域的交叉融合。

Baker 教授的实验室早期研究蛋白质的折叠速率问题,通过噬菌体展示技术发现,一些人工设计的随机序列虽然稳定性较差,但可能比天然序列折叠得更快<sup>[4]</sup>,蛋白质中接触残基之间的平均序列距离(contact order)与折叠速率密切相关<sup>[5]</sup>。这些研究使他认识到有些局部结构的片段作为孤立的肽段是



王初 北京大学化学与分子工程学院教授,北大清华—生命联合中心高级研究员,北京大学前沿交叉研究院院长。主持国家杰出青年科学基金、科技部重点研发计划等项目;获得“国际化学生物学会青年化学生物学家奖”“药明康德生命科学研究学者奖”等。其课题组主要研究方向为化学和计算驱动的功能蛋白质组学。



刘源 北京大学化学与分子工程学院专职副研究员。先后获得科技部重点研发项目、国家自然科学基金面上项目等项目的资助。曾在梁好均课题组、David Baker 课题组和王初课题组进行学习或研究。主要研究方向为生物大分子体系的建模与模拟。

稳定的,多肽链的每个片段都会根据其局部氨基酸序列采样可能的局部构象,当取得正确的结构和方向时,非极性残基被埋藏在疏水核心中,也就发生了蛋白质折叠。同时这也启发他的课题组逐步转向蛋白质结构从头预测的问题,并开发出著名的蛋白质结构预测软件 Rosetta。经过数年的算法开发和模型迭代,Rosetta 不仅在蛋白质结构预测领域取得成功,还被进一步成功地应用到蛋白质序列设计的任务<sup>[6]</sup>。“序列设计”意味着通过人为设计的序列使氨基酸链折叠到目标的蛋白结构(图1)。早在1988年,Bill DeGrado<sup>[7]</sup>首先在四螺旋束模型上证明理性设计蛋白序列的可能。1997年,Steve Mayo<sup>[8]</sup>开创

收稿日期:2024-11-01;修回日期:2024-11-01

\* 通信作者,Email: wendao@pku.edu.cn; chuwang@pku.edu.cn

性地展示可以利用计算算法和统计势能进行自动化的序列设计。2003 年, Baker 实验室的 Brian Kuhlman 利用 Rosetta 设计了 TOP7 蛋白<sup>[9]</sup>, 以原子级别精度设计出自然界从未出现过的新拓扑。随后在 2008 年设计出的两个自然界中不存在的催化功能的酶<sup>[10, 11]</sup>, 再到 2011 年可以结合流感病毒刺突蛋白保守区域的结合蛋白<sup>[12]</sup>, 一系列成功让 Baker 实验室声名鹊起, 也让 Rosetta 成为领域内广为人知的蛋白质设计软件。2016 年, 在取得了一系列从头设计蛋白的技术突破后<sup>[13-16]</sup>, David Baker 在 *Nature* 杂志上发表综述文章<sup>[17]</sup>, 宣告蛋白设计新时代的到来。

虽然计算蛋白质方法取得了巨大成功, 此时的设计成功率仍然较低, 设计序列后需要大量计算和实验筛选和验证。关键在于, 蛋白质结构预测的能力仍然不足。“结构预测”意味着根据氨基酸序列, 预测其生理条件下折叠形成的蛋白质三维结构。CASP 竞赛 (the Critical Assessment of protein Structure Prediction) 则是专门评估各种预测方法性能的“奥林匹克”。2016 年正值 CASP12 举行, 传统方法的整体表现已到达瓶颈, 而基于共进化的方法开始展现出巨大潜力。这时就要提到计算机科学家 Demis Hassabis, 他于 2010 年成立人工智能公司 DeepMind, 2016 年领导开发的 AlphaGo 在有“人类智慧最后堡垒”之称的围棋项目中战胜世界顶尖棋手<sup>[18]</sup>, 展示出人工智能技术的巨大进步。同时, DeepMind 也开始涉足医疗健康领域。2017 年, John M. Jumper 博士毕业后加入 DeepMind 公司并参与 AlphaFold 的开发项目, 他早在 2010 年就在 D. E. Shaw Research 参与蛋白质的分子动力学模拟研究<sup>[19]</sup>, 随后在芝加哥大学利用机器学习研究蛋白质折叠并获得化学硕士及博士学位。在 2018 年的 CASP13 中, AlphaFold 初露锋芒, 一举摘得桂冠<sup>[20]</sup>, 但此时所采用的方法并没有摆脱学术界的框架<sup>[21]</sup>。接下来, 在 Jumper 领导下的团队突破性引

入 transformer 模型, 使其可以更灵活地在海量数据中找到相关模式。在两年后的 CASP14 竞赛中, 横空出世的 AlphaFold2 以其极高的计算效率和惊人的预测精度震撼了世界<sup>[22]</sup>, 使蛋白质结构预测的水平在很大比例达到了实验级别的精度。随后他们开源了 AlphaFold2 算法并耗费庞大资源预测了几乎全部已知的蛋白质序列所对应的结构<sup>[23]</sup>。这对整个分子生物学和生物化学的研究都具有重大推动作用。

“成务”就是把事情办好, AlphaFold 系列工具在蛋白质结构预测领域取得巨大的成功且为设计具有特定功能的新型蛋白质提供了助推力<sup>[24]</sup>。“开物”则意味着弄清事物的规律, 神经网络确实学习到了结构与信息之间的内在联系, 但对科学家而言, 蛋白质折叠的基本原理和动力学过程仍然充满未知。这要求科学家们继续在理论和实验上进行深入探索, 以揭示蛋白质折叠背后的深层次机制, 进而推动生物医学、疾病治疗和基因工程等领域的突破。

2013 年诺贝尔化学奖颁发给“在开发多尺度复杂化学系统模型方面所做的贡献”的三位科学家, 这代表了由第一性原理出发的还原论研究范式。而 2024 年获奖的则是“连结主义”所代表的新的研究范式。人工智能作为一门多学科交叉的技术, 已经与基础科学研究深度融合, 成为推动科学研究和技术进步的重要工具。未来的科学研究将更多依赖于数据驱动的方法, 通过处理学习海量数据, 加速科学研究和新的科学发现的进程。随着 AI 技术不断迭代, 科学研究的门槛可能会降低, 使得更多的研究者能够参与到前沿科学的探索中, 推动科学知识的普及和应用。同时也对教育体系提出新的要求, 需要培养能够适应 AI 时代的新型人才, 以满足未来科学研究的需求。

## 参 考 文 献

- [1] Dill KA, MacCallum JL. The protein-folding problem, 50 years on. *Science*, 2012, 338(6110): 1042-1046.
- [2] Levinthal C. Are there pathways for protein folding? *Journal de Chimie Physique*, 1968, 65: 44-45.
- [3] Anfinsen CB. Principles that govern the folding of protein chains. *Science*, 1973, 181(4096): 223-230.
- [4] Riddle DS, Santiago JV, Bray-Hall ST, et al. Functional rapidly folding proteins from simplified amino acid sequences. *Nature Structural Biology*, 1997, 4(10): 805-809.
- [5] Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology*, 1998, 277(4): 985-994.

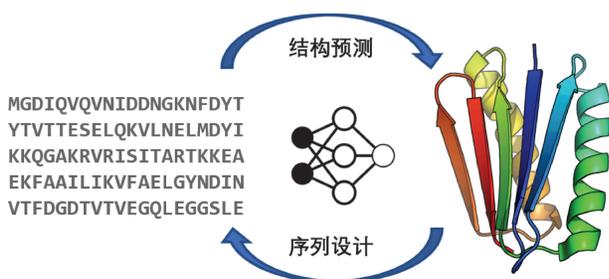


图 1 蛋白质结构预测与序列设计

- [6] Das R, Baker D. Macromolecular modeling with Rosetta. *Annual Review of Biochemistry*, 2008, 77: 363—382.
- [7] Regan L, DeGrado WF. Characterization of a helical protein designed from first principles. *Science*, 1988, 241(4868): 976—978.
- [8] Dahiyat BI, Mayo SL. De novo protein design: fully automated sequence selection. *Science*, 1997, 278(5335): 82—87.
- [9] Kuhlman B, Dantas G, Ireton GC, et al. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 2003, 302(5649): 1364—1368.
- [10] Jiang L, Althoff EA, Clemente FR, et al. De novo computational design of retro-aldol enzymes. *Science*, 2008, 319(5868): 1387—1391.
- [11] Röthlisberger D, Khersonsky O, Wollacott AM, et al. Kemp elimination catalysts by computational enzyme design. *Nature*, 2008, 453(7192): 190—195.
- [12] Fleishman SJ, Whitehead TA, Ekiert DC, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*, 2011, 332(6031): 816—821.
- [13] Koga N, Tatsumi-Koga R, Liu GH, et al. Principles for designing ideal protein structures. *Nature*, 2012, 491(7423): 222—227.
- [14] Huang PS, Oberdorfer G, Xu CF, et al. High thermodynamic stability of parametrically designed helical bundles. *Science*, 2014, 346(6208): 481—485.
- [15] King NP, Bale JB, Sheffler W, et al. Accurate design of co-assembling multi-component protein nanomaterials. *Nature*, 2014, 510(7503): 103—108.
- [16] Bhardwaj G, Mulligan VK, Bahl CD, et al. Accurate *de novo* design of hyperstable constrained peptides. *Nature*, 2016, 538(7625): 329—335.
- [17] Huang PS, Boyken SE, Baker D. The coming of age of *de novo* protein design. *Nature*, 2016, 537(7620): 320—327.
- [18] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge. *Nature*, 2017, 550(7676): 354—359.
- [19] Shaw DE, Maragakis P, Lindorff-Larsen K, et al. Atomic-level characterization of the structural dynamics of proteins. *Science*, 2010, 330(6002): 341—346.
- [20] Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 2020, 577(7792): 706—710.
- [21] Wang S, Sun SQ, Li Z, et al. Accurate *de novo* prediction of protein contact map by ultra-deep learning model. *PLoS Computational Biology*, 2017, 13(1): e1005324.
- [22] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021, 596(7873): 583—589.
- [23] Tunyasuvunakool K, Adler J, Wu Z, et al. Highly accurate protein structure prediction for the human proteome. *Nature*, 2021, 596(7873): 590—596.
- [24] Watson JL, Juergens D, Bennett NR, et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 2023, 620(7976): 1089—1100.

## Unlocking Nature's Secrets: The 2024 Nobel Prize in Chemistry for “Computational Protein Design and Structure Prediction”

Yuan Liu<sup>1\*</sup>    Chu Wang<sup>1, 2\*</sup>

1. College of Chemistry and Molecular Engineering, Peking University, Beijing 100871
2. Advanced Academy for Interdisciplinary Studies, Peking University, Beijing 100871

**Abstract** The 2024 Nobel Prize in Chemistry was awarded to professor David Baker from the University of Washington in the United States, in recognition of his outstanding contributions to the field of “computational protein design”; and to two scientists, Demis Hassabis and John M. Jumper, from the UK’s DeepMind company, in recognition of their groundbreaking achievements in the field of “protein structure prediction”.

**Keywords** Nobel Prize in Chemistry; protein folding; protein structure prediction; protein design; AlphaFold; Rosetta

(责任编辑 张强)

\* Corresponding Authors, Email: wendao@pku.edu.cn; chuwang@pku.edu.cn