·综 述·

蛋白质结构预测模型 AlphaFold2 的应用进展

张弘[1]，王慧洁[1]，鲁睿捷[1]，兰家靖[1]，陈林洁[1,2]，何小柏[1,3*]

1 杭州医学院检验医学院与生物工程学院，浙江 杭州 310053
2 检验诊断关键技术浙江省工程研究中心，浙江 杭州 310053
3 浙江省生物标志物与体外诊断转化重点实验室，浙江 杭州 310053

张弘, 王慧洁, 鲁睿捷, 兰家靖, 陈林洁, 何小柏. 蛋白质结构预测模型 AlphaFold2 的应用进展[J]. 生物工程学报, 2024, 40(5): 1406-1420.
ZHANG Hong, WANG Huijie, LU Ruijie, LAN Jiajing, CHEN Linjie, HE Xiaobai. Advances in the application of AlphaFold2: a protein structure prediction model[J]. Chinese Journal of Biotechnology, 2024, 40(5): 1406-1420.

摘 要：蛋白质结构预测是生命科学和医学的重要研究领域，也是人工智能在科学研究中的重要应用场景。AlphaFold2 是由 DeepMind 开发的一种基于深度学习的蛋白质结构预测系统，可以从氨基酸序列中高效地生成原子级精度的蛋白质空间结构。由于 AlphaFold2 优越的性能，自问世以来对蛋白质结构预测方面的研究提供了前所未有的助力，因此备受关注和研究。本文介绍了 AlphaFold2 的模型架构、亮点、局限性和应用进展，列举了几种其他类型的蛋白质结构预测模型，并讨论了其能力、优势及局限性并思考该蛋白质结构预测模型的未来发展方向。

关键词：AlphaFold2；蛋白质结构；蛋白质结构预测；深度学习

资助项目：浙江省医药卫生科技项目(2023KY650, 2021KY132, 2020KY107)；浙江省大学生创新创业训练计划(S202313023085)
This work was supported by the Zhejiang Provincial Medical and Health Science and Technology Program (2023KY650, 2021KY132, 2020KY107), and the Innovation and Entrepreneurship Training Program for College Students of Zhejiang Province (S202313023085).
*Corresponding author. E-mail: shining0206@163.com
Received: 2023-10-06; Accepted: 2024-01-23; Published online: 2024-01-26

Advances in the application of AlphaFold2: a protein structure prediction model

ZHANG Hong[1], WANG Huijie[1], LU Ruijie[1], LAN Jiajing[1], CHEN Linjie[1,2], HE Xiaobai[1,3*]

1 School of Laboratory Medicine and Bioengineering, Hangzhou Medical College, Hangzhou 310053, Zhejiang, China
2 Zhejiang Engineering Research Centre for Key Technology of Diagnostic Testing, Hangzhou 310053, Zhejiang, China
3 Key Laboratory of Biomarkers and In Vitro Diagnosis Translation of Zhejiang Province, Hangzhou 310053, Zhejiang, China

Abstract: Protein structure prediction is an important research field in life sciences and medicine, and it is also a key application scenario of artificial intelligence in scientific research. AlphaFold2 is a protein structure prediction system developed by DeepMind based on deep learning, capable of efficiently generating the atomic-scale spatial structure of a protein from the amino acid sequence. It has demonstrated superior performance in the prediction of protein structures since its inception, thus attracting much attention and research. This paper introduces the model architecture, highlights, limitations, and application progress of AlphaFold2. Furthermore, it briefs the capabilities, highlights, and limitations of several other types of protein structure prediction models and prospects the future development direction in this field.

Keywords: AlphaFold2; protein structure; protein structure prediction; deep learning

蛋白质作为生物的重要组成成分和生命过程的主要承担者，其空间结构与其功能密切相关，提供了宝贵的生物信息。因此准确地预测蛋白质结构对蛋白质设计、结构生物学、基于蛋白质结构的药物开发等生命科学和医学的研究和发展具有重大意义[1]。

蛋白质结构预测(protein structure prediction, PSP)是根据蛋白质的氨基酸序列推测其三维结构的方法[2]，目前 PSP 算法主要有 3 种[3]：同源建模、从头建模与基于机器学习的建模，它们的共同理论基础为 Anfinsen 法则[4]，即氨基酸序列决定蛋白质的空间结构，而空间结构又决定了蛋白质的生物功能。同源建模基于同源蛋白质结构相似的原理，以已知同源蛋白质结构为模板构建目的蛋白质结构，该法简单快速，但不能预测同源结构尚未确定的蛋白质结构。从头建模基于第一性原理和蛋白质的天然结构对应于其自由能的全局最小值这一事实[5]，该方法不依赖于数据库信息，通过能量函数模拟构象从而发现新结构，但能量函数难建立，且算力需求大，目前只适用于氨基酸残基数量在 10−80 之间的小蛋白。

作为机器学习算法中的一员，深度学习(deep learning, DL)算法近年来在大数据、算力、算法上不断优化和提升，同时也在蛋白质结构预测中发挥了巨大作用[3]。基于 DL 的蛋白质结构预测算法相比于上述两种，在各方面都融入了神经网络，是一种最新的数据驱动算法，具有更准确的预测结果，已逐步取代同源建模和从头建模[2]。目前用于蛋白质结构预测的 DL 建模方法主要有 AlphaFold2、RoseTTAFold、ESMFold 等。此外，我们还总结了 1967 年至今蛋白质结构预测方法的发展历程(图 1)。

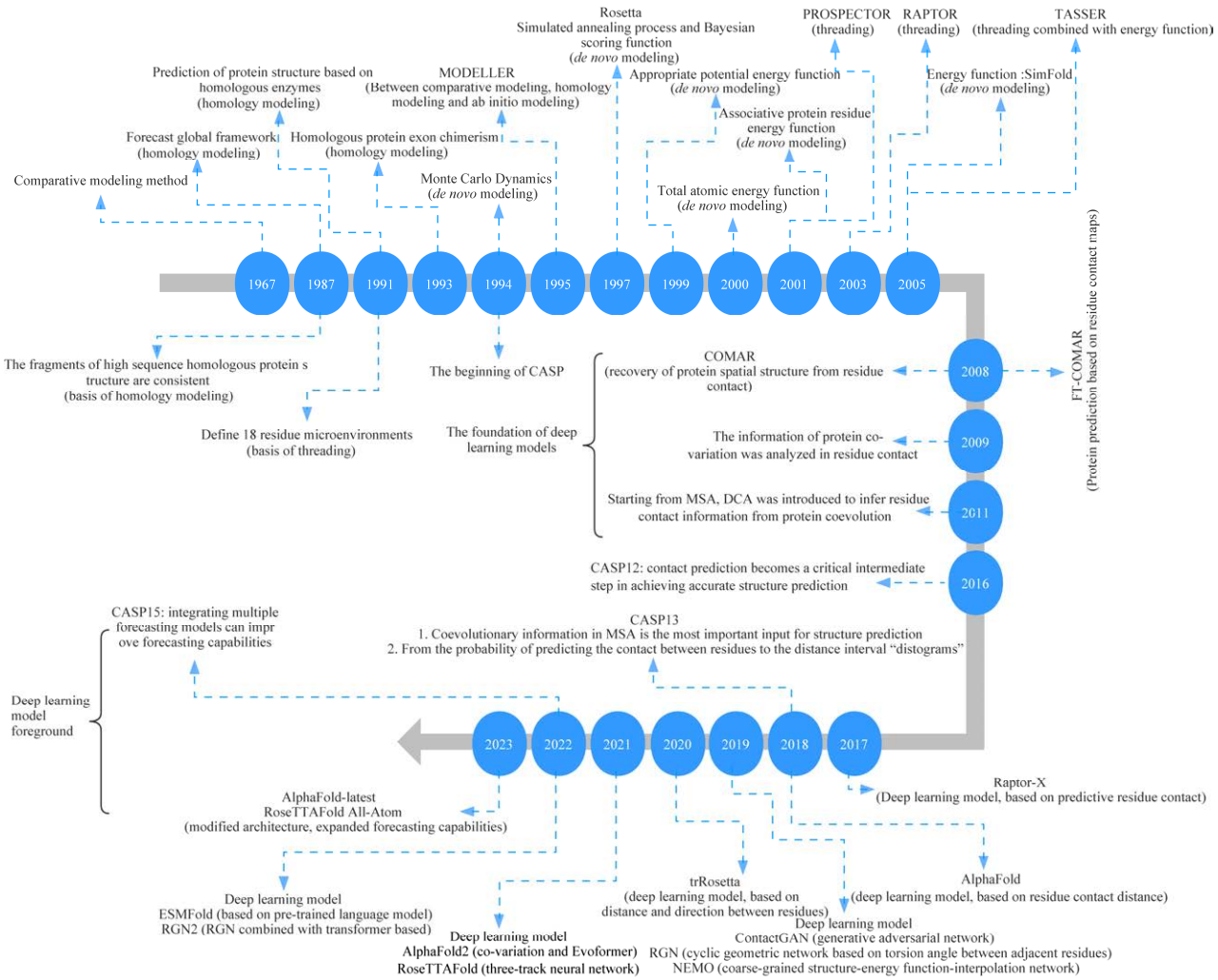

图 1　蛋白质结构预测方法的发展历程

Figure 1　The development of protein structure prediction methods.

其中，AlphaFold2 (AF2)是由 DeepMind 开发的一项基于人工智能深度学习的蛋白质结构预测方法，其利用深度学习算法，根据给定的氨基酸序列，使用神经网络架构模型 Evoformer，通过利用同源蛋白质的信息和多序列比对(multiple sequence alignment, MSA)，以原子级精度从氨基酸序列中训练预测蛋白质结构[3,6]。因具有优秀的预测精度，在第 14 届结构预测的关键评估(critical assessment of structure prediction, CASP14)大赛中获得了冠军(排名对比可见此链接 https://predictioncenter.org/casp14/zscores_final.cgi)[7]；而在 2022 年举行的 CASP15 的预测评估中[8-9]，由于参赛者采用的各类预测方法普遍整合了 AF2，各模型结果的整体折叠和界面接触预测方面表现出色，相比于 CASP14 期间 31%的成功率，实现了令人印象深刻的 90%的成功率。本文对 AF2 的模型架构、亮点、应用等方面进行综述，并讨论了目前 AF2 存在的局限性及其未来发展方向。

1 AlphaFold2 概述

1.1 AF2 模型架构

AF2 是 DeepMind 最先进的蛋白质结构预测方法，其利用一种新颖的神经网络架构模型——Evoformer，结合蛋白质结构的进化机理、物理和几何约束规则，实现了优越的蛋白结构预测性能。

Evoformer 的开发灵感来源于 MSA-Transformer (图 2)[10-11]。Transformer 是一种新兴的自注意力模型，利用自注意力机制来提取序列数据的内在特征，具有广泛的人工智能应用潜力[12]。在 Transformer 的基础上延伸出的 MSA-Transformer[11]使用 MSA 表示(MSA representation)作为输入，通过注意力(attention)机制来处理蛋白质序列的信息。Evoformer 使用了 2 组类似 MSA-Transformer 的结构，分别用于捕捉氨基酸残基间的多序列比对信息和结构约束信息特征，从而提高了预测质量[6]。

AF2 的生物学原理[13]是利用蛋白质共进化中包含的结构信息。在多序列比对中，可以观察到一些残基在不同的蛋白质中有协同变异[14]的行为，这可能与空间邻近或功能联系有关。这种共进化可以发生在同一蛋白质内部，也可以发生在不同蛋白质之间，尤其是在相互作用的蛋白质界面上。AF2 通过多序列比对，从与目的蛋白质具有共进化关系的蛋白质序列中获取保守性和协变性(co-variation)[14]信息，并结合氨基酸残基

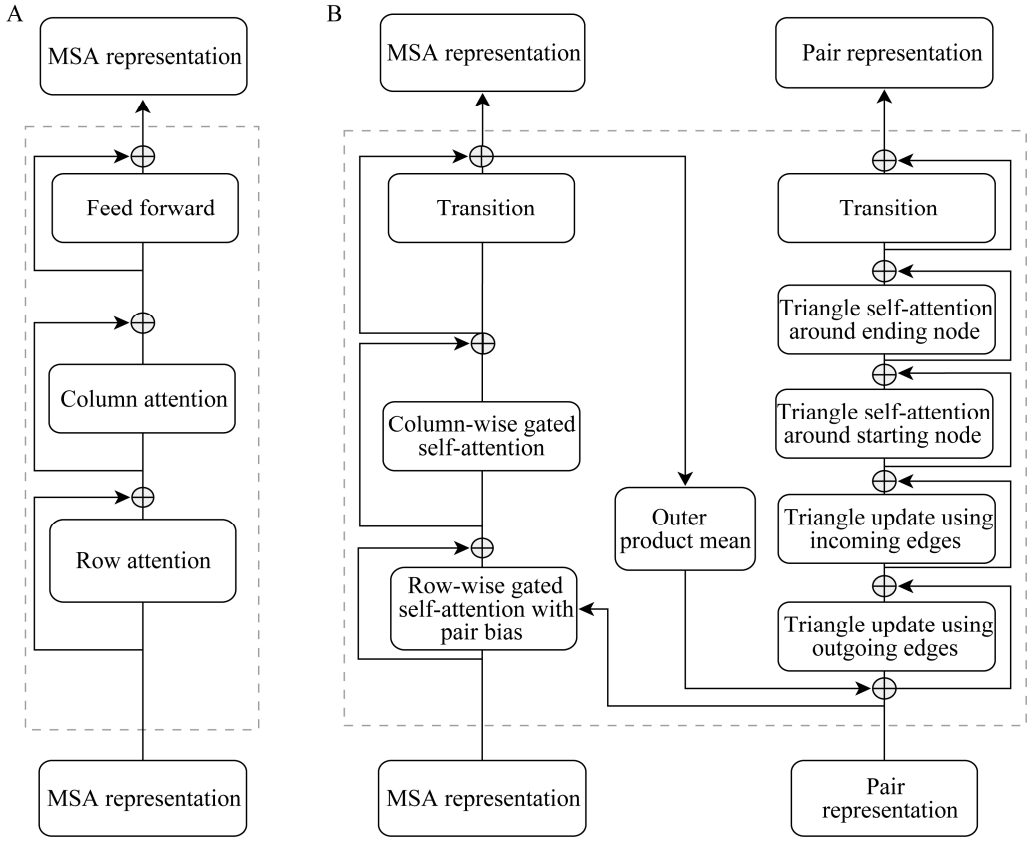

图 2 MSA-Transformer 与 Evoformer 的架构对比图[10]

Figure 2 The architecture comparison of MSA-Transformer and Evoformer[10]. A: The architecture of MSA-Transformer. B: The architecture of Evoformer.

间结构约束信息特征，通过神经网络模型预测目的蛋白质结构。基于此原理进行的蛋白质预测不仅提高了预测的准确性，而且提高了预测的效率[3]。

AF2 还利用中间损失(intermediate loss)函数，包括帧点对齐误差损失(frame-aligned point error, FAPE)、辅助损失(auxiliary loss)、违规损失(violation loss)等来优化模型预测性能[6]。其中，FAPE 是 DeepMind 独立构建的损失函数，用于测量蛋白质结构中预测原子坐标与真实原子坐标之间的误差，确保了侧链相互作用的准确性，同时也为 AF2 提供了手性的主要来源。辅助损失用于将额外的信息引入模型中，可以帮助模型从数据中学习到更多的特征和关系。违规损失用于在微调过程中惩罚违反肽键几何约束的情况，使得预测的结构满足立体化学要求。此外，AF2 还采用了循环机制、自蒸馏(self-distillation)[15]和自估计准确度(self-estimate of accuracy)[6]等方法，进一步提升了模型的预测性能。

AF2 总体框架[3,6]见图 3，分为特征提取模块(feature extraction module)、编码模块(encode module)、结构解码模块(structure decode module)。输入模块根据给定的氨基酸序列，在序列数据库中寻找其同源序列，并进行多序列比对。MSA 可以反映出蛋白质序列之间的相似性和共进化信息，这些信息对于预测蛋白质结构尤为重要。同时输入模块检查是否有任何同源序列存在已知的三维结构，并在蛋白质结构数据库中查找；如有，输入模块会构建一个两两距离矩阵，表示每一对氨基酸之间的空间距离。随后输入模块生成 MSA 表示和成对表示(pair representation)，其中 MSA 表示是一个三维矩阵，表示每个氨基酸在 MSA 中的位置、频率和共进化信息；成对表示也是三维矩阵，表示氨基酸之间结构约束信息特征，包括每一对氨基酸之间的距离、角度和相互作用[6]。

生成的 MSA 表示和成对表示输入到 AF2 的核心模块——由 Evoformer 组成的编码模块。Evoformer 模块能够利用 MSA 和残基对之间的共进化信息来推理蛋白质的空间和进化关系[3,6]。AF2 使用 48 个不共享权重的 Evoformer 模块，每个模块有一个 MSA 表示和一个成对表示作为

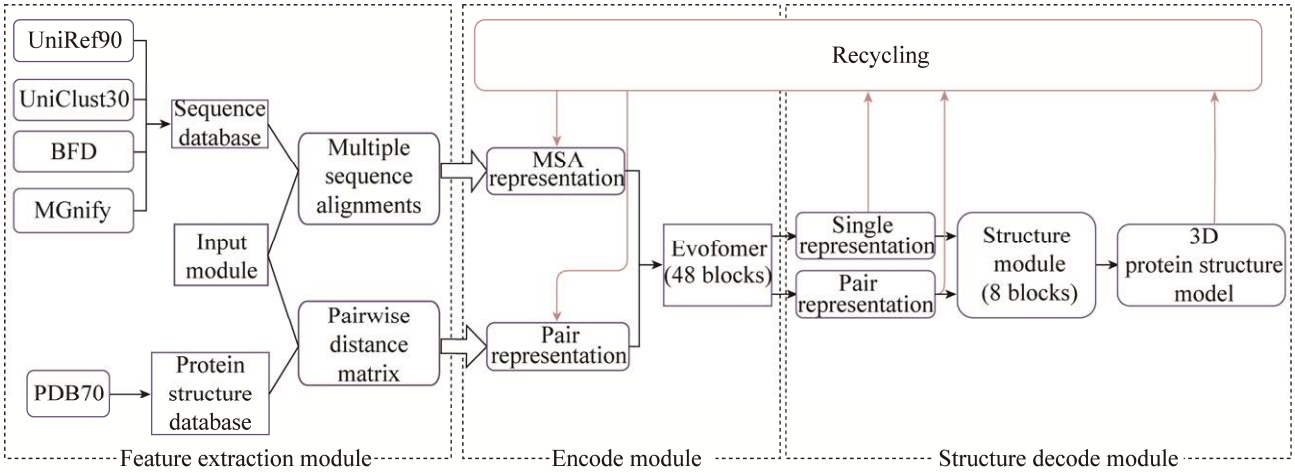

图 3 AlphaFold2 总体框架[6]

Figure 3 The overall framework of AlphaFold2[6].

输入，并输出更新后的 MSA 表示和成对表示。每个 Evoformer 模块通过三角自注意力机制和几何变换来更新这两个表示，在计算过程中还使用了 Dropout 方法来减轻过拟合问题[6]。

最后是结构解码模块，可以将编码模块的输出结果转换为目标蛋白质的三维结构。结构解码模块的关键组件是不变点注意力(invariant point attention, IPA)，它是一种几何感知的注意力机制，用于更新单一表示。IPA 操作的最终注意力值在三维空间中是等变的，这意味其不受全局刚体运动(包括旋转和平移)的影响，即不管蛋白质结构在三维空间中如何变换，不变点注意力都能保持相同的输出，这有助于提高蛋白质结构预测的准确性和稳定性[6]。

1.2 AF2 的亮点

AF2 利用了自注意力机制、自蒸馏的训练方式，结合高效的搜索算法，从大量的蛋白质序列和结构数据中深度学习蛋白质的特征和规律，高效地预测出未知的蛋白质结构，达到了接近实验水平的精度。

亮点一：AF2 采用的神经网络架构。AF2 使用了 Evoformer，从不同方面学习蛋白质序列和结构的特征。Evoformer 的两组类 MSA-Transformer 模块分别作用于原始的多序列比对(MSA)和成对信息。Evoformer 将门控[16]引入到自注意力机制，实现根据输入的信息动态地调节网络的输出，从而提高网络的灵活性和适应性。MSA 逐行门控自注意力机制使模型能够捕捉氨基酸序列和蛋白质结构中的长程依赖关系，MSA 逐列门控自注意力机制是一种“保守性感知”的注意力机制，能够让不同物种之间的元素交换信息。Evoformer 的三角自注意力机制使模型能够学习蛋白质分子内部的几何约束。结构解码模块部分还使用了一个类 MSA-Transformer 模块，将残基几何编码为三维空间中的有向参考框架[6]。简言之，基于注意力机制的神经网络架构能够有效地捕捉蛋白质的关键特征，从而使 AF2 具备极强的预测能力。

亮点二：AF2 采用的数据与搜索算法。数据来源于 Uniref90、Uniclust30、MGnify 和 Big Fantastic Database 等高质量的蛋白质序列数据库[17-19]，用于训练和模板的结构数据库，则是分别采用了布鲁克海文蛋白质数据库(BrookHaven Protein Data Bank, PDB)和 PDB70[20]。大量的氨基酸序列和结构数据使深度学习神经网络能够探索蛋白质序列和结构之间的各种依赖关系[3]，有助于提高 AF2 预测结果的精确性。AF2 还利用了一些高效的搜索算法，包括用于基因搜索的 JackHMMER[21]和 HHBlits[22]，以及用于模板搜索的 HHSearch[20]。

亮点三：AF2 的训练方法。开发者利用了自蒸馏的思想，即采用有监督学习进行知识蒸馏[15]。将 PDB 和一个新的自蒸馏无标签数据集(由预测的蛋白质结构组成)作为训练数据来训练 AF2，其中 25%的训练样本来自 PDB 中已知结构，75%的数据来自新的自蒸馏数据集，目的是让 AF2 通过使用不同的训练数据增强方法来重现之前难以预测的蛋白质结构。这种集成数据及数据集的训练方法可以让 AF2 学习到更多的蛋白质特征，从而预测出更复杂和多样的蛋白质结构[6]。

AF2 还使用端到端框架减少信息的损失和噪声，并通过循环机制让模型多次更新和优化其输出，从而达到更好的收敛和稳定性[3]。这些亮点共同提高了 AF2 预测蛋白质结构的精确性，为生物学和医学的发展提供了强大的工具。

1.3 AF2 的局限性

AF2 对蛋白质的精细结构的预测能力不足[23]，比如 He 等[23]利用 AF2 对 G 蛋白偶联受体的结构进行预测，但在细胞外和跨膜结构域的组装、

配体结合口袋的形状以及转导结合界面的构象与实验结构在许多方面存在差异；在我们使用过程中发现严重急性呼吸综合征冠状病毒 2 (severe acute respiratory syndrome coronavirus-2, SARS-CoV-2)的受体结合区(receptor binding domain, RBD)结构域无法“弹出”。我们推测可能是由于 AF2 是基于水溶液中蛋白质结构开发的，无法模拟蛋白质所处环境中溶剂条件、pH 值、离子强度等外力对结构的影响[24]。

对于预测结构和各结构域的相对位置，无论置信度高低，AF2 预测结果都存在不确定性，其产生的原因[25]包括：(1) 被视为无序片段的低置信度区域，是由于 AF2 训练使用的 X 射线数据中存在无法解析的蛋白质紊乱部分。蛋白质紊乱导致了 AF2 预测产生低置信度的无序片段；(2) 部分高置信度的结构域之间的连接具有灵活性，使结构域的相对位置存在误差。这种不确定性可能导致 AF2 在结构相似性、口袋、变异效应或模型构建等应用中产生错误的结果或识别。这表明蛋白质结构的实验研究不会被 AF2 替代，AF2 预测的结构需要结合实验数据进行人工检查和纠正，故实验数据收集与人工智能的结合可能是一个未来的趋势。

2 AlphaFold2 的应用

AF2 不仅可以预测单链蛋白质的三维结构，还具有预测多蛋白质复合物结构的潜能，为蛋白质功能研究、蛋白质设计和药物发现提供了有力的工具。本节将综述近期关于 AF2 在不同领域的应用现状，展示其在解决一些重要的生物学和医学问题方面的潜力。

AF2 可以用于解析一些难以通过传统实验方法获得结构信息的蛋白质或复合物[26]，也能与 X 射线晶体学、冷冻电镜和核磁共振等实验方法相结合，提高结构解析的效率和准确度。例如，Xiao 等[27]利用 AF2 预测 69 个大肠杆菌主要协同转运蛋白超家族(major facilitator superfamily, MFS)蛋白在 N 端和 C 端结构域界面残基选择性突变后的不同构象结构，为研究 MFS 转运蛋白的转运机制提供了结构基础；Hu 等[28]利用 AF2 预测并通过 X 射线晶体学测定了轮状病毒棘突蛋白聚糖结合结构域的新折叠，说明了进化如何在同一病毒属内的同源蛋白质中加入具有相似功能的结构不同的褶皱；Yang 等[29]利用 AF2 预测了 SARS-CoV-2 变体 Omicron 的 S、M 和 N 蛋白的结构，并详细研究了突变如何影响 S 蛋白及其 S1 亚基的 N 末端结构域和 RBD 部分。Wayment-Steele 等[30]利用基于 AF2 开发的 AF-Cluster，预测变性蛋白 KaiB 的多种构象状态，还对 628 个蛋白家族进行筛选，发现了一个结核分枝杆菌的分泌氧化还原酶 Mpt53 的一个可能的新构象。这些例子表明，AF2 可以为一些难以解析或缺乏实验数据的蛋白质或复合物提供可靠的结构信息，为理解其功能和机制提供重要的线索。

AF2 可以用于评估和优化已有的实验结构数据，提高其准确性和可信度。例如，Fowler 和 Williamson[31]评估了核磁共振结构和 AF2 预测结构的准确性，发现 AF2 往往比核磁共振集合更准确，但在一些动态区域，核磁共振结构更好，建议用 AF2 辅助核磁共振研究蛋白质结构；Xiao 等[27]预测结果相较于使用 X 射线晶体学获得的结构具有较小的均方根偏差(root mean square deviation, RMSD)值，表明 AF2 预测结果的高精度。这些例子表明，AF2 相较于传统的生物结构学研究方法具有独到的优势，可以为已有的实验结构数据提供一个有效的补充和验证手段，帮助发现其中可能存在的误差或不足，并提高其在后续应用中的可靠性。

AF2 可以批量分析大规模的相关蛋白质，为

寻找蛋白质相互作用的关键空间结构提供高效的工具。例如，Ibrahim 等[32]利用基于 AF2 开发的 ColabFold 预测分析了自噬相关基因 8/微管相关蛋白 1 轻链 3 (autophagy-related gene 8/microtubule-associated protein 1 light chain 3, ATG8/LC3)蛋白家族的空间结构，精确预测发现典型及非典型的 ATG8/LC3 作用基序/区域形成的口袋，进一步分析了这些结构在细胞自噬通路中的功能和作用；Lorenz 等[33]利用 AF2 预测了一些 Krüppel 相关性保守盒(Krüppel-associated-boxes, KRAB)的结构域，发现它们有 2 个由 α 螺旋构成的 L 形结构，随氨基酸变化与现代 KRAB 域 B 亚型的第 3 个螺旋形成一种典型的特征空间排列，为 KRAB 如何与含三方基序 28 形成复合物提供了基本结构认识。这些例子表明，AF2 可以对大量的相关蛋白质进行快速的结构预测和分析，为揭示蛋白质相互作用的关键空间结构和功能提供了强有力的支持，为生物学研究和药物开发提供了新的思路和方法。

AF2 可以用于从头设计新型或改良型的蛋白质或复合物，为蛋白质工程和药物开发提供创新的方案。例如，Jendrusch 等[34]开发的一种基于 AF2 的从头蛋白质设计框架，称为 AlphaDesign。它通过反转 AF2 网络，根据目标的蛋白质结构从随机序列开始预测完全新颖的蛋白质单体和复合物，生成能够折叠成该结构的蛋白质序列并使用了预测权重集和损失函数来提高生成序列的质量和多样性。这表明 AF2 已经学习了足够多的蛋白质折叠原理，可以用于从头蛋白质设计，并为解决从头蛋白质设计中仍然存在的挑战提供了潜在的解决方案；Goverde 等[35]设计了一种基于 AF2 的蛋白设计流程(图 4)，通过反转 AF2 网络，以蛋白质结构生成目标折叠的序列，并进行表面优化和体外验证，该方法设计的蛋白质能够在溶液中折叠成预期的结构，并且具有高熔点温度，但没有充分考虑蛋白质表面的疏水性和亲水性分布，需额外干预改善表面特性和功能性能；Zeng 等[36]利用 AF2 设计了血凝素干细胞

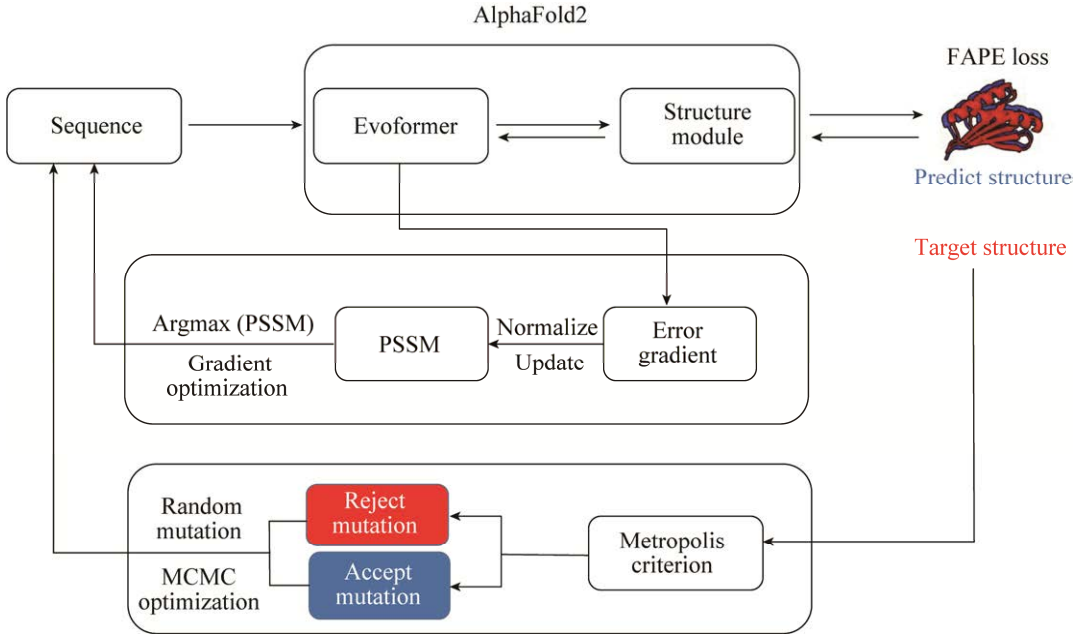

图 4　基于 AF2 的蛋白设计流程[35]

Figure 4　Protein design process based on AF2[35]. FAPE: Frame-aligned point error; PSSM: Position specific scoring matrix; MCMC: Markov chain Monte Carlo.

疫苗 B60-Stem-8070，与原血凝素干细胞抗原相比，该疫苗表现出更好的性能。这些例子表明，AF2 可以为蛋白质设计提供强大的支持，帮助创造出具有特定功能或特性的新型蛋白质。然而，AF2 也不是万能的，它仍然需要与其他计算方法或实验技术相结合，以克服一些固有的局限性或不确定性。

AF2 也能够预测蛋白质与其他分子之间的相互作用，辅助药物靶点的开发与研究。例如，Weng 等[37]利用 AF2 预测了一种新的潜在抗癌靶点 WD40 重复和 SOCS 盒结合蛋白 1 (WD40 repeat and SOCS box containing 1, WSB1)的三维结构并通过分子动力学模拟对预测的结构进行优化，用优化后的 WSB1 三维结构作为受体结构进行分子对接，筛选 WSB1 抑制剂，最后得到了一些潜在的活性化合物；Ren 等[38]将 AF2 嵌入到他们的端到端人工智能药物发现引擎中，在 30 d 内从靶点选择到合成 7 个化合物后，发现了当时第一个针对细胞周期依赖性激酶 20 的命中小分子化合物。这些例子表明，AF2 在药物开发领域具有巨大的应用潜能，有助于实现药物靶点开发与小分子药物大规模筛选。

AF2 还可以用于预测和分析蛋白质之间的相互作用网络，为蛋白质功能和调控机制的研究提供重要的信息。例如，本课题组 Fang 等[39]利用 ColabFold——一个搭载了 AF2 的免费蛋白质结构预测平台[40]，来预测 SARS-CoV-2 棘突蛋白不同变体与潜在受体相互作用形成的复合体。他们从 ColabFold 预测结果中提取预测模板建模得分(predicted template modeling-score, pTM)、接口 pTM (interface pTM, ipTM)、Confidence 等相关特征，并利用这些特征可以分析出 SARS-CoV-2 棘突蛋白不同变体与不同受体结合作用的强度，如野生型 SARS-CoV-2 与跨膜丝氨酸蛋白酶 2 (transmembrane protease serine 2, TMPRSS2)预测结果的 ipTM 和 Confidence 明显高于 Omicron 型 SARS-CoV-2 的预测结果，这表明 Omicron 型与 TMPRSS2 结合作用的强度低于野生型，而这也与实验结果相符；Bartas 等[41]通过搜索含有 Zα 域(Z-DNA/Z-RNA 结合蛋白域)的蛋白质，从 AF2 预测结构数据库中识别出 185 种可能与 Z-DNA/Z-RNA 结合并在多种细胞过程中发挥重要作用的蛋白质；Yin 等[42]使用 AF2 的多聚体版本 AlphaFold-Multimer[43]，预测了一个包含 4 000 多个最新的蛋白质复合物的大型数据集的结构，评估了所有非冗余的低模板界面的相似性，对于异源性界面，成功预测了 67% 的结构，并高精度预测 23%的结构，而同源性界面，则成功预测了 69%的结构，并高精度预测 34%的结构，这些结果都表明与现有的常规方法(基于相互作用位点的方法、基于分子对接的方法)相比，以 AlphaFold-Multimer 为代表的基于端到端的方法，具有更优越的性能。Gao 等[44]开发了一种基于 AF2 但不依赖于成对 MSA 的系统，称为 AF2Complex。该模型可以预测多聚体蛋白质的直接物理相互作用，并以细胞色素 c 生物发生系统 I 为例，验证了其优异的性能。这些例子表明，AF2 可以为蛋白质相互作用网络提供有价值的信息，帮助揭示蛋白质之间如何协同工作。

此外 AF2 还可以用于构建大规模的蛋白质结构数据库，为生物学和医学研究提供丰富和可靠的资源。例如，Wang 等[45]利用 AF2 预测的人类蛋白质组中所有蛋白质的三维结构，使用 CAVITY 程序从中检测出潜在的配体结合位点，并构建了 CavitySpace 数据库。CavitySpace 数据库可以用于通过反向对接的方法进行蛋白质功能预测和药物再利用研究，发现了一些潜在的新靶标和新用途；Ma 等[46]研究发现：基于真实蛋白结构的蛋白质功能预测模型可以使用由 AF2

预测的结构组成的虚拟训练数据进行训练，甚至仅使用 AF2 预测的虚拟训练数据训练的模型与基于真实蛋白质结构的模型的性能相当。Varadi 等[47]建立了一个开放的 AlphaFold 蛋白质结构数据库-AlphaFold DB，该数据库收录了大量高精度的蛋白质结构预测模型，为生物学研究提供了宝贵的资源；Hekkelman 等[48]利用实验测定的蛋白质结构中的小分子、离子，对 AlphaFold 蛋白质结构数据库中的蛋白质模型进行了“移植”补充，建立了 AlphaFill 数据库。该数据库收录了 995 411 个 AlphaFold 模型的 12 029 789 次“移植”结果，并提供了相关的验证指标和可视化界面，丰富了 AlphaFold 数据库中的模型信息，为研究人员提供了新的蛋白质功能假设的线索。这些例子表明，AF2 可以为构建大规模的蛋白质结构数据库提供高质量和高效率的生成和分析方法，为生物学和医学研究提供宝贵的资源。

AF2 蛋白质结构数据库的出现还可以将突变信息与结构数据集融合，释放这些预测数据集的潜力。例如，Blaabjerg 等[49]创新了一种利用深度学习表征快速准确预测蛋白质稳定性变化的模型，称为 RaSP，RaSP 可以对每个残基在不到 1 s 的时间内实现饱和诱变稳定性预测。RaSP 模型不仅可以接受实验测定的蛋白质结构作为输入，还可以接受 AF2 预测的结构作为输入，且预测性能相当，甚至稍有提高。因此，AF2 为 RaSP 模型提供了强大的支持，使其能够实现大规模的蛋白质稳定性预测；Sun 等[50]提出来一个专门用于进行零样本自由能变化(ΔΔG)的预测的自监督图神经网络，称为 Pythia。Pythia 利用 AlphaFold 数据库中的 2 600 万个预测结构，计算了所有可能的单点突变，展示了 Pythia 在大规模突变预测方面的高效性和潜力。Cheng 等[51]基于 AF2 改进了一个模型，称为 AlphaMissense。该模型在 AlphaFold 的基础上进行了微调，利用人类和灵长类的变异频率数据作为弱标签，避免了使用人工注释的循环。AlphaMissense 可以对所有可能的单氨基酸突变进行替换，将 89%的错义变异区分为可能致病或可能良性。这些例子表明，AF2 生成的高质量的蛋白质结构在预测蛋白质稳定性变化方面的应用价值，有助于深入剖析蛋白质的复杂性。

AF2 是一种基于深度学习的蛋白质结构预测方法，它在不同领域的应用中展示了其巨大的应用前景。AF2 不仅可以提供可靠的蛋白质或复合物的三维结构信息，还可以用于评估和优化已有的实验结构数据，在寻找蛋白质功能关键空间结构、从头设计新型或改良型的蛋白质或复合物、预测和分析蛋白质之间或蛋白质与其他分子之间的相互作用网络、构建和分析大规模的蛋白质结构数据库等多方面提供高效的研究工具。AF2 为蛋白质科学领域带来了新的机遇和挑战，也为人工智能在科学研究中发挥更大作用提供了新的范例。

3 其他蛋白质结构预测的模型

目前蛋白质结构预测模型是结构生物学与深度学习交叉领域的研究热点，除了 AlphaFold2 外，有如 RoseTTAFold[52]、ESMFold[53]、Umol (预测蛋白质-小分子复合物)[54]等数量众多蛋白质结构预测模型。这些模型基于不同的原理和架构，具有独特的优势和缺陷，本节将简述几种蛋白质结构预测的模型。

3.1 RoseTTAFold

RoseTTAFold[52]是受 DeepMind 启发基于深度学习的蛋白质结构预测方法，能够根据蛋白质的氨基酸序列推断其三维空间结构。它使用了三轨神经网络，将残基间的距离和方向、序列和原子坐标联系起来，从而提高了预测的准确性和效率。它还可以利用多序列比对和共进化信息，增

强了对蛋白质结构的理解和建模能力，能够快速预测蛋白质-蛋白质复合物的结构。

但是，RoseTTAFold 仍然存在一些问题[55]：(1) 依赖于已知的蛋白质结构数据库和多序列比对，可能无法有效地处理新颖或稀有的蛋白质序列；(2) 对于低质量或过长的序列，可能无法给出可靠的结果；(3) 计算时资源要求仍然较高，且仍然需要后续验证和筛选最优解。

3.2 ESMFold

ESMFold[53,56]是一种基于预训练语言模型的蛋白质结构预测方法，它可以从单个蛋白质序列直接生成原子级的三维空间结构，而不需要依赖多序列比对或外部建模程序。它利用大规模预训练的蛋白质语言模型 ESM-2 (训练参数达到 150 亿，是目前最大的蛋白质语言模型)[56]来替代 MSA。ESMFold 的结构预测效果随着语言模型的规模和对序列的理解程度而提高(与困惑度呈负相关)。ESMFold 的预测速度比基于 MSA 的方法快一个数量级，可以高效地探索大规模的蛋白质结构空间。

ESMFold 仍存在一些问题[56]：(1) ESMFold 的预测准确性与对序列的困惑度呈负相关，即当语言模型无法理解一条序列时将难以推断出它的结构；(2) 目前对于多链或复合物等更复杂的结构预测能力同 AlphaFold2 相比尚有差距，ESMFold 还需要进一步地改进和优化。

3.3 Umol

Umol[54]是一种基于神经网络的蛋白质-配体结构预测方法，它可以从蛋白质序列信息、结合位点在序列中的位置和配体的化学图直接生成原子级的三维空间结构，而不需要任何结构信息。它也利用 Evoformer 来处理蛋白质和配体的特征，并在结构模块中预测蛋白质-配体复合物的构象。Umol 没有对蛋白质和配体的柔性做任何限制，因此可以预测出完全动态的结构。作者在 PoseBusters 基准集上评估了 Umol，Umol 和 RoseTTAFold All-Atom (RoseTTAFold 的全原子版本)[57]是为数不多的不需要已知蛋白质结构作为输入的方法，Umol (45.3%)在成功率(配体 RMSD≤2Å) 上优于 RoseTTAFold All-Atom (42%)，且优于部分需要已知的蛋白质结构作为输入的蛋白质-配体对接方法。

但是 Umol 仍存在一些问题[54]：(1) Umol 的预测准确性与配体的大小和复杂度有关，对于较大的配体，预测的方向可能不正确；(2) 目前 Umol 在 PoseBusters 基准集的成功率仍然低于需要已知的蛋白质结构作为输入的蛋白质-配体对接方法如 AutoDock Vina；(3) 使用的训练数据集 PDBbind 是从 2019 年的 PDB 中提取的，而在此之后，PDB 中又提交了许多新的蛋白质-配体复合物，这表明更高的准确性是可能实现的。

4 总结与展望

AlphaFold2 作为基于深度学习的蛋白质结构预测模型，通过独特的原理和架构，实现了高准确度的快速蛋白质结果预测，并在生物学和医学的研究中发挥多方面的作用。但是 AF2 还存在严苛的算力需求的缺陷，难以开展大规模应用，并且在某些蛋白的结构预测应用上存在着结构误差，有待相关技术工作者对其进行改良以满足更广泛的应用场景。

目前基于深度学习开发的模型在生物学领域有着广阔的应用前景，不仅用于预测蛋白质结构，还为生物学各方面研究提供了工具，例如预测化合物-蛋白质相互作用的 DeepDTA、DeepCPI、WideDTA[58]。AlphaFold2 具有极强的可塑性、无穷的潜能及广阔的应用前景，目前基于 AlphaFold2 已开发出多种预测模型以期实现更多的功能，如上文提到的 AlphaFold-Multimer[32,42-43]、AF2Complex[44]、ColabFold[40]和 AlphaMissense[51]，

以及 DeepMind 联合 Isomorphic Labs 团队开发出可以实现预测蛋白-核酸、蛋白-配体、抗体-抗原相互作用的最新一代 AlphaFold2：AlphaFold-latest (https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/a-glimpse-of-the-next-generation-of-alphafold/alphafold_latest_oct2023.pdf)。

在 CASP15 单体结构预测前三名 Yang-server、UM-TBM 和 PEZYFolding 都通过将包括 AlphaFold2 在内的多个预测算法整合到自己的预测算法中实现算法优化[9]，这表明整合多模型可能是提升预测准确率的方向。所以，未来 AlphaFold2 的开发方向可以聚焦在优化和改进的架构，尝试多种模型联合运用，以实现预测能力的提升和预测功能的多元化，为更加复杂的问题提供高效的研究手段。

REFERENCES

[1] TUNYASUVUNAKOOL K, ADLER J, WU Z, GREEN T, ZIELINSKI M, ŽÍDEK A, BRIDGLAND A, COWIE A, MEYER C, LAYDON A, VELANKAR S, KLEYWEGT GJ, BATEMAN A, EVANS R, PRITZEL A, FIGURNOV M, RONNEBERGER O, BATES R, KOHL SAA, POTAPENKO A, et al. Highly accurate protein structure prediction for the human proteome[J]. Nature, 2021, 596(7873): 590-596.

[2] ALQURAISHI M. Machine learning in protein structure prediction[J]. Current Opinion in Chemical Biology, 2021, 65: 1-8.

[3] YANG ZY, ZENG XX, ZHAO Y, CHEN RS. AlphaFold2 and its applications in the fields of biology and medicine[J]. Signal Transduction and Targeted Therapy, 2023, 8(1): 115.

[4] HIRATA F, SUGITA M, YOSHIDA M, AKASAKA K. Perspective: structural fluctuation of protein and Anfinsen's thermodynamic hypothesis[J]. The Journal of Chemical Physics, 2018, 148(2): 020901.

[5] DORN M, E SILVA MB, BURIOL LS, LAMB LC. Three-dimensional protein structure prediction: methods and computational strategies[J]. Computational Biology and Chemistry, 2014, 53: 251-276.

[6] JUMPER J, EVANS R, PRITZEL A, GREEN T, FIGURNOV M, RONNEBERGER O, TUNYASUVUNAKOOL K, BATES R, ŽÍDEK A, POTAPENKO A, BRIDGLAND A, MEYER C, KOHL SAA, BALLARD AJ, COWIE A, ROMERA-PAREDES B, NIKOLOV S, JAIN R, ADLER J, BACK T, PETERSEN S, et al. Highly accurate protein structure prediction with AlphaFold[J]. Nature, 2021, 596(7873): 583-589.

[7] KRYSHTAFOVYCH A, SCHWEDE T, TOPF M, FIDELIS K, MOULT J. Critical assessment of methods of protein structure prediction (CASP)-Round XIV[J]. Proteins, 2021, 89(12): 1607-1617.

[8] OZDEN B, KRYSHTAFOVYCH A, KARACA E. The impact of AI-based modeling on the accuracy of protein assembly prediction: insights from CASP15[J]. Proteins: Structure, Function, Bioinformatics, 2023, 2023: 548341.

[9] 曹卫, 潘宪明. 蛋白质结构预测进展[J]. 生物化学与生物物理进展, 2023, 50(5): 1190-1194.
CAO W, PAN XM. Advances in protein structure prediction[J]. Progress in Biochemistry and Biophysics, 2023, 50(5): 1190-1194 (in Chinese).

[10] HU M, YUAN FJ, YANG KK, JU FS, SU JY, WANG HY, YANG F, DING QY. Exploring evolution-aware &-free protein language models as protein function predictors[J]. Advances in Neural Information Processing Systems, 2022, 35: 38873-38884.

[11] RAO RM, LIU J, VERKUIL R, MEIER J, CANNY J, ABBEEL P, SERCU T, RIVES A. MSA transformer[C]//Proceedings of the 38th International Conference on Machine Learning. Vienna, Austria: PMLR, 2021, 139: 8844-8856.

[12] VASWANI A, SHAZEER NM, PARMAR N, USZKOREIT J, JONES L, GOMEZ AN, KAISER L, POLOSUKHIN I. Attention is all you need[C]//Neural Information Processing Systems. Long Beach, California, USA: Curran Associates Inc, 2017: 6000-6010.

[13] PAZOS F, VALENCIA A. Protein co-evolution, co-adaptation and interactions[J]. The EMBO journal, 2008, 27(20): 2648-2655.

[14] ASHENBERG O, LAUB MT. Using analyses of amino acid coevolution to understand protein structure and function[J]. Methods in Enzymology, 2013, 523: 191-212.

[15] XIE QZ, LUONG MT, HOVY E, LE QV. Self-training with noisy student improves ImageNet classification[C]//

2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 10684-10695.

[16] MAKKUVA A, OH S, KANNAN S, VISWANATH P. Learning in gated neural networks[C]//International Conference on Artificial Intelligence and Statistics. Vienna, Austria: PMLR, 2020: 3338-3348.

[17] SUZEK BE, WANG YQ, HUANG HZ, MCGARVEY PB, WU CH, UNIPROT C. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches[J]. Bioinformatics, 2015, 31(6): 926-932.

[18] MIRDITA M, von den DRIESCH L, GALIEZ C, MARTIN MJ, SÖDING J, STEINEGGER M. Uniclust databases of clustered and deeply annotated protein sequences and alignments[J]. Nucleic Acids Research, 2017, 45(D1): D170-D176.

[19] MITCHELL AL, ALMEIDA A, BERACOCHEA M, BOLAND M, BURGIN J, COCHRANE G, CRUSOE MR, KALE V, POTTER SC, RICHARDSON LJ, SAKHAROVA E, SCHEREMETJEW M, KOROBEYNIKOV A, SHLEMOV A, KUNYAVSKAYA O, LAPIDUS A, FINN RD. MGnify: the microbiome analysis resource in 2020[J]. Nucleic Acids Research, 2020, 48(D1): D570-D578.

[20] STEINEGGER M, MEIER M, MIRDITA M, VÖHRINGER H, HAUNSBERGER SJ, SÖDING J. HH-suite3 for fast remote homology detection and deep protein annotation[J]. BMC Bioinformatics, 2019, 20(1): 473.

[21] JOHNSON LS, EDDY SR, PORTUGALY E. Hidden Markov model speed heuristic and iterative HMM search procedure[J]. BMC Bioinformatics, 2010, 11: 431.

[22] REMMERT M, BIEGERT A, HAUSER A, SÖDING J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment[J]. Nature Methods, 2011, 9(2): 173-175.

[23] HE XH, YOU CZ, JIANG HL, JIANG Y, XU HE, CHENG X. AlphaFold2 versus experimental structures: evaluation on G protein-coupled receptors[J]. Acta Pharmacologica Sinica, 2023, 44(1): 1-7.

[24] REY J, MURAIL S, de VRIES S, DERREUMAUX P, TUFFERY P. PEP-FOLD4: a pH-dependent force field for peptide structure prediction in aqueous solution[J]. Nucleic Acids Research, 2023, 51(W1): W432-W437.

[25] AKDEL M, PIRES DEV, PARDO EP, JäNES J, ZALEVSKY AO, MéSZáROS B, BRYANT P, GOOD LL, LASKOWSKI RA, POZZATI G, SHENOY A, ZHU WS, KUNDROTAS P, SERRA VR, RODRIGUES CHM, DUNHAM AS, BURKE D, BORKAKOTI N, VELANKAR S, FROST A, et al. A structural biology community assessment of AlphaFold2 applications[J]. Nature Structural & Molecular Biology, 2022, 29(11): 1056-1067.

[26] 廖世玉, 刘庆培, 陈福生. 基于 AlphaFold 2 和分子对接探讨非还原型聚酮合酶的碳甲基化程序[J]. 微生物学报, 2024, 64(1): 143-160.
LIAO SY, LIU QP, CHEN FS. C-methylation programming of non-reducing polyketide synthases: based on AlphaFold 2 and molecular docking[J]. Acta Microbiologica Sinica, 2024, 64(1): 143-160 (in Chinese).

[27] XIAO QJ, XU MX, WANG WW, WU TT, ZHANG WZ, QIN WM, SUN B. Utilization of AlphaFold2 to predict MFS protein conformations after selective mutation[J]. International Journal of Molecular Sciences, 2022, 23(13): 7235.

[28] HU LY, SALMEN W, SANKARAN B, LASANAJAK Y, SMITH DF, CRAWFORD SE, ESTES MK, PRASAD BVV. Novel fold of rotavirus glycan-binding domain predicted by AlphaFold2 and determined by X-ray crystallography[J]. Communications Biology, 2022, 5(1): 419.

[29] YANG QZ, SYED AAS, FAHIRA A, SHI YY. Structural analysis of the SARS-CoV-2 Omicron variant proteins[J]. Research (Washington, DC), 2021, 2021: 9769586.

[30] WAYMENT-STEELE HK, OJOAWO A, OTTEN R, APITZ JM, PITSAWONG W, HöMBERGER M, OVCHINNIKOV S, COLWELL L, KERN D. Predicting multiple conformations *via* sequence clustering and AlphaFold2[J]. Nature, 2023: 1-3.

[31] FOWLER NJ, WILLIAMSON MP. The accuracy of protein structures in solution determined by AlphaFold and NMR[J]. Structure, 2022, 30(7): 925-933.e2.

[32] IBRAHIM T, KHANDARE V, MIRKIN FG, TUMTAS Y, BUBECK D, BOZKURT TO. AlphaFold2-multimer guided high-accuracy prediction of typical and atypical ATG8-binding motifs[J]. PLoS Biology, 2023, 21(2): e3001962.

[33] LORENZ P, STEINBECK F, KRAUSE L, THIESEN H-J. The KRAB domain of ZNF10 guides the identification of specific amino acids that transform the

ancestral KRAB-A-related domain present in human PRDM9 into a canonical modern KRAB-a domain[J]. International Journal of Molecular Sciences, 2022, 23(3): 1072.

[34] JENDRUSCH M, KORBEL JO, SADIQ SK. AlphaDesign: a *de novo* protein design framework based on AlphaFold [J]. bioRxiv, 2021, 2021: 463937.

[35] GOVERDE CA, WOLF B, KHAKZAD H, ROSSET S, CORREIA BE. *De novo* protein design by inversion of the AlphaFold structure prediction network[J]. Protein Science: a Publication of the Protein Society, 2023, 32(6): e4653.

[36] ZENG D, XIN JB, YANG KY, GUO SX, WANG Q, GAO Y, CHEN HQ, GE JQ, LU Z, ZHANG LM, CHEN JY, CHEN YB, XIA NS. A hemagglutinin stem vaccine designed rationally by AlphaFold2 confers broad protection against influenza B infection[J]. Viruses, 2022, 14(6): 1305.

[37] WENG Y, PAN CH, SHEN ZY, CHEN SK, XU L, DONG XW, CHEN J. Identification of potential WSB1 inhibitors by AlphaFold modeling, virtual screening, and molecular dynamics simulation studies[J]. Evidence-Based Complementary and Alternative Medicine: eCAM, 2022, 2022: 4629392.

[38] REN F, DING X, ZHENG M, KORZINKIN M, CAI X, ZHU W, MANTSYZOV A, ALIPER A, ALADINSKIY V, CAO ZY, KONG SS, LONG X, MAN LIU BH, LIU YT, NAUMOV V, SHNEYDERMAN A, OZEROV IV, WANG J, PUN FW, POLYKOVSKIY DA, et al. AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor[J]. Chemical Science, 2023, 14(6): 1443-1452.

[39] FANG Q, HE XB, ZHENG XG, FU Y, FU T, LUO JY, DU YQ, LAN JJ, YANG J, LUO YN, CHEN XP, ZHOU NM, WANG Z, LYU JX, CHEN LJ. Verifying AXL and putative proteins as SARS-CoV-2 receptors by DnaE intein-based rapid cell-cell fusion assay[J]. Journal of Medical Virology, 2023, 95(7): e28953.

[40] MIRDITA M, SCHÜTZE K, MORIWAKI Y, HEO L, OVCHINNIKOV S, STEINEGGER M. ColabFold: making protein folding accessible to all[J]. Nature Methods, 2022, 19(6): 679-682.

[41] BARTAS M, SLYCHKO K, BRáZDA V, ČERVEŇ J, BEAUDOIN CA, BLUNDELL TL, PEČINKA P. Searching for new Z-DNA/Z-RNA binding proteins based on structural similarity to experimentally validated Zα domain[J]. International Journal of Molecular Sciences, 2022, 23(2): 768.

[42] YIN R, FENG BY, VARSHNEY A, PIERCE BG. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants[J]. Protein Science: a Publication of the Protein Society, 2022, 31(8): e4379.

[43] EVANS R, O'NEILL M, PRITZEL A, ANTROPOVA N, SENIOR A, GREEN T, ŽíDEK A, BATES R, BLACKWELL S, YIM J. Protein complex prediction with AlphaFold-Multimer[J]. bioRxiv, 2021, 2021: 463034.

[44] GAO M, NAKAJIMA AN D, PARKS JM, SKOLNICK J. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning[J]. Nature Communications, 2022, 13(1): 1744.

[45] WANG SW, LIN HY, HUANG ZX, HE YF, DENG XB, XU YJ, PEI JF, LAI LH. CavitySpace: a database of potential ligand binding sites in the human proteome[J]. Biomolecules, 2022, 12(7): 967.

[46] MA WJ, ZHANG SG, LI Z, JIANG MJ, WANG S, LU WG, BI XP, JIANG HS, ZHANG HG, WEI ZQ. Enhancing protein function prediction performance by utilizing AlphaFold-predicted protein structures[J]. Journal of Chemical Information and Modeling, 2022, 62(17): 4008-4017.

[47] VARADI M, ANYANGO S, DESHPANDE M, NAIR S, NATASSIA C, YORDANOVA G, YUAN D, STROE O, WOOD G, LAYDON A, ŽíDEK A, GREEN T, TUNYASUVUNAKOOL K, PETERSEN S, JUMPER J, CLANCY E, GREEN R, VORA A, LUTFI M, FIGURNOV M, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models[J]. Nucleic Acids Research, 2022, 50(D1): D439-d444.

[48] HEKKELMAN ML, de VRIES I, JOOSTEN RP, PERRAKIS A. AlphaFill: enriching AlphaFold models with ligands and cofactors[J]. Nature Methods, 2023, 20(2): 205-213.

[49] BLAABJERG LM, KASSEM MM, GOOD LL, JONSSON N, CAGIADA M, JOHANSSON KE, BOOMSMA W, STEIN A, LINDORFF-LARSEN K. Rapid protein stability prediction using deep learning representations[J]. eLife, 2023, 12: e82593.

[50] SUN JY, ZHU T, CUI YL, WU B. Structure-based self-supervised learning enables ultrafast prediction of

stability changes upon mutation at the protein universe scale[J]. bioRxiv, 2023, 2023: 552725.

[51] CHENG J, NOVATI G, PAN J, BYCROFT C, ŽEMGULYTĖ A, APPLEBAUM T, PRITZEL A, WONG LH, ZIELINSKI M, SARGEANT T, SCHNEIDER RG, SENIOR AW, JUMPER J, HASSABIS D, KOHLI P, AVSEC Ž. Accurate proteome-wide missense variant effect prediction with AlphaMissense[J]. Science, 2023, 381(6664): eadg7492.

[52] WANG J, LISANZA S, JUERGENS D, TISCHER D, WATSON JL, CASTRO KM, RAGOTTE R, SARAGOVI A, MILLES LF, BAEK M, ANISHCHENKO I, YANG W, HICKS DR, EXPÒSIT M, SCHLICHTHAERLE T, CHUN JH, DAUPARAS J, BENNETT N, WICKY BIM, MUENKS A, et al. Scaffolding protein functional sites using deep learning[J]. Science, 2022, 377(6604): 387-394.

[53] MENG QZ, GUO F, TANG JJ. Improved structure-related prediction for insufficient homologous proteins using MSA enhancement and pre-trained language model[J]. Briefings in Bioinformatics, 2023, 24(4): bbad217.

[54] BRYANT P, KELKAR A, GULJAS A, CLEMENTI C, NOé F. Structure prediction of protein-ligand complexes from sequence information with Umol[J]. bioRxiv, 2023, 2023: 565471.

[55] LIANG TJ, JIANG C, YUAN JY, OTHMAN Y, XIE XQ, FENG ZW. Differential performance of RoseTTAFold in antibody modeling[J]. Briefings in Bioinformatics, 2022, 23(5): bbac152.

[56] LIN ZM, AKIN H, RAO R, HIE B, ZHU ZK, LU WT, dos SANTOS COSTA A, FAZEL-ZARANDI M, SERCU T, CANDIDO S, RIVES A. Language models of protein sequences at the scale of evolution enable accurate structure prediction[J]. bioRxiv, 2022, 2022: 500902.

[57] KRISHNA R, WANG J, AHERN W, STURMFELS P, VENKATESH P, KALVET I, LEE GR, MOREY-BURROWS FS, ANISHCHENKO I, HUMPHREYS IR, McHUGH R, VAFEADOS D, LI XT, SUTHERLAND G, HITCHCOCK A, HUNTER N, BAEK M, DiMAIO F, BAKER D. Generalized biomolecular modeling and design with RoseTTAFold all-atom[J]. bioRxiv, 2023, 2023: 561603.

[58] DU BX, QIN Y, JIANG YF, XU Y, YIU SM, YU H, SHI JY. Compound-protein interaction prediction by deep learning: databases, descriptors and models[J]. Drug Discovery Today, 2022, 27(5): 1350-1366.

(本文责编 陈宏宇)