

人工智能赋能DNA计算：探索生物信息的新前沿

杨润^{1,†}, 庞华婕^{1,†}, 臧惠萍¹, 张瑞中¹, 张志成^{1,*}, 李希艳^{2,*}, 张立兵^{1,*}

¹天津大学理学院化学系, 有机集成电路教育部重点实验室, 天津市分子光电科学重点实验室, 天津 300072

²南开大学电子信息与光学工程学院, 光电子薄膜器件与技术研究所, 天津 300350

摘要: DNA计算作为一种开创性的技术, 利用生物分子执行数据存储、问题求解和逻辑操作等计算任务, 为突破传统计算的瓶颈提供了新途径。随着其技术的不断发展, DNA计算在复杂性、错误率及效率等方面的挑战愈加凸显。人工智能(AI)的发展为DNA计算带来了优化与提升的重要契机。特别是在数据分析、模型优化及错误修正领域, AI技术的运用显著提升了DNA计算的效率、精确性和稳定性。本文综述了AI算法的分类, 深入探讨了AI如何赋能DNA计算, 特别是在优化计算流程、强化逻辑门功能及改进错误修正机制方面。此外, 本文还总结了AI与DNA计算在生物信息学前沿的重要应用, 包括基因组学、蛋白质结构预测、疾病诊断与精准医疗等领域。展望未来, 随着AI技术与DNA计算的持续革新, 两者结合的应用范围将进一步拓宽, 有望成为推动生物科学发展的重要驱动力。

关键词: DNA计算; 人工智能; 生物分析; 生物信息学

中图分类号: G64; O6-39; Q523

Artificial Intelligence-Enabled DNA Computing: Exploring New Frontiers in Bioinformatics

Run Yang^{1,†}, Huajie Pang^{1,†}, Huiping Zang¹, Ruizhong Zhang¹, Zhicheng Zhang^{1,*}, Xiyan Li^{2,*}, Libing Zhang^{1,*}

¹ Key Laboratory of Organic Integrated Circuit, Ministry of Education & Tianjin Key Laboratory of Molecular Optoelectronic Sciences, Department of Chemistry, School of Science, Tianjin University, Tianjin 300072, China.

² Institute of Photoelectronic Thin Film Devices and Technology, College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300350, China.

Abstract: DNA computing, an innovative technology that utilizes biological molecules to execute computational tasks such as data storage, problem-solving, and logical operations, offers novel pathways to transcend the limitations of conventional computing. As DNA computing continues to evolve, challenges pertaining to complexity, error rates, and computational efficiency have become increasingly pronounced. The advent of artificial intelligence (AI) presents significant opportunities to optimize and enhance DNA computing. Particularly in the realms of data analysis, model optimization, and error correction, the integration of AI technologies has markedly improved the efficiency, accuracy, and stability of DNA computing. This paper provides an in-depth review of AI algorithm classifications and examines how AI empowers DNA computing, with a particular emphasis on optimizing computational workflows, enhancing logic gate functionality, and refining error correction mechanisms. Additionally, we summarize the pivotal applications of AI and DNA computing at the forefront of bioinformatics, including genomics, protein structure prediction, disease diagnosis, and precision medicine. Looking ahead, the ongoing innovation in AI and DNA computing is anticipated to

收稿: 2024-12-30; 录用: 2025-02-12; 网络发表: 2025-06-06

[†]共同第一作者, 对本文工作同等贡献

^{*}通讯作者, Emails: zczhang19@tju.edu.cn (张志成); xiyan.li@nankai.edu.cn (李希艳); libing.zhang@tju.edu.cn (张立兵)

基金资助: 2024年天津大学理学院研究生教育改革研究计划项目和国家自然科学基金面上项目(22474085, 22474088, 22375142)

further broaden their application scope, positioning them as a transformative force in the advancement of biosciences.

Key Words: DNA computing; Artificial intelligence; Bioanalysis; Bioinformatics

1 引言

DNA计算作为信息科学与生物科学交叉融合的一种新型计算范式^[1], 利用DNA分子执行一系列连续的生化反应来模拟复杂的计算流程。这一概念最早由Adleman于1994年提出, 他基于碱基配对的热力学平衡机制, 成功解决了数学领域中的经典问题——非确定性多项式(NP)完全问题中的哈密顿路径问题^[2]。随后, 在2004年, Okamoto等人将数字逻辑与DNA分子结合, 构建了DNA逻辑门, 为DNA计算的应用与发展奠定了基础^[3]。2009年, IBM宣布了一项利用DNA和纳米技术开发下一代微处理芯片的计划, 开启了DNA计算的新时代。自此之后, DNA计算技术取得了显著的进展, 并被认为是一种能够突破传统电子计算机局限的全新方法。与传统计算机不同, DNA计算通过模拟生物分子之间的相互作用和反应来进行信息的存储和处理, 具有极高的并行计算能力^[4]。这使得DNA计算在解决组合优化问题、大规模数据存储和传统计算机无法高效处理的复杂问题时展现出了巨大的潜力。

虽然DNA计算在理论上具有许多优势, 但在实际应用中仍然面临诸多挑战, 尤其是复杂性和固有错误问题^[5]。具体而言, DNA计算依赖于生物分子之间的化学反应, 但由于环境条件的多变性, 这些反应可能会产生不确定性, 导致计算结果出现误差。此外, DNA分子在实验中可能会受到污染甚至是降解, 进一步增加了计算过程中的数据丢失或错误的风险。因此, 如何提高DNA计算的稳定性、减少错误率并优化计算效率, 成为推动该领域发展的关键所在。

人工智能(AI)的迅猛发展为DNA计算的优化和提升带来了新的契机^[6]。AI凭借其强大的数据分析、预测以及优化能力, 革新了DNA计算领域^[7-9]。AI技术不仅能够高效处理DNA计算中的大规模数据集, 还能识别潜在问题并快速反馈优化结果^[10,11]。此外, AI算法的介入, 可以优化DNA序列设计、预测DNA结构, 从而提升计算效率和准确性, 同时降低实验成本^[12-14]。更为重要的是, AI在DNA计算中发挥着错误检测与修正的关键作用, 能够有效应对复杂生物反应和环境变化导致的计算误差, 从而能够增强DNA计算的稳定性和准确性^[15]。

AI与DNA计算的结合能够使研究人员更精准地操控计算流程并预测实验结果, 显著提高了DNA计算的效率与应用潜力, 进而为生物信息学开辟了新的研究视角。AI在数据分析、模式识别等方面的优势, 与DNA计算利用生物分子解决复杂计算问题的能力相得益彰, 共同推动了生物信息学的数据处理、分析与应用向高效、精准的方向发展。随着科技发展, AI将持续推动DNA计算的创新与发展, 并进一步拓展其应用。本文重点综述了AI算法的分类及其在DNA计算中的赋能作用, 特别是在优化计算过程、增强逻辑门功能以及改进错误修正机制方面的应用。此外, 我们还总结了AI与DNA计算在生物信息前沿领域的重要应用, 包括基因组学、蛋白质结构预测、疾病诊断与精准治疗等领域。

2 AI算法分类及其在DNA计算中的应用

2.1 AI算法的分类

AI算法主要可以分为监督学习、无监督学习、强化学习和深度学习四类(图1 A), 这些算法在DNA计算中展现出独特优势, 能够有效提升计算效率、优化实验设计, 并解决生化反应中的不确定性和误差问题。监督学习是最普遍的机器学习方法, 其依赖于已标记的数据集进行模型训练, 使模型能够基于历史数据学习并进行预测。常用的监督学习算法包括线性判别分析(LDA)^[16]、支持向量机(SVM)^[17]和随机森林(RF)^[18]。在DNA计算中, 监督学习被广泛用于DNA序列的设计与优化, 以及计算路径的高效规划。与监督学习不同, 无监督学习无需依赖于标注的数据集, 其重点在于自动分

析数据中的内在结构和模式来进行学习。常见的无监督学习算法包括主成分分析(PCA)^[19]、聚类分析(CA)^[20,21]和t分布随机邻居嵌入(t-SNE)^[22]。在DNA计算领域,无监督学习被广泛应用于模式识别、异常检测和数据降维等任务。此外,无监督学习还可以应用于DNA计算中的多任务处理,帮助模型从大量未标记的数据中找到潜在的计算规律。而强化学习是一种基于奖励机制的学习方法,机器通过与环境的交互获得反馈并调整行为策略。在DNA计算中,强化学习可以用于优化计算过程,通过实时反馈来调整DNA序列和实验设计。深度学习是机器学习的一个重要分支,通过多层神经网络进行特征提取和决策支持。常用的深度学习算法包括人工神经网络(ANN)^[23]、深度神经网络(DNN)^[24]、卷积神经网络(CNN)^[25]、循环神经网络(RNN)^[26]、长短时记忆网络(LSTM)^[27,28]以及生成对抗网络(GAN)^[29,30]等。在DNA计算中,深度学习被广泛应用于处理复杂的DNA计算任务,尤其是那些涉及大量数据和高维特征的计算问题。

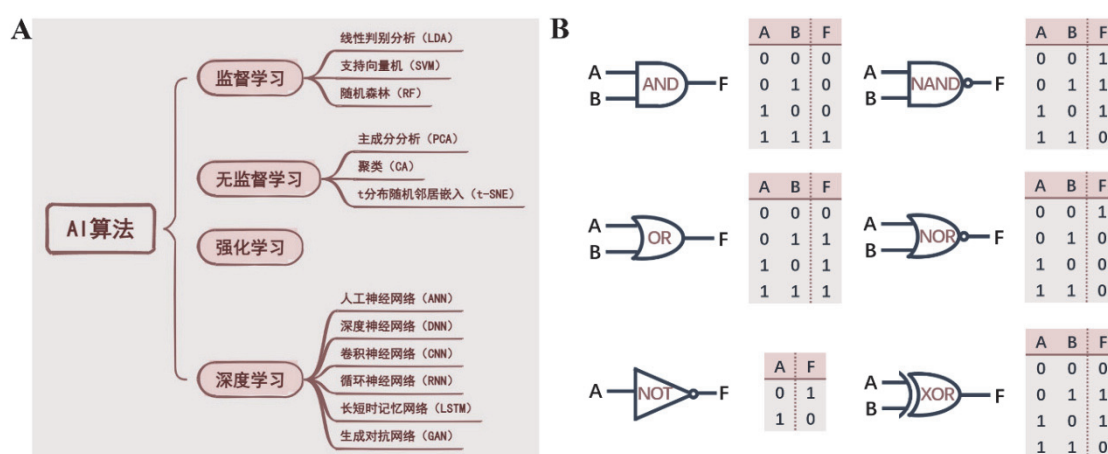


图1 (A) AI算法的分类; (B) 常见逻辑门的示意图和真值表

2.2 AI算法在DNA计算中的应用

DNA计算的核心在于利用DNA分子执行逻辑运算和数据处理任务,但在模型设计与优化上仍面临挑战。传统的DNA计算模型通常依赖于实验设置和简单生物学反应(如DNA杂交、酶催化等)的精确控制。尽管沃森-克里克碱基配对和单链置换反应(SDR)为输入和输出的调节提供了理论基础^[31,32],但这些传统方法在面对大规模数据处理和复杂计算任务时,受限于实验条件,难以同时实现高效与精准。特别是在实验路径选择和反应条件设定上,通常依赖于实验人员的手动调整。这种方式不仅效率低下,还可能因实验误差导致结果偏差。而AI算法的引入为解决这些问题提供了全新思路。AI通过优化DNA计算模型,能够显著提升计算速度与精确度。其中,强化学习和深度学习技术,凭借其自动学习数据模式与规律的能力,为DNA计算过程的优化提供了强大支持。例如,深度神经网络能有效预测DNA序列的反应效率,从而优化实验设计。通过实验结果的反馈学习,AI还能精准调整DNA计算模型的参数,进一步提高计算的精确性和效率^[33,34]。此外,AI在识别DNA计算模型中潜在瓶颈方面发挥着重要作用,特别是在多步骤DNA计算任务中,某些反应步骤可能会导致计算延迟或错误^[13,35],AI算法能够模拟这些步骤,深入分析反应效率,并据此自动优化实验条件,从而确保计算过程的流畅性。

分子逻辑门是一种模拟电子晶体管的生物化学器件,具有原位分子识别、逻辑计算和信号输出的功能。而DNA逻辑门则是基于DNA分子构建的逻辑运算装置,它利用DNA链的互补配对原则及链置换反应等生物机制,将生物分子或其他外界信息作为输入,通过DNA结构变化引发的各种表征结果作为输出,从而实现逻辑运算。这些逻辑功能包括AND、OR、NOT等基本运算,以及更复杂的

NAND、NOR、XOR运算(图1 B)^[36-38]。尽管DNA逻辑门的运算原理与传统计算机中的逻辑门类似,但其依赖生物分子而非电子信号,因而其面临着关键DNA序列设计的复杂性以及其在运行过程中的稳定性的重大挑战。每个DNA逻辑门的设计都需要根据其功能要求精确地设计特定的DNA序列及特定的二级结构,而传统的序列设计需要借助实验数据的反馈来进行,常受到计算复杂性、数据不完全以及人工的限制。随着AI的发展,机器学习已经成为核酸序列设计与结构预测的一个有力助手^[39]。具体来说,深度神经网络和卷积神经网络已经被成功应用于RNA二级结构的预测。基于深度学习的网络(如InceptionNet和DeepFold)通过直接学习核酸序列和二级结构之间的映射关系,能够提供比传统热力学方法更高的预测精度^[40,41]。目前,许多基于机器学习的工具已被开发用于核酸结构预测,如NUPACK,其结合热力学模型和动态规划算法,可实现RNA和DNA序列的二级结构预测、设计与分析^[42,43]。此外,Rtfold等基于深度学习的RNA二级结构预测工具,通过引入层次递归和自注意力机制,提升了模型的表达能力,在多项结构预测评估指标上表现优异^[44]。

由于DNA分子在不同环境条件下可能会发生变化,导致计算过程中出现不确定性甚至是错误运行。AI算法可通过模拟DNA逻辑门在不同条件下的反应,预测其稳定性。这些预测结果可以帮助优化DNA序列,避免因环境变化引发的不稳定性问题^[45]。此外,AI还能够帮助检测DNA序列中可能存在的设计缺陷,并在计算之前进行调整,确保计算过程的稳定性和可靠性^[46]。

DNA计算的显著特点之一是其反应过程具有高度并行性,这也使得错误可能在多个步骤中同时出现。所以AI的一个重要应用是实时检测这些错误并在不依赖人工干预的情况下,能够通过算法自动优化设计出有效的错误修正机制。例如,通过学习历史数据,AI能够在出现误差时自动识别出错误的来源,并迅速进行调整^[11,13]。

另外,深度神经网络(DNN)能够分析大规模DNA计算数据,学习其中的误差模式,并自动开发修正算法。这种深度学习模型不仅能够识别常见的错误,还能够敏锐地发现一些微小错误,甚至能够预测错误发生的概率,以便提前采取预防措施^[47]。凭借其卓越的自适应能力,DNN能够在不同实验条件下灵活调整DNA计算过程,从而最大限度地降低误差,提高计算精度。

3 AI与DNA计算的结合: 前沿研究

AI与DNA计算的结合为生物信息学领域开辟了全新的研究方向。DNA计算通过利用DNA分子的独特计算特性,致力于解决复杂的计算难题,而AI则在数据分析、模式识别及优化算法等方面提供了强大的技术支撑。AI与DNA计算深度融合,能够推动生物信息学在数据处理、分析及应用领域迈向更高效、更精准的全新阶段,特别是在基因组学研究、蛋白质结构预测以及疾病诊断等关键领域。

3.1 基因组学

基因组学是生物信息学中极具挑战性的领域之一,包含了海量的基因序列数据。基因组测序是解码生物体完整DNA序列的过程,是基因研究的基石。AI在基因组测序中的应用主要包括识别基因调控机制、定位疾病相关基因,以及解析疾病状态下的基因表达模式^[44]。例如,人工智能技术已被应用于绘制与癌症、糖尿病及心血管疾病相关的基因变异图谱,为疾病诊断、药物开发和基因编辑提供了重要依据。Yang等人采用聚类分析技术,聚焦于结肠腺癌(COAD)中的差异表达基因、调控网络、相关性及基因变异,成功界定了COAD中的分子亚型,有望为COAD的风险管理和个性化治疗策略提供新的见解与方向^[48]。Zheng等人提出了一种基于迁移学习的增强方法(Tlsub),能够精准检测广谱亚克隆比例的突变。此方法通过分析信号数据来解构测序信号与克隆比例之间关系的方法,并运用此构建了一个新的复制希尔伯特空间,能够过滤假阳性信号,实现了精准检测(图2 A)^[49]。此外,Liu等人聚焦于肾透明细胞癌(KIRC)中表皮生长调节基因的变异与数量变化,并运用基于差异表达的表皮生长调节因子的聚类分析,深入探究了其临床与免疫特征,从而阐述了表皮生长在肾透明细胞癌中的功能角色,进而构建了预测个体患者预后的列线图风险评估模型,展现了AI在临床精准医学中的应用潜力^[50]。

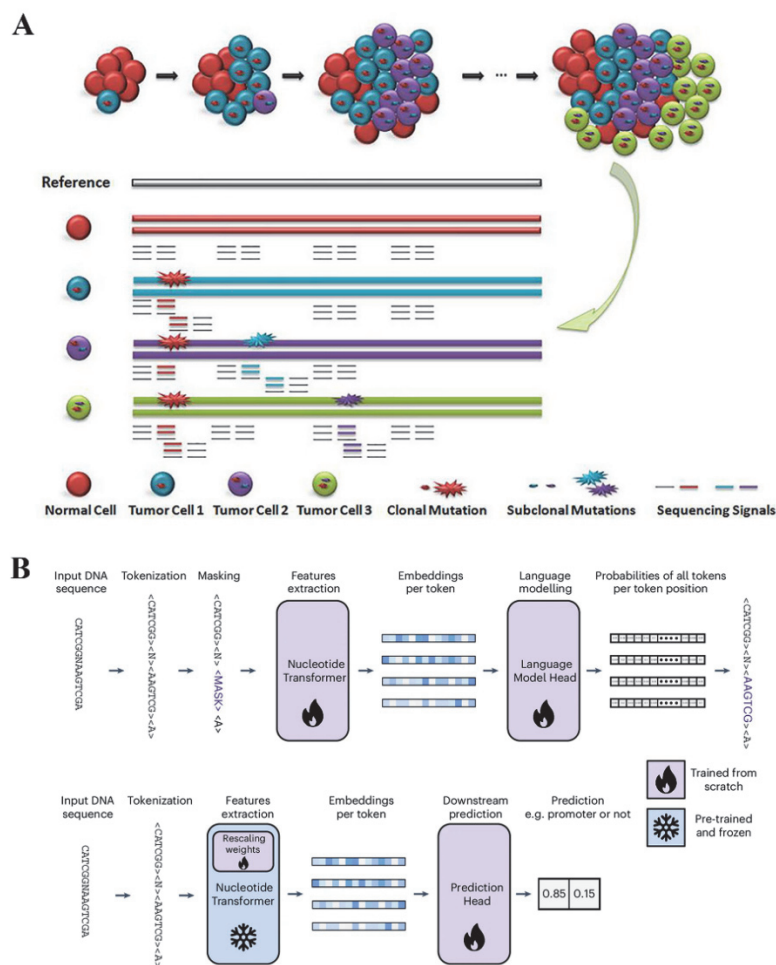


图2 (A) 亚克隆变异和检测信号示意图, 基于CC BY协议引自T Zheng, 2022^[49];
 (B) Nucleotide Transformer训练概述图及通过微调应用于下游基因组预测任务示意图,
 基于CC BY-NC-ND 4.0协议引自H Dalla-Torre等, 2024^[51]

最近一种基于自监督学习的AI模型(Nucleotide Transformer)被开发出来, 其能够通过大规模基因组数据的训练, 提取核苷酸序列的上下文特征, 并用于准确预测分子表型(图2 B)^[51]。该模型在基因组数据的处理上进行了创新, 训练过程使用了128个显示芯片(GPU)和16个计算节点, 模型规模可达25亿参数。与传统模型通常针对特定任务的限制不同, 其具有通用性, 能够应用于多种基因组学任务, 例如基因表达预测、功能元件预测以及序列设计等, 从而解决了基因组学中关于如何从基因型预测表型的关键问题, 突破了传统基因组学模型的局限。

尽管AI在基因组学中展示了广泛的应用前景, 但仍面临细胞类型特异性信息不足及跨任务迁移能力有限的问题。未来研究应通过多模态数据(如分子数据的多层次整合)提升模型性能, 同时开发全面的评估标准。DNA计算的并行计算能力与天然优势与AI技术结合, 必将为基因组数据分析的智能化与精准化带来革命性突破。

3.2 蛋白质组学

蛋白质组学研究长期面临的核心挑战是揭示氨基酸序列与其复杂三维结构之间的转化关系^[52], 突破这个瓶颈的关键在于揭示特定结构如何允许蛋白质执行特殊功能, 从而推动合成生物学和个性化医学等新兴领域的快速发展。

蛋白质在几乎所有生物过程中都发挥着重要作用，其功能与三维结构密切相关。因此，准确预测蛋白质结构对于理解生命机制、设计新的治疗药物、酶类及生物材料至关重要。而传统确定蛋白质结构的方法主要依赖于实验技术，如X光散射技术、核磁共振(NMR)光谱和冷冻电子显微镜(cryo-EM)等。尽管这些方法提供了有价值的研究数据，但它们通常耗时昂贵，并且受到复杂蛋白质结构的限制，因此阻碍了它们的广泛应用。而AI技术的应用标志着这一领域的重大突破。自从机器学习引入这个领域，已经有各种各样的模型来协助完成这项任务。

在众多计算方法中，深度学习技术的引入为蛋白质结构预测带来了革命性的变化，使得预测精度达到了前所未有的水平^[53]。谷歌DeepMind团队开发的AlphaFold便是这一领域的里程碑式成果。其创新的深度学习算法能够以原子级别的精度预测蛋白质结构，并且与传统实验方法的精确度相媲美^[54]。这一突破为科学探索和实际应用提供了全新的可能。随后，AlphaFold2的大规模架构优化进一步推动了基于AI的结构建模发展，并显著提高了模型的准确性^[55]。在此基础上，DeepMind还推出了多项更新，包括AlphaFold Multimer，支持多链蛋白复合物的预测^[56]。最新版本AlphaFold2.3进一步扩展训练数据，纳入更多的新型和大型蛋白质结构^[57]。受此启发，Baek等人开发了RoseTTAFold10模型，通过结合蛋白质一维序列、二维氨基酸相互作用和三维结构信息，进一步提升了结构预测能力^[58]。最近，RoseTTAFold All-Atom问世，成为下一代蛋白质、核酸、小分子、金属和化学修饰组件预测和设计工具，能够处理包括蛋白质-小分子、蛋白质-金属复合物等多种生物组装的结构预测^[59]。同时，DeepMind与同构实验室合作推出的AlphaFold3 (AF3)不仅突破了蛋白质组学分析的局限，还支持DNA、RNA及小分子配体等生物分子的结构建模，为药物开发提供了强大支持(图3 A, B)^[60]。然而，AlphaFold2模型由于缺乏必要的代码和数据，限制了新应用的开发。因此，OpenFold联盟应运而生，旨在开发开放源代码的AI蛋白质建模工具。近期发布的OpenFold模型在准确性和鲁棒性方面与AlphaFold2相当，已成为蛋白质建模领域的重要资源。这些工具突显了人工智能在精确预测蛋白质结构方面的巨大潜力，能够通过生成新的蛋白质序列和结构预测基因组变异对功能的影响，加速新药物和治疗方法的开发，并更高效地识别和验证新的药物靶点，尤其是那些缺乏实质性结构数据的靶点。

蛋白质的功能由其一级序列决定，遵循序列-结构-功能的黄金法则。因此，蛋白质的功能是由其氨基酸序列所决定的。深入理解序列与功能之间的关系，对于快速识别新蛋白质功能，并通过序列修饰进行从头设计具有重要意义。近年来，低成本高效的测序技术促使蛋白质序列数据量急剧增加^[61,62]，因此，如何确立序列与功能之间的关系，成为蛋白质设计中的关键问题。随着AI技术的发展，许多算法已被应用于蛋白质功能预测领域。例如，ProLanGO2方法借鉴自然语言翻译的设计原理，利用基于序列的递归神经网络进行蛋白质功能预测，其性能与其他序列方法及基于网络的NetGO2.0相当^[63]。Gligorijević等人提出的DeepFRI模型利用图形卷积网络和深度学习方法，结合更多序列信息，显著提升了预测准确性^[64]。同年，Yong等提出的DeepGraphGO方法，基于图神经网络，结合序列与高阶网络信息，超越众多先进技术(图3 C)^[65]。此外，张阳实验室整合多项技术，推出I-TASSER-MTD平台，实现从蛋白质序列到多结构域蛋白质结构和功能的全自动预测，涵盖序列结构域描述、空间几何约束预测、单域结构建模、多域组装及功能注释^[66]。另外，Oliveira等人提出了一种名为TEMPROT的创新方法^[67]，该方法通过对蛋白质序列进行预训练的可用架构中嵌入层进行微调 and 提取来实现。随后，他们又进一步优化出TEMPROT+来提升预测性能，这是前一种方法与分析序列相似性的局部比对工具(BLASTp)的集成，负责基于局部序列相似性进行同源性搜索。

除了基于序列的智能预测，基于蛋白质结构域^[68]、同源蛋白质功能转移^[69]及蛋白质网络依赖^[70]的自动化功能预测方法以及基于多种数据的综合蛋白质功能预测方法也取得了显著进展。然而，对于一些新测序或研究较少的蛋白质，缺乏关键的实验数据仍然是一个挑战。未来蛋白质功能预测的发展应集中在以下三方面：一是精确表征蛋白质结构与功能之间的关系，以提升预测的准确性；二

是通过功能间相关性精确定位多功能蛋白；三是整合蛋白质的结构特征、序列特征(包括全局和局部特征)以及基因组上下文特征，进一步提高预测的准确性。

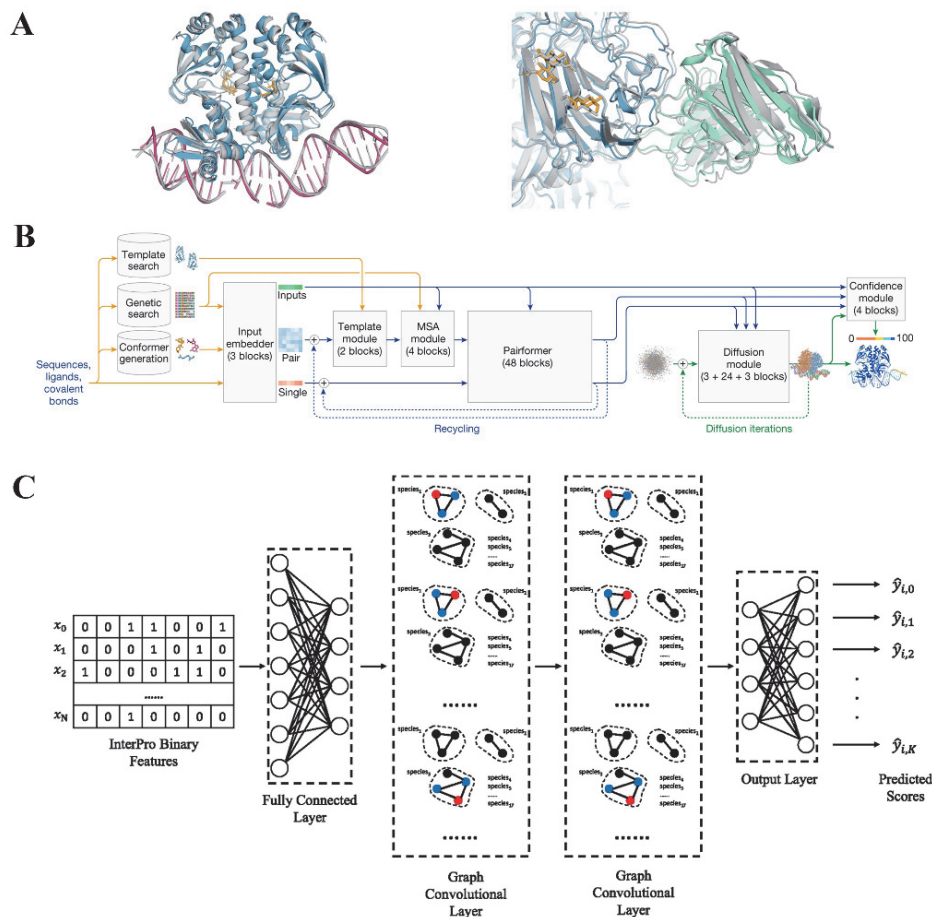


图3 AF3预测的生物分子复合物实例结构图(A)和用于推理的AF3架构示意图(B)，基于CC BY 4.0协议引自J Abramson等, 2024^[60]；(C) DeepGraphGO的程序示意图，基于CC BY 4.0协议引自R You等, 2021^[65]

尽管目前尚未有DNA计算直接应用于蛋白质组学，但凭借其高精度、大规模并行处理能力以及长期稳定性等优势，结合AI技术，DNA计算在蛋白质组学中的应用前景广阔。通过这种结合，可以高效提供蛋白质序列与结构的精确数据，快速处理海量信息，同时保证数据的长期保存与传输，从而深化我们对蛋白质功能与相互作用的理 解，推动这一领域的进一步发展。

3.3 疾病诊断与精准医疗

精准医疗的核心在于整合庞大的临床与生物标志物数据集，以实现患者的精确分类，从而为患者提供量身定制的预防和治疗方案^[71]。在此过程中，构建大规模数据库、采用高效的特征描述方法及运用尖端计算技术处理海量数据，为精准医疗的实施提供了重要支持。然而，数据异质性成为一大挑战，妨碍了数据的深入分析、有效整合、直观呈现和精确理解。

为了克服这一挑战，精准医疗迫切需要一种能够精准捕捉临床特征，并针对疾病特异性分子进行分类的创新方法^[72,73]。在此背景下，科研人员正全力投入生物信息学算法与机器学习模型的研发，这些前沿技术能够深度挖掘复杂数据背后的潜在规律，精准识别出与特定疾病密切相关的分子标记物，为精准医疗的深入实践奠定坚实基础。随着AI分类器的发展以及基于DNA反应的DNA计算的实

现, AI与DNA计算结合为精准医疗提供了一种强大且潜力巨大的通用分子分类方法^[74,75]。

Seelig及其团队设计了一种基于AI的分类器模型, 该模型能够将参数和数学函数转化为一类DNA探针报告基因, 用于实现早期癌症和呼吸道感染的多基因分类(图4 A, B, C)^[76]。此外, Han等人展示了一种分子分类器, 可以通过分析肺癌血清样本中不同microRNA (miRNAs)进而进行癌症诊断, 其诊断准确率为86.4%^[77]。在一个目标(DNA/RNA)与多个单链DNA报告基因之间的结合事件被统一地转化为一个权重分配的AI分析。然而, 由于这些结合过程的异质性, 将这种方法扩展到蛋白质或代谢小分子领域仍面临困难。因此, 基于DNA的多维分子分类器面临的挑战之一是开发一种能够以可编程方式将异构的多维分子信息转化为统一输出信号的信号报告器。

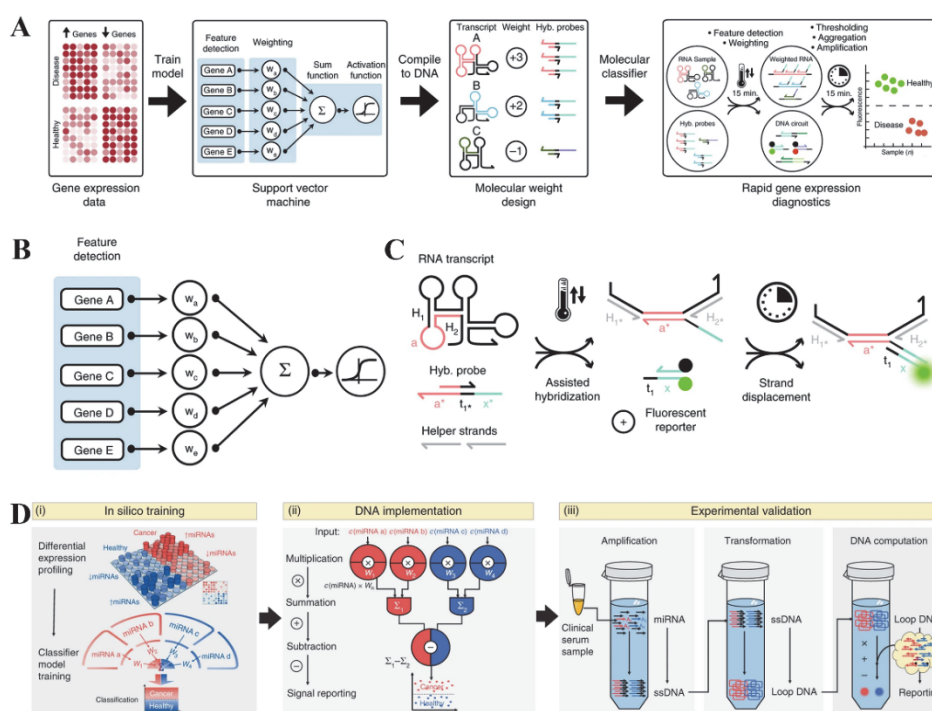


图4 (A, B, C) 用于基因表达诊断的分子分类快速成型通用框架示意图, 在参考文献[75]许可下转载, Springer Nature 2018年版权所有; (D) 用于非小细胞肺癌(NSCLC)诊断的DNA计算平台工作流程图, 在参考文献[79]许可下转载, Springer Nature 2020年版权所有

(A) 计算机模拟分类器对基因表达数据进行训练和验证并在分子水平上实现分类器示意图; (B) DNA 链置换级联检测特定RNA转录物的系统方法示意图; (C) 耦合基于DNA的电路与内源RNA转录物的分子机制的反应步骤组成示意图

在最近的研究中, DNA分子算术装置已被开发为机器学习(ML)分类器, 用于生物样本的疾病诊断^[78-80]。基于DNA的支持向量机(SVM)分类器通过级联加权乘法、求和和阈值来处理内源性mRNA信息, 从而有效区分病毒感染和细菌感染^[81]。具体来说, 此SVM分类器首先通过对包含7个宿主基因的RNA谱进行训练, 然后在一个由6名病毒感染患者和6名细菌感染患者组成的小队列中进行验证, 结果显示良好的分类性能。除了mRNA检测外, Zhang等人还构建了一个DNA算术计算平台, 用于分析血清样本中的miRNA谱, 以实现癌症诊断(图4 D)^[80]。鉴于miRNA序列的短促(~22 nt), 且在临床血液样本中的含量较低(通常低于皮摩尔水平), 首先使用指数后线性(LATE)-PCR扩增miRNA靶点, 然后将PCR产物转化为环DNA, 以激活分子计算过程。随后基于AI计算的miRNA谱训练的SVM分类器, 最终被解码为DNA计算方案, 包括乘法、求和、减法和信号报告。除此之外, Yang等人构建了空间定位的DNA集成电路分类器(DNA IC-CLA), 其可以在分子水平上执行基于神经形态架构的

计算,用于医学诊断^[82]。该分类器以二维DNA折纸为框架,内置局部处理模块作为计算核心执行基本的算术运算,从而能够高效地线性分类miRNA复杂模式。相比传统DNA电路,DNA IC-CLA在合成及临床样本中能够更迅速(约3小时)且更准确地诊断癌症。这些研究展示了AI算法与DNA分子计算在廉价、非侵入性、自动诊断工作流程中的巨大潜力。AI与DNA计算的结合为疾病诊断和个性化医疗领域带来了前所未有的机遇。DNA计算机能够高效处理大规模基因分析,而AI则能从海量数据中提取有价值的信息,为疾病的早期诊断和治疗提供精准支持。

4 挑战与未来展望

人工智能与DNA计算的融合,为DNA计算领域和生物信息领域带来了颠覆性的变革。AI在优化DNA计算模型、提升DNA逻辑门效能、确保计算精准度与稳定性方面,扮演了至关重要的角色。尤其是在破解DNA计算中的复杂难题方面,AI提供了全新的解决思路,极大推动了DNA计算技术的进步。AI与DNA计算的结合,不仅提升了大规模计算任务的处理效率,也在药物研发、疾病诊断等实际应用领域展现出巨大的潜力。随着DNA计算技术的不断发展,它正在逐步进入更广泛的实际应用场景,成为生命科学研究和医疗健康领域中的重要工具。

然而,DNA计算技术的发展仍面临许多挑战,尤其是在可扩展性、错误修正技术的精进、AI模型的可解释性和透明度等方面。当前,DNA计算在处理大规模数据时,仍然受到存储容量和计算速度的限制,如何有效应对这些瓶颈,成为了该技术进一步发展的关键。同时,尽管AI的引入为DNA计算提供了强大的支持,但AI模型的“黑箱”特性,仍然使得其在某些复杂应用中的可解释性和透明度问题成为障碍,如何提升AI模型的可解释性,帮助研究人员深入理解计算过程和结果,仍是未来研究的重要方向。

此外,随着技术的不断演进,跨学科的合作将变得尤为重要。未来的研究应聚焦于DNA计算与AI的深度融合,推动计算理论、生命科学与信息技术的跨界创新。在推动DNA计算技术突破的同时,AI可以进一步优化DNA计算的效率和准确性,提升其在药物筛选、疾病早期诊断、个性化治疗等领域的实际应用价值。结合量子计算等新兴技术,DNA计算的未来有望在大规模数据处理、智能分析和生物学应用等多个领域实现更为广泛的突破。

参 考 文 献

- [1] Kumar, S. N. *Am. J. Nanomater.* **2015**, *3* (1), 1.
- [2] Adleman, L. M. *Science* **1994**, *266* (5187), 1021.
- [3] Okamoto, A.; Tanaka, K.; Saito, I. *J. Am. Chem. Soc.* **2004**, *126* (30), 9458.
- [4] 高琳, 许进, 张军英. 电子学报, **2001**, *29* (7), 973.
- [5] Chen, X.; Ellington, A. D. *Curr. Opin. Biotechnol.* **2010**, *21* (4), 392.
- [6] Aldossary, A.; Campos-Gonzalez-Angulo, J. A.; Pablo-García, S.; Leong, S. X.; Rajaonson, E. M.; Thiede, L.; Tom, G.; Wang, A.; Avagliano, D.; Aspuru-Guzik, A. *Adv Mater.* **2024**, *36* (30), e2402369.
- [7] 鲍艳春, 石彩霞, 张传强, 谷明娟, 朱琳, 刘在霞, 周乐, 马凤英, 娜日苏, 张文广. 遗传, **2024**, *46* (9), 701.
- [8] Nguyen, E.; Poli, M.; Durrant, M. G.; Kang, B.; Katrekar, D.; Li, D. B.; Bartie, L. J.; Thomas, A. W.; King, S. H.; Brix, G.; *et al.* *Science* **2024**, *386* (6723), eado9336.
- [9] Sanabria, M.; Hirsch, J.; Joubert, P. M.; Poetsch, A. R. *Nat. Mach. Intell.* **2024**, *6* (8), 911.
- [10] Avsec, Ž.; Agarwal, V.; Visentin, D.; Ledsam, J. R.; Grabska-Barwinska, A.; Taylor, K. R.; Assael, Y.; Jumper, J.; Kohli, P.; Kelley, D. R. *Nat. Methods.* **2021**, *18* (10), 1196.
- [11] Kabir, A.; Bhattarai, M.; Peterson, S.; Najman-Licht, Y.; Rasmussen, K. Ø.; Shehu, A.; Bishop, A. R.; Alexandrov, B.; Usheva, A. *Nucleic Acids Res.* **2024**, *52* (19), e91.

- [12] Zhang, P.; Wei, L.; Li, J.; Wang, X. *Natl. Sci. Rev.* **2024**, *11* (11), nwae343.
- [13] Yang, A.; Zhang, W.; Wang, J.; Yang, K.; Han, Y.; Zhang, L. *Front. Bioeng. Biotechnol.* **2020**, *8*, 1032.
- [14] Dixit, S.; Kumar, A.; Srinivasan, K.; Vincent, P. M. D. R.; Ramu-Krishnan, N. *Front. Bioeng. Biotechnol.* **2024**, *11*, 1335901.
- [15] Xie, R.; Zan, X.; Chu, L.; Su, Y.; Xu, P.; Liu, W. *BMC Bioinformatics.* **2023**, *24* (1), 111.
- [16] Pang, S.; Ozawa, S.; Kasabov, N. *IEEE. Trans. Syst. Man. Cybern. B. Cybern.* **2005**, *35* (5), 905.
- [17] Suykens, J. A. K.; Vandewalle, J. *Neural Process. Lett.* **1999**, *9*, 293.
- [18] Díaz-Uriarte, R.; Alvarez de Andrés, S. *BMC Bioinformatics.* **2006**, *7*, 1.
- [19] Lever, J.; Krzywinski, M. *Nat. Methods.* **2017**, *14* (7), 641.
- [20] Lu, J.; Getz, G.; Miska, E. A.; Alvarez-Saavedra, E.; Lamb, J.; Peck, D.; Sweet-Cordero, A.; Ebert, B. L.; Mak, R. H.; Ferrando, A. A.; *et al.* *Nature* **2005**, *435* (7043), 834.
- [21] D'haeseleer, P. *Nat. Biotechnol.* **2005**, *23* (12), 1499.
- [22] Belkina, A. C.; Ciccolella, C. O.; Anno, R.; Halpert, R.; Spidlen, J.; Snyder-Cappione, J. E. *Nat. Commun.* **2019**, *10* (1), 5415.
- [23] Schmidhuber, J. *Neural. Netw.* **2015**, *61*, 85.
- [24] Larochelle, H.; Bengio, Y.; Louradour, J.; Lamblin, P. *J. Mach. Learn. Res.* **2009**, *10* (1).
- [25] 周飞燕, 金林鹏, 董军. *计算机学报*, **2017**, *40* (6), 1229.
- [26] Grossberg, S. *Scholarpedia* **2013**, *8* (2), 1888.
- [27] Hochreiter, S.; Schmidhuber, J. *Neural Comput.* **1997**, *9* (8), 1735.
- [28] Chambers, J. D.; Cook, M. J.; Burkitt, A. N.; Grayden, D. B. *Front Neurosci.* **2024**, *18*, 1472747.
- [29] Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A. A. *IEEE Signal Process. Mag.* **2018**, *35* (1), 53.
- [30] Li, C.; Xu, K.; Zhu, J.; Liu, J.; Zhang, B. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44* (12), 9629.
- [31] Wang, S. S.; Ellington, A. D. *Chem. Rev.* **2019**, *119* (10), 6370.
- [32] Simmel, F. C.; Yurke, B.; Singh, H. R. *Chem. Rev.* **2019**, *119* (10), 6326.
- [33] 纪守领, 李进锋, 杜天宇, 李博. *计算机研究与发展*, **2019**, *56* (10), 2071.
- [34] Yue, T.; Wang, Y.; Zhang, L.; Gu, C.; Xue, H.; Wang, W.; Lyu, Q.; Dun, Y. *Int. J. Mol. Sci.* **2023**, *24* (21), 15858.
- [35] Robinson, T.; Harkin, J.; Shukla, P. *Bioinformatics* **2021**, *37* (13), 1785.
- [36] Fu, R.; Hou, J.; Wang, Z.; Xianyu, Y. *ACS Nano.* **2024**, *18* (22), 14754.
- [37] Xiao, M.; Wang, X.; Yao, Q.; Li, L.; Fan, C.; Pei, H. *Chem* **2024**, *10* (12), 3634.
- [38] Bardales, A. C.; Smirnov, V.; Taylor, K.; Kolpashchikov, D. M. *ChemBiochem.* **2024**, *25* (8), e202400080.
- [39] 胡子昂, 高利明, 余文颖. *中国药科大学学报*, **2024**, *55* (3), 335.
- [40] Fu, L.; Cao, Y.; Wu, J.; Peng, Q.; Nie, Q.; Xie, X. *Nucleic Acids Res.* **2022**, *50* (3), e14.
- [41] 刘元宁, 臧子楠, 张浩, 刘震. *吉林大学学报(工学版)*, **2025**, *55* (1), 297.
- [42] Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. *J. Comput. Chem.* **2011**, *32* (1), 170.
- [43] Fornace, M. E.; Huang, J.; Newman, C. T.; Porubsky, N. J.; Pierce, M. B.; Pierce, N. A. *ChemRxiv*. 2022; doi:10.26434/chemrxiv-2022-xv981
- [44] Lee, J. W.; Won, J. H.; Jeon, S.; Choo, Y.; Yeon, Y.; Oh, J. S.; Kim, M.; Kim, S.; Joung, I.; Jang, C.; *et al.* *Bioinformatics* **2023**, *39* (12), btad712.
- [45] Zhang, P.; Wei, L.; Li, J.; Wang, X. *Natl. Sci. Rev.* **2024**, *11* (11), nwae343.
- [46] Ma, Y.; Chen, S.; Qi, X.; Lu, Z.; Bi, K. *IEEE Trans. Nanobioscience* **2025**, *24* (1), 89.
- [47] Wang, R.; Chen, J. *Peer J. Comput. Sci.* **2024**, *10*, e2160.
- [48] Yang, W.; Luo, C.; Chen, S. *Front. Genet.* **2022**, *13*, 986325.
- [49] Zheng, T. *Front. Genet.* **2022**, *13*, 981269.
- [50] Liu, A.; Li, Y.; Shen, L.; Li, N.; Zhao, Y.; Shen, L.; Li, Z. *Front. Genet.* **2022**, *13*, 983445.
- [51] Dalla-Torre, H.; Gonzalez, L.; Mendoza-Revilla, J.; Lopez Carranza, N.; Grzywaczewski, A. H.; Oteri, F.; Dallago, C.; Trop, E.; de Almeida, B. P.; Sirelkhatim, H.; *et al.* *Nat. Methods.* **2025**, *22* (2), 287.
- [52] Chaaban, S.; Ratkevičiūtė, G.; Lau, C. *The Biochemist.* **2024**, *46* (2), 7.

- [53] 王冯璋, 刘源, 王初. *大学化学*, **2025**, *40* (1), 75.
- [54] Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A. W. R.; Bridgland, A.; *et al.* *Nature* **2020**, *577* (7792), 706.
- [55] Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; *et al.* *Nature* **2021**, *596* (7873), 583.
- [56] Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Židek, A.; Bates, R.; Blackwell, S.; Yim, J.; *et al.* Protein complex prediction with AlphaFold-Multimer. [2024-12-25]. <https://doi.org/10.1101/2021.10.04.463034>
- [57] Zidek, A. Technical_note_v2.3.0.Md. [2024-12-25]. https://github.com/google-deepmind/alphafold/blob/main/docs/technical_note_v2.3.0.md (2022)
- [58] Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; *et al.* *Science* **2021**, *373* (6557), 871.
- [59] Krishna, R.; Wang, J.; Ahern, W.; Sturmfels, P.; Venkatesh, P.; Kalvet, I.; Lee, G. R.; Morey-Burrows, F. S.; Anishchenko, I.; Humphreys, I. R.; *et al.* *Science* **2024**, *384* (6693), ead12528.
- [60] Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; *et al.* *Nature* **2024**, *630* (8016), 493.
- [61] Steinegger, M.; Söding, J. *Nat. Biotechnol.* **2017**, *35* (11), 1026.
- [62] Steinegger, M.; Söding, J. *Nat. Commun.* **2018**, *9* (1), 2542.
- [63] Hippe, K.; Gbenro, S.; Cao, R. ProLanGO2: Protein Function Prediction with Ensemble of Encoder-Decoder Networks//Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. ACM BCB 2020, Virtual Conference, September 21–24, 2020. New York: Association for Computing Machinery (ACM), 2020: 1–6.
- [64] Gligorijević, V.; Renfrew, P. D.; Kosciolk, T.; Leman, J. K.; Berenberg, D.; Vatanen, T.; Chandler, C.; Taylor, B. C.; Fisk, I. M.; Vlamakis, H.; *et al.* *Nat. Commun.* **2021**, *12* (1), 3168.
- [65] You, R.; Yao, S.; Mamitsuka, H.; Zhu, S. *Bioinformatics* **2021**, *37* (Suppl. 1), i262.
- [66] Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; *et al.* Language models of protein sequences at the scale of evolution enable accurate structure prediction. [2024-12-25]. <https://doi.org/10.1101/2022.07.20.500902>
- [67] Oliveira, G. B.; Pedrini, H.; Dias, Z. *BMC Bioinformatics* **2023**, *24* (1), 242.
- [68] Koo, D. C. E.; Bonneau, R. *Bioinformatics* **2019**, *35* (10), 1737.
- [69] Törönen, P.; Medlar, A.; Holm, L. *Nucleic Acids Res.* **2018**, *46* (W1), W84.
- [70] Gligorijević, V.; Barot, M.; Bonneau, R. *Bioinformatics* **2018**, *34* (22), 3873.
- [71] Vargas, A. J.; Harris, C. C. *Nat. Rev. Cancer* **2016**, *16* (8), 525.
- [72] Collins, F. S.; Varmus, H. *N. Engl. J. Med.* **2015**, *372* (9), 793.
- [73] Thomasian, N. M.; Kamel, I. R.; Bai, H. X. *Nat. Rev. Endocrinol.* **2022**, *18* (2), 81.
- [74] Benenson, Y.; Gil, B.; Ben-Dor, U.; Adar, R.; Shapiro, E. *Nature* **2004**, *429* (6990), 423.
- [75] Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. *Science* **2006**, *314* (5805), 1585.
- [76] Lopez, R.; Wang, R.; Seelig, G. *Nat. Chem.* **2018**, *10* (7), 746.
- [77] Zhang, C.; Zhao, Y.; Xu, X.; Xu, R.; Li, H.; Teng, X.; Du, Y.; Miao, Y.; Lin, H. C.; Han, D. *Nat. Nanotechnol.* **2020**, *15* (8), 709.
- [78] Su, H.; Xu, J.; Wang, Q.; Wang, F.; Zhou, X. *Nat. Commun.* **2019**, *10* (1), 5390.
- [79] Moerman, P. G.; Schulman, R. *Nat. Nanotechnol.* **2020**, *15* (8), 626.
- [80] Zhang, C.; Zhao, Y.; Xu, X.; Xu, R.; Li, H.; Teng, X.; Du, Y.; Miao, Y.; Lin, H. C.; Han, D. *Nat. Nanotechnol.* **2020**, *15* (8), 709.
- [81] Lopez, R.; Wang, R.; Seelig, G. *Nat. Chem.* **2018**, *10* (7), 746.
- [82] Yang, L.; Tang, Q.; Zhang, M.; Tian, Y.; Chen, X.; Xu, R.; Ma, Q.; Guo, P.; Zhang, C.; Han, D. *Nat. Commun.* **2024**, *15* (1), 4583.