

综述



丛尧, 中国科学院分子细胞科学卓越创新中心(上海生物化学与细胞生物学研究所)研究员、博士生导师, 曾兼任国家蛋白质科学研究(上海)设施冷冻电镜系统副总设计师。长期从事基于冷冻电镜的蛋白质质量控制大分子机器和纤毛运动调控大分子机器的动态与原位结构研究, 新冠病毒及肠道病毒的结构研究及抗体和抑制药物发展, 并致力于冷冻电镜重构算法及亲和载网等新方法的发展, 取得了一系列有国际影响力的研究成果。近五年来, 在*Nature*、*Mol Cell*、*Sci Adv*、*Nat Commun*、*Cell Res*、*Proc Natl Acad Sci USA*等国际知名学术期刊发表通讯及共同通讯作者论文20余篇。中国科学院“百人计划”、基金委首届“优秀青年”及上海市“优秀学术带头人”获得者, 并荣获中国生物物理学会“女科学家优秀科研成果奖”等。承担多项科技部、基金委、中国科学院的科研项目, 担任*Sci Rep*、*QRB Discov*、*ABBS*等期刊编委。任中国生物物理学会第十二届理事会理事、上海市生物物理学会第十二届理事会理事和中国生物物理学会分子生物物理分会委员等。

RNA的冷冻电镜结构研究与结构预测方法发展

彭嘉豪^{1#}, 李祖洋^{1#}, 丛尧^{1,2*}

(¹中国科学院分子细胞科学卓越创新中心, 上海生物化学与细胞生物学研究所, 核糖核酸功能与应用重点实验室, 中国科学院大学, 上海 200031; ²国科大杭州高等研究院生命与健康科学学院, 中国科学院大学, 浙江省系统健康科学重点实验室, 杭州 310024)

摘要: 近十年来冷冻电镜技术飞速发展, 应用冷冻电镜解析的蛋白质结构数量迅速增加, 各类结构预测软件也日臻成熟, 蛋白质高级结构的准确性和预测精度和速度得到了巨大提升。作为功能载体和遗传信息的中间体, RNA的潜在应用正不断地被发掘, 相关药物和疫苗研究蓬勃发展。但相较于蛋白质结构研究, RNA三维结构解析和预测仍面临挑战和机遇。本文将从RNA的冷冻电镜结构解析及其解析新策略、RNA结构预测方法和软件发展三个方面入手, 介绍其高级结构的研究进展, 旨在为RNA新结构的解析、功能型RNA的发现和RNA药物设计提供参考。

关键词: RNA; 冷冻电镜; 三维结构; 结构预测

Cryo-EM analysis and development of structure prediction methods on RNA

PENG Jiahao^{1#}, LI Zuyang^{1#}, CONG Yao^{1,2*}

(¹Key Laboratory of RNA Innovation, Science and Engineering, Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, University of Chinese Academy of Sciences,

收稿日期: 2024-08-08

基金项目: 国家自然科学基金重点项目(32130056); 中国科学院战略性先导科技专项B类(XDB0570000); 中国科学院上海基础研究试点计划(JCYJ-SHFY-2022-008)

[#]共同第一作者: 彭嘉豪, E-mail: pengjiahao2023@sibcb.ac.cn; 李祖洋, E-mail: lizuyang2020@sibcb.ac.cn

*通信作者: E-mail: cong@sibcb.ac.cn;

Shanghai 200031, China; ²Key Laboratory of Systematic Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China)

Abstract: Over the past decade, cryo-electron microscopy (cryo-EM) has rapidly advanced, becoming a powerful tool in structural biology. The number of protein structures resolved using cryo-EM has increased significantly, with notable improvements in the accuracy, prediction precision, and speed of protein tertiary structure determination. As functional carriers and intermediates of genetic information, RNA's potential applications are continuously being explored, leading to a boom in related drug and vaccine research. However, compared to cryo-EM studies on protein structures, the structural determination and prediction of RNA's three-dimensional structure still face challenges and opportunities. This review introduces the latest progress in RNA structure studies from three perspectives: RNA cryo-EM structure analysis, new strategies for RNA cryo-EM study, and the development of method and software for RNA structure prediction. It is hoped that these advancements will contribute to resolving new RNA structures, discovering functional RNA, and developing RNA-based drug designs.

Key Words: RNA; cryo-EM; three-dimensional structure; structure prediction

随着近一个世纪以来电子显微技术的发明和发展,以及冷冻电镜技术四十余年来的发展,尤其是伴随着近十年来冷冻电镜分辨率革命的到来,我们现在得以在原子分辨水平观测各类生物大分子的三维结构^[1]。并且2017年诺贝尔化学奖授予了在冷冻电镜技术领域做出杰出贡献的三位科学家: Jacques Dubochet、Joachim Frank和Richard Henderson,以表彰他们在冷冻电镜发展过程中做出的杰出贡献^[2]。近年来,自动化数据收取流程的建立、图像处理和三维结构重建软件的发展以及基于人工智能(artificial intelligence, AI)的结构模型预测方法和软件的飞速发展也极大地促进了冷冻电镜技术的应用,使得该技术成为生物大分子结构解析与功能阐释之间的重要桥梁和工具。根据蛋白质数据库(Protein Data Bank, PDB)的统计数据,通过冷冻电镜技术解析的蛋白质结构数量正以惊人的速度逐年增长。截至2024年5月25日,已有20 171个蛋白质结构被成功解析,凸显了冷冻电镜技术在蛋白质结构研究中的重要作用。

RNA作为DNA和蛋白质的信息媒介,其潜在功能也得到广泛的关注。除了众所周知的转运RNA(tRNA)、核糖体RNA(rRNA)和信使RNA(mRNA),还存在着一系列非编码但仍具生物学功能的RNA,称为非编码RNA(non-coding RNA, ncRNA)^[3]。自20世纪80年代发现第一个具有调控

功能的ncRNA基因—大肠杆菌中的micF RNA基因以来,ncRNA在生命活动中的独特功能逐渐被人们所认知,其在疾病发生与治疗中展现出潜在的应用价值^[4,5]。随着对RNA功能研究的不断深入,有关RNA结构的研究也随之增长。截至2024年5月25日,PDB中共计有1 830个纯RNA(RNA-Only)的结构,而其中通过冷冻电镜解析的RNA-Only结构仅有114个。这一数字远远小于蛋白质结构数据量,侧面反映了RNA结构研究起步较晚,且具有复杂高级结构的RNA不如蛋白质常见。同时,众多RNA功能依赖于序列特异性的一级结构或简单的二级结构,缺乏足够的长程相互作用以稳定整体结构^[6],极为动态,因此也给结构解析带来挑战。

PDB上最早的RNA-Only结构数据可以追溯到1978年,当时研究人员通过X-射线晶体衍射技术解析了酿酒酵母苯丙氨酸tRNA的结构,分辨率达2.50 Å和2.70 Å^[7,8]。而第一个应用冷冻电镜技术解析的RNA-Only结构则是在2000年,基于已有的核糖体模型,解析了大肠杆菌中23S与5S rRNA的结构^[9]。此后,对RNA的结构研究逐渐增多,但具有高级结构且功能明确的主要为tRNA和rRNA,后续的结果也多集中在这二者上。自2013年以来,随着冷冻电镜分辨率革命的到来和相关方法学的发展,RNA-Only结构的解析数量迅速增加,分辨率

也得到了显著提高, 最佳分辨率已达到3 Å以内。此外, 研究对象的种类也随之增多, 包括核酶^[10-18]、核糖开关^[10,19-22]以及人工设计的RNA高级结构等^[23,24], 为深入探究RNA的高级结构提供了新的视角和机会。除了直接的结构观察, RNA高级结构预测方法也在近年来迅速发展^[25], 不仅为RNA结构的解析提供了强有力的工具, 也为理解RNA的功能机制奠定了基础。有关利用核磁共振和X-射线晶体学方法解析RNA结构及应用冷冻电镜的RNA三维结构测定与药物研发等相关内容已有论述^[26,27]。本文将重点回顾冷冻电镜技术在RNA-Only结构研究中的应用, 探讨RNA结构解析策略以及RNA结构预测方法和软件的新发展, 旨在为未来的RNA结构与功能研究提供参考和启示。

1 基于冷冻电镜的RNA-Only结构研究进展

在2000年, PDB中记录了首个仅由RNA组成的分辨率为7.5 Å的冷冻电镜结构。这一分辨率水平尚不足以区分单个碱基的清晰密度, 因此无法

直接根据电镜密度进行精确建模。研究人员根据该冷冻电镜密度, 并结合蛋白与核酸的交联质谱数据, 模拟构建了23S和5S rRNA的原子模型(atomic model)^[28]。同年, PDB还公布了其他两个中低分辨率的冷冻电镜结构: 大肠杆菌中的tRNA-核糖体复合物在易位前后的结构, 以及φ29噬菌体的RNA包装马达, 分辨率分别为17.0 Å和20.0 Å^[29,30]。从2002年到2013年, PDB数据库中增加了24个RNA-Only结构, 这些结构主要涉及各种生物的tRNA或rRNA, 并且集中在它们的功能区域的结合位点或功能执行过程中的动态变化。其中, 2013年的10个RNA结构为同一实验室报道, 解析了大肠杆菌30S核糖体亚基的动态组装过程^[31]。并且由于分辨率有限, 直接建模困难, 研究者借助电镜密度, 结合RNA序列信息和交联质谱数据进行了建模分析。2013年之后, 随着直接电子探测相机的引入和三维重构软件性能的提升, 冷冻电镜的分辨率得到了显著提高^[32-35], 部分RNA的结构分辨率达到10 Å以内(表1)。这使得RNA的高分辨率结构解析成为可能。这不仅促进了对tRNA和rRNA等传统

表1 分辨率高于10 Å的RNA-Only结构

RNA名称	分辨率范围	参考文献	RNA名称	分辨率范围	参考文献
大肠杆菌23S/5S rRNA	7.50 Å	[9]	嗜热四膜虫核酶的错误构象	3.90 Å	[13]
酿酒酵母rRNA	8.60 Å	[40]	RNA折纸	4.08~5.71 Å	[23]
猪瘟病毒内部核糖体进入位点	8.50 Å	[41]	嗜热四膜虫核酶第二步自我剪切的过程	v2.35~3.41 Å	[14]
乳酸乳球菌II型内含子	4.50 Å	[39]	嗜热四膜虫I型内含子不同状态	2.84~4.05 Å	[16]
HIV-1 RNA二聚化信号	9.00 Å	[42]	broccoli-pepper适配体	4.43~4.55 Å	to be published
噬菌体phi29 prohead RNA	4.60 Å	[43]	嗜热四膜虫核酶第一步自我剪切的过程	2.98~4.14 Å	[18]
枯草芽孢杆菌T-box核糖开关与Gly-tRNA复合物	4.90 Å	[19]	地下热厌氧杆菌腾冲亚种腺苷钴胺素核糖开关	3.00~5.30 Å	[21]
分枝杆菌SAM-IV核糖开关	3.70~4.10 Å	[20]	抗核酸外切酶RNA接于核酶支架	3.40 Å	[22]
具核梭杆菌甘氨酸核糖开关	7.4~10.0 Å	[10]	石油热袍菌氟化物核糖开关接于核酶支架	2.99 Å	[22]
分枝杆菌SAM-IV核糖开关	4.70~4.80 Å	[10]	四膜虫核酶支架	2.44 Å	[22]
嗜热四膜虫核酶	6.80 Å	[10]	冠状病毒HKU5基因组5'茎环结构	4.70~7.30 Å	[44]
霍乱弧菌甘氨酸核糖开关	4.80~5.70 Å	[10]	RNA基序纳米结构	5.39~6.56 Å	[45]
新冠病毒移码刺激元件	6.90 Å	[46]	新冠病毒移码刺激元件	6.10 Å	[47]
与p10A结合的双链核表达元件B	5.60 Å	[48]	甲烷菌吡咯赖氨酸-tRNA	2.80 Å	[49]
L-21Sca I 核酶	3.05~3.14 Å	[11]	RNA聚合酶核酶	5.00 Å	[17]
雀麦花叶病毒tRNA类结构	4.30 Å	[50]	冠状病毒基因组5'茎环结构	5.90~7.10 Å	[51]
嗜热四膜虫I型内含子	4.10 Å	[15]	RNA折纸	5.94 Å	[24]
嗜热四膜虫核酶各构象	3.01~4.01 Å	[12]	DNA内切核酶HYER1及HYER2	2.58~3.41 Å	[52]

RNA结构的深入研究,也使得核酶、核糖开关等功能性RNA的研究得到了快速发展。

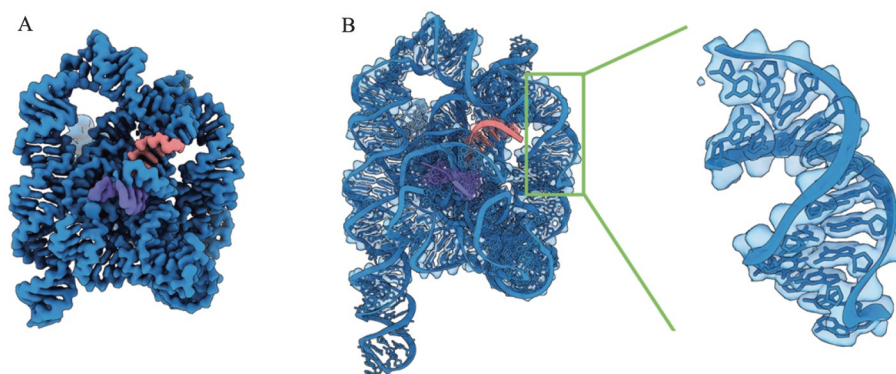
1.1 核酶

核酶(ribozyme)是一类具备催化功能的RNA分子,它们像蛋白质一样可以高效地催化特定的化学反应,有力地支撑了RNA世界假说(RNA World hypothesis)^[36,37]。核酶在生物体内扮演着至关重要的角色,参与RNA的剪接和加工等生物过程,其研究对于深入理解生命活动的分子机制具有重要意义^[38]。2016年,乳酸乳球菌中的Ⅱ型内含子的结构被成功解析^[39],并获得了其独立结构及其与LtrA蛋白形成复合体的高分辨率冷冻电镜结构,分别为4.50 Å和3.80 Å。2020年,研究人员借助新开发的Ribosolve工具,应用冷冻电镜解析了四膜虫核酶截短体L-21 *Sca* I核酶的6.80 Å结构^[10]。2021年,研究人员进一步解析了四膜虫Ⅰ型内含子(Tetrahymena group I intron)全长的冷冻电镜结构^[11],揭示了其独立与结合RNA底物的状态(分辨率达到3.14 Å和3.05 Å),并发现了两个全新的三级相互作用。2022年,研究人员发展了接吻环(RNA oligomerization-enabled cryo-EM via installing kissing loops, ROCK)技术^[15],使单个的四膜虫核酶形成稳定的二聚体或三聚体。通过精巧的构建策略并结合冷冻电镜技术,该团队解析了4.10 Å的三聚体核酶结构、2.85 Å的pre-2S结构和3.68 Å的post-2S结构,展现了四膜虫Ⅰ型内含子剪接过程中的动态变化,其中pre-2S结构是首个达到3 Å以内分辨率的RNA-Only结构。同年,四膜虫L-21

Sca I核酶体外组装过程中错误折叠形成的多个构象得以解析^[12],分辨率在3.53 Å至4.01 Å之间,并观察到了分辨率为3.01 Å的天然正确构象,为理解体外条件下错误构象向正确构象转变的过程提供了线索。与此同时,另一团队也独立解析了L-21 *Sca* I核酶错误折叠的构象^[13],分辨率为3.90 Å。2023年,四膜虫Ⅰ型内含子核酶在剪接过程中的不同构象得以解析,其中最佳分辨率达到了2.35 Å(图1)^[14,16,18],这也是迄今为止分辨率最高的RNA-Only冷冻电镜结构。此外,RNA聚合酶核酶(RNA polymerase ribozyme),这类在体外通过进化获得的、能以RNA为模板催化RNA复制的功能性RNA,也在2024年通过冷冻电镜技术得以解析,分辨率为5.00 Å,揭示了数种新的RNA基序,为未来的RNA结构设计和功能研究提供了新思路^[17]。

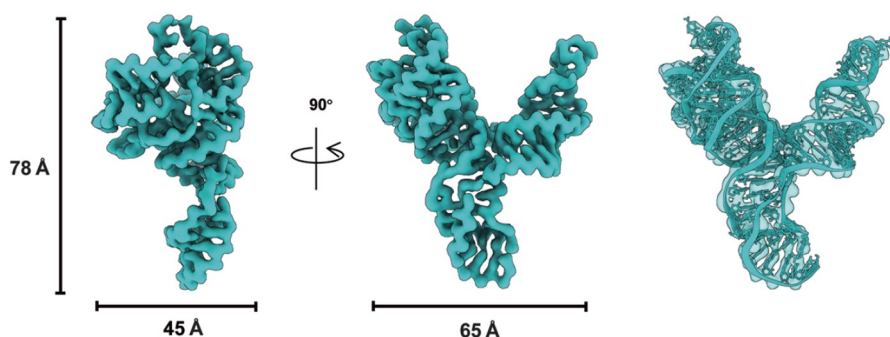
1.2 核糖开关

核糖开关(riboswitch)是定位于某些mRNA非翻译区的调控序列,能够形成特定的结构域。这些结构域当与相应的配体结合后,会发生构象变化,从而调控mRNA的表达^[53]。2019年,研究人员成功解析了枯草芽孢杆菌中的一种核糖开关——特异性识别甘氨酸的T-box及其与tRNA-Gly结合的全长结构,分辨率达到4.90 Å,为药物靶点的设计提供了新方向^[19]。同年,分支杆菌中SAM-IV核糖开关在结合与不结合SAM配体时的结构得以解析^[20],分辨率分别为4.10 Å与3.70 Å(图2)。值得注意的是,解析的SAM-IV核糖开关相对分子质量约为40 000,小于先前冷冻电镜所测定的最小蛋



A: 四膜虫Ⅰ类内含子的冷冻电镜结构(EMDB: EMD-33813),蓝色为内含子骨架,紫色代表P10,红色代表P1。B: 四膜虫Ⅰ类内含子的原子模型(PDB ID: 7yga)及电镜密度图

图1 四膜虫Ⅰ类内含子的冷冻电镜结构



EMDB: EMD-20755, PDB ID: 6ues

图2 分支杆菌SAM-IV型核糖开关的冷冻电镜结构及原子模型

白——相对分子质量为52 000的链霉素亲和素^[54], 展示了冷冻电镜在测定小相对分子质量RNA结构方面的潜力。2020年, 研究人员进一步解析了分支结核杆菌中的SAM-IV核糖开关在有无配体状态下的结构, 以及霍乱弧菌与具核梭杆菌甘氨酸核糖开关在有无配体状态下的结构, 分辨率范围从10 Å到4.7 Å不等^[10]。2023年, 腺苷钴胺核糖开关在结合配体时的四种构型以及无配体时的一种构型被报道^[21], 其冷冻电镜结构的分辨率在3.00 Å到3.50 Å之间, 而无配体时的分辨率为5.30 Å。同年, 研究人员利用四膜虫I型内含子作为支架, 将热孢菌的相对分子质量为17 000的氟化物核糖开关连接其上, 成功解析到了整体分辨率达2.99 Å的冷冻电镜结构, 其局部核糖开关的分辨率则达到了4.46 Å^[22]。这一成果表明, 利用特定RNA作为支架的方法在辅助解析小RNA结构方面具有广泛的应用前景。

1.3 其他的RNA结构

在2016年之后, tRNA冷冻电镜结构的分辨率得到了显著提高。2021年, 雀麦花叶病毒的tRNA结构被报道, 分辨率为4.30 Å^[50]。2022年, 研究人员揭示了重组tRNA在大肠杆菌核糖体结合位点中的变化^[55], 分辨率在2.60 Å至3.00 Å之间。研究显示, C8:A45碱基对在稳定tRNA结构中起着关键作用, 并影响核糖体的合成能力。通过突变该碱基对, 翻译效率得到了显著提升。2024年, 该团队进一步解析了结合核糖体A位点的吡咯赖氨酸tRNA结构^[49], 分辨率为2.80 Å。

此外, 其他功能性RNA的研究也取得了一系列成果。2016年, 研究人员应用冷冻电镜结合核

磁共振技术解析了HIV DIS二聚体的结构^[56], 分辨率为9.00 Å。2019年, ϕ29噬菌体中的pRNA环的结构得以解析, 分辨率为4.60 Å^[43], 发现该结构由5个而非6个pRNA分子组成, 更新了领域内对pRNA环的认知。2021年, 研究人员合作解析了相对分子质量仅28 000的新冠病毒移码刺激元件的冷冻电镜结构^[46], 分辨率为6.90 Å, 为反义寡核苷酸药物的设计提供了方向。该结构于2024年被进一步解析^[47], 分辨率提升至6.10 Å。2021年, 水稻转座元件TWIFB1与poly(A)结合时的构象被报道^[48], 分辨率达到5.60 Å。2024年, 新冠病毒基因组5'端元件stem-loop5 RNA的三级结构被报道^[44], 分辨率为4.70 Å。并同时解析了非典病毒、中东呼吸综合征冠状病毒和伏翼蝠冠状病毒HKU5的stem-loop 5的构象, 分辨率从5.90 Å到7.30 Å不等。2024年, 研究人员对新冠病毒、人类冠状病毒OC43、中东呼吸综合征冠状病毒和果蝠冠状病毒GCCDC1的stem-loop5结构进行了解析, 分辨率在5.90 Å至7.10 Å之间^[51]。

2 RNA结构解析策略

冷冻电镜适合解析相对分子质量较大的生物大分子或复合体的结构, 在解析相对分子质量较小的生物体系时存在一定的局限性。如在1995年理论预测提出对于相对分子质量大于38 000的分子, 它们的取向(orientation)可以通过完美图像进行定位, 但是随着图像质量的下降, 这种精确定位的难度急剧增加^[57]。这也被称为“冷冻电镜的38 000理论下限”。2019年, 应用冷冻电镜技术解析了52 000的链霉素亲和素结构, 这是当时应用该技术解

析的最小蛋白质结构^[54]。为了解析小蛋白的结构, 可以将其融合表达至作为支架的已知蛋白的结构上, 或者使用抗体增大蛋白整体相对分子质量^[58,59]。在RNA结构研究方面, 冷冻电镜解析的结构相对分子质量已经小于50 000, 但分辨率仍然受限^[20,22,46]。为了拓展可研究RNA体系的相对分子质量下限, 可以采用类似蛋白质研究的策略, 例如将小RNA接在已知结构的RNA分子支架上^[22], 或者让RNA形成多聚体^[15], 这不仅解决了相对分子质量小难以确定其颗粒取向的问题, 也有助于稳定其动态结构^[12,15]。

2.1 RNA分子支架

四膜虫 I 型内含子的结构已经被广泛解析, 分辨率达到3 Å以内^[15]。将其作为支架并融合表达小的RNA进行结构解析具有重要的应用潜力。通过对四膜虫 I 型内含子进行改造, 在不影响结构组装的位点融合表达待研究的RNA, 研究者已经获得了寨卡病毒的相对分子质量为23 000抗核酸外切酶RNA(exorib-onuclease-resistant RNA, xrRNA)结构(5.04 Å分辨率), 并发现该结构的P4区域与一种脑炎病毒晶体结构中的xrRNA存在很大差异。此外, 还解析到了另外两种相对分子质量约17 000的RNA结构, 分别是*Tamara*蝙蝠病毒的xrRNA以及*Thermotoga petrophila*的氟化物核糖开关, 分辨率分别为4.95 Å和4.46 Å^[22]。虽然这些结构尚为中等分辨率水平, 但大沟和小沟的结构已经可见, 对于评估RNA的折叠与行为信息具有重要作用。在采用该RNA分子支架策略时, 应注意尽量不影响待研究RNA分子的自然状态和组装。

2.2 RNA多聚体组装

ROCK是一种通过在RNA功能冗余区安装可配对突出环, 使RNA单体结构互补配对自组装形成多聚体的结构解析策略^[15]。由于多聚体的形成, RNA相对分子质量增大, 在电镜下衬度更为明显, 且摆动区变得更为稳定, 有助于后续的颗粒提取与结构解析。应用该策略, 研究人员选择性地四膜虫 I 型内含子的不同区域接上配对环, 从而在体外配对自组装得到该结构的多聚体形式, 解析了2.85 Å分辨率的四膜虫 I 型内含子核心结构。这也是第一个RNA-Only的分辨率好于3 Å的冷冻电镜结构, 实现了从头构建完整的原子模

型。并且电镜密度图中可以清晰地辨识出镁离子, 反映其在RNA催化中的配体作用。

上述RNA结构研究策略为研究人员提供了一条以更高的清晰度和精度探索RNA结构的新途径。

3 RNA三维结构预测

自AlphaFold模型问世以来, 深度学习算法在蛋白质结构预测方面取得了革命性突破, 即便在缺乏已知类似结构的情况下, 也能预测蛋白质的三维结构^[60]。截至2024年, AlphaFold蛋白质结构数据库已从最初的36万个预测结构激增至超过2.14亿个, 标志着蛋白质结构预测技术的爆炸性增长^[61,62]。相较之下, RNA结构预测明显滞后, 仍面临一些挑战, 尤其是在局部与全局结构的同步优化以及原子尺度精确预测方面^[25]。早在2011年, Laing等^[63]对RNA结构预测、分析和设计的方法学进行了全面调研, 指出在预测中大尺寸RNA(50-130 nt)的结构时, 即便是最佳方法也存在较大的均方根偏差(root-mean-square-deviation, RMSD), 平均达到了20 Å, 对于长度超过50 nt的RNA, 现有程序在预测长程接触和螺旋排列方面仍有改进空间。

尽管存在挑战, 近十年来RNA三维(3D)结构预测领域仍取得了显著进展, 预测精度不断提升, 新模型层出不穷, 展现出该领域巨大的发展潜力。下文中将介绍3dRNA、RNA-FM与E2efold-3D等软件在RNA结构预测方面的最新进展。

3.1 3dRNA

2012年开发的3dRNA软件, 通过将最小二级元件组装为发夹或双链体的策略, 不仅能自动和快速地预测RNA三级结构, 而且有效解决了预测大尺寸RNA和复杂拓扑结构的准确性问题^[64]。在RNA三维结构预测竞赛RNA-puzzles中, 3dRNA的RMSD值排名通常在前2或高于平均值。当使用真实的二级结构作为输入, 而非基于预测的二级结构时, 3dRNA的预测精度得到了显著提升, 显示了准确二级结构信息对于提高预测准确性的重要性。2015年, 新开发的3dRNAscore^[65]计算方法在传统的基于原子间距离评分的基础上, 加入了相邻碱基对中的RNA堆叠相互作用和骨架扭转角的

贡献。不仅能够识别合适的三级折叠, 还可以识别具有非经典碱基对的三级基序, 在性能上优于当时已有的方法。

2017年, Wang等^[66]进一步采用基于核苷酸协同进化的直接偶联分析(DI约束)来代替局部的3D模板, 优化了RNA 3D结构预测的准确性, 尤其在预测RNA多分支连接构象这一业内主要瓶颈问题上取得了显著进步。2020年, 3dRNA的模版库得到了改进^[67], 对先前预测的32种RNA数据集(12-101 nt)进行计算, 大多数RNA的预测精度明显提高。然而, 新模版库中仍有一半的结果不理想, 主要因为模版库中缺乏相应的模版, 侧面反映了RNA结构研究领域的数据匮乏问题。

2022年, 3dRNA推出了预测环形RNA的新功能^[68]。由于目前尚无环状RNA的实验三维结构, 3dRNA的该功能无法直接验证。为了评估其环化结构预测性能, 研究者采用了PKR, 一种含有两个RNA结构结合位点的蛋白。使用3dRNA预测POLR2A RNA的结构后, 发现其线性结构比环形结构更松散, 线性结构中只有一个可结合PKR的区域, 而环形结构则有三个结合位点, 这与自然状态下环形RNA结合效果更好的结论一致, 侧面印证了3dRNA在预测环形RNA结构上的潜力。

3.2 RNA-FM与E2Efold-3D

2022年, 研究人员开发了全新的RNA-FM模型(RNA foundation model)^[69]。该模型不同于以往依赖于RNA序列标签信息的计算方法, 而是采用自我监督的方式进行深度学习训练。通过利用未标记的2300万个ncRNA序列数据进行训练, RNA-FM可以在不使用任何标签的情况下推断出非编码RNA的序列和进化信息, 有效解决了缺乏带注释数据的困境。在三维结构预测方面, 基于RNA-FM的单一模型在长程预测的精度上甚至超过了使用100个模型的集成方法30%。

同年, 研究人员开发了第一个端到端的RNA 3D结构预测算法E2Efold-3D^[70], 利用深度学习架构, 能够直接从RNA序列进行从头的RNA结构预测。这种一体化的方法大大简化了预测流程, 提高了效率。E2Efold-3D除使用常用的包含进化信息的多序列比对外, 还利用了上述RNA-FM中的序列信息, 缓解了数据稀缺的问题。除训练三维结构

外, 损失函数中还引入了二级结构信息, 增加了模型预测的准确性。通过以上策略, E2Efold-3D在RNA 3D结构预测的精度及效率上均有提高。上述优势不仅为大规模的数据分析提供了有力的技术支撑, E2Efold-3D所生成的预测结果还能够为建模分析提供高质量的初始模型。

3.3 AlphaFold3

分别于2021年7月及2024年5月发布的AlphaFold2^[60]和AlphaFold3^[71]迅速推动了整个蛋白质结构预测领域的繁荣。AlphaFold3在AlphaFold2的基础上进行了改进, 旨在将多种生物分子(蛋白质、DNA、RNA、配体等)结构预测工具集成到单一神经网络中, 预测几乎所有生命大分子的结构和相互作用。相比于AlphaFold2, 研究团队做了如下改进, 以便包含更广泛的化学结构并提高数据使用效率: (1)减少多序列比对(multiple sequence alignments, MSAs)模块, 大幅提升计算速度; (2)引入了扩散模块, 实现了对生物分子原子坐标的直接预测, 无需依赖氨基酸框架或侧链角度。此外, 扩散模型在无结构区域可能产生幻觉的问题, 即模型可能在无结构的区域生成看似合理的结构。为了解决这一挑战, AlphaFold3引入了交叉蒸馏法, 通过把AlphaFold-Multimer v2预测的无序区域的结构添加到AlphaFold3的训练数据集中, 显著降低了AlphaFold3在预测过程中产生不合理结构的可能性, 提高了其整体的预测可靠性。

为了全面评估AlphaFold3在不同生物分子结构预测上的性能, 研究者们在蛋白质-配体相互作用、蛋白质-核酸复合体、RNA结构、共价修饰和蛋白复合体五个关键领域进行了准确性测试。在蛋白质-配体相互作用、共价修饰和蛋白复合体及蛋白质-RNA复合体结构预测方面, AlphaFold3均表现出色。在RNA结构预测领域, 尽管AIchemy_RNA2方法借助人类专家的辅助取得了领先, 但AlphaFold3在CASP15竞赛中对10个RNA结构的测试中, 其结果优于基于深度学习算法的其他预测软件, 这充分证明了AlphaFold3在生物分子结构预测方面的卓越性能。尽管RNA结构预测在某些方面取得了进展, 但其准确性仍然远低于蛋白质结构预测。这一差距的主要原因在于目前已知的RNA分子结构数量有限, 现有数据量不足以开发

出性能卓越的RNA结构预测模型。因此,利用冷冻电镜等实验方法来解析更多的RNA分子结构在当前阶段至关重要。随着RNA分子结构数据量的不断扩充,将为RNA结构预测技术的发展提供新的动能和突破。

4 总结与展望

经过长达二十余年的不懈努力, RNA的冷冻电镜结构解析取得了显著进步, 分辨率从不足7 Å提升至如今最佳的2.35 Å, 解析得到的结构数量也正逐年加速增长。然而, 这并不意味着结构解析已经到达了高速发展期。目前, 多数RNA结构的分辨率仍难以突破3 Å的水平。对于3.7 Å分辨率的电镜密度, 核酸碱基的特征几乎没有分离, 主链特征也没有得到很好描绘, 这使得模型构建严重依赖于计算机建模。因此, 针对RNA的电镜结构研究还有巨大的提升空间。值得一提的是, 近年来解析的RNA结构还包括了人工设计的RNA, 这从侧面反映出RNA的应用潜力正被逐步认识和发掘。例如, Φ29噬菌体的pRNA被设计为稳定的三聚体结构, 无需依赖金属盐即可形成稳定的纳米颗粒。经过进一步优化后还能形成四聚体, 行使药物递送的功能。在小鼠模型中, 这些结构未表现出明显的毒性或免疫原性, 展示了RNA作为有效抗癌药物载体的巨大潜力^[72]。RNA结构预测算法的发展同样令人瞩目。从十多年前预测结果的RMSD值普遍远大于5 Å, 到如今多数预测结果的RMSD值已控制在4 Å乃至3 Å以内。尽管在预测长RNA的结构上仍有不足, 但跨越性的进步为RNA结构预测提供了前所未有的准确性。预测速度方面也从过去的数小时以上减少到1秒以内, 这对于高通量RNA功能序列的快速发现具有重要意义。RNA多聚体或结构设计在助力RNA结构解析方面具有重要作用。但需要注意的是, 各种方法策略也有其局限性。例如, ROCK策略中接吻环的设计和放置位置依赖于RNA结构的初步预测, 且碱基对的增多或减少会对最终成环效果造成比较大的影响。虽然也可通过实验探索, 但无疑会增加实验周期。AlphaFold3和一系列新型预测工具的出现或许可以在一定程度上缓解这个问题。因此, 在实验中需要根据实际情况选择和调整方法。

RNA与蛋白质相互作用预测方面发展颇多。如RPI-Pred^[73], 在前人序列特征的基础上结合蛋白质和RNA的高阶结构, 可用于更高精度地获得蛋白质与RNA的相互作用信息。RBPsuite^[74], 一种基于深度学习方法开发的自动识别RNA中结合序列与结构基序的工具, 能预测线性和环形RNA上的蛋白质结合位点。RPI-EDLCN^[75], 用于预测ncRNA与蛋白质的相互作用。在RNA动态结构预测方面, PrismNet^[76]可捕获不同细胞类型中RNA的特异性结构变化及其与体内蛋白质的相互作用, 与仅使用RNA序列或预测的RNA结构来评估RNA与蛋白质相互作用的方法相比, 该模型综合考虑了RNA序列与体内结构数据, 揭示了可变位点对于疾病发生的影响。关于其他的RNA与蛋白互作预测模型, 目前已多有论述和比较^[77,78]。

RNA的原位结构研究在当前相对滞后, 主要由于RNA本身的动态性以及生物体内具有特定结构的RNA相对稀缺。虽然已有部分RNA病毒的原位结构与组装进程得到解析, 但与明显的蛋白密度相比, 其中的RNA密度难以辨别, 无法从头建模^[79,80]。

RNA自身结构研究正在为药物研发开辟新的方向。例如, 程序性-1核糖体移码(-1 ribosomal frameshifting, -1 RF)在SARS病毒合成重要复制蛋白时起关键作用, 而RNA假结(RNA pseudoknot)的存在和结构完整是实现-1 RF的前提。通过构建假结的三维结构模型并辅以化合物筛选, 研究人员成功鉴定了一种显著抑制-1 RF的配体^[81], 助力开发新型抗病毒药物。此外, 通过对SARS-CoV-2全基因组的结构研究, 发现不同冠状病毒间存在明显的相关变异^[82], 暗示这些区域在功能上的保守作用, 为设计相应的反义寡核苷酸药物提供了方向。随着全球对RNA相关研究投入的增加, 未来预计会涌现出更多的RNA相关药物。这些药物可能会以RNA结构为靶点进行创新设计, 也可能在脂质体或纳米载体等递送系统上进一步改进。冷冻电镜技术在这些研究中将扮演重要的角色。随着冷冻电镜硬件技术的进步和结构解析及RNA结构预测相关方法学 and 软件的发展, 有望解析更多具有重要生物学功能的RNA的冷冻电镜结构, 并通过预测软件快速捕捉RNA中的关键结构信息。

参考文献

- [1] Rüdénberg R. Origin and background of the invention of the electron microscope[M]. *Advances in Imaging and Electron Physics*. Elsevier, 2010, 160: 171-205
- [2] 李雪明. 冷冻电镜: 在原子尺度上观察生命. *物理*, 2017, 46(12): 809-816
- [3] Fu XD. Non-coding RNA: a new frontier in regulatory biology. *Funct Integr Genomics*, 2014, 1(2): 190-204
- [4] Delihás N. Discovery and characterization of the first non-coding RNA that regulates gene expression, micF RNA: A historical perspective. *World J Biol Chem*, 2015, 6(4): 272
- [5] Zhou R, Joshi P, Katsushima K, et al. The emerging field of noncoding RNAs and their importance in pediatric diseases. *J Pediatr*, 2020, 221: S11-S19
- [6] Zhang J, Ferré-D'Amaré AR. New molecular engineering approaches for crystallographic studies of large RNAs. *Curr Opin Struct Biol*, 2014, 26: 9-15
- [7] Sussman JL, Holbrook SR, Warrant RW, et al. Crystal structure of yeast phenylalanine transfer RNA: I. Crystallographic refinement. *J Mol Biol*, 1978, 123(4): 607-630
- [8] Hingerty B, Brown RS, Jack A. Further refinement of the structure of yeast tRNA^{Phe}. *J Mol Biol*, 1978, 124(3): 523-534
- [9] Mueller F, Sommer I, Baranov P, et al. The 3D arrangement of the 23 S and 5 S rRNA in the *Escherichia coli* 50 S ribosomal subunit based on a cryo-electron microscopic reconstruction at 7.5 Å resolution. *J Mol Biol*, 2000, 298(1): 35-59
- [10] Kappel K, Zhang K, Su Z, et al. Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat Methods*, 2020, 17(7): 699-707
- [11] Su Z, Zhang K, Kappel K, et al. Cryo-EM structures of full-length *Tetrahymena* ribozyme at 3.1 Å resolution. *Nature*, 2021, 596(7873): 603-607
- [12] Li S, Palo MZ, Pintilie G, et al. Topological crossing in the misfolded *Tetrahymena* ribozyme resolved by cryo-EM. *Proc Natl Acad Sci USA*, 2022, 119(37): e2209146119
- [13] Bonilla SL, Vicens Q, Kieft JS. Cryo-EM reveals an entangled kinetic trap in the folding of a catalytic RNA. *Sci Adv*, 2022, 8(34): eabq4144
- [14] Li S, Palo MZ, Zhang X, et al. Snapshots of the second-step self-splicing of *Tetrahymena* ribozyme revealed by cryo-EM. *Nat Commun*, 2023, 14(1): 1294
- [15] Liu D, Thélot FA, Piccirilli JA, et al. Sub-3-Å cryo-EM structure of RNA enabled by engineered homomeric self-assembly. *Nat Methods*, 2022, 19(5): 576-585
- [16] Luo B, Zhang C, Ling X, et al. Cryo-EM reveals dynamics of *Tetrahymena* group I intron self-splicing. *Nat Catal*, 2023, 6(4): 298-309
- [17] McRae EKS, Wan CJK, Kristoffersen EL, et al. Cryo-EM structure and functional landscape of an RNA polymerase ribozyme. *Proc Natl Acad Sci USA*, 2024, 121(3): e2313332121
- [18] Zhang X, Li S, Pintilie G, et al. Snapshots of the first-step self-splicing of *Tetrahymena* ribozyme revealed by cryo-EM. *Nucleic Acids Res*, 2023, 51(3): 1317-1325
- [19] Li S, Su Z, Lehmann J, et al. Structural basis of amino acid surveillance by higher-order tRNA-mRNA interactions. *Nat Struct Mol Biol*, 2019, 26(12): 1094-1105
- [20] Zhang K, Li S, Kappel K, et al. Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat Commun*, 2019, 10(1): 5511
- [21] Ding J, Deme JC, Stagno JR, et al. Capturing heterogeneous conformers of cobalamin riboswitch by cryo-EM. *Nucleic Acids Res*, 2023, 51(18): 9952-9960
- [22] Langeberg CJ, Kieft JS. A generalizable scaffold-based approach for structure determination of RNAs by cryo-EM. *Nucleic Acids Res*, 2023, 51(20): e100
- [23] McRae EKS, Rasmussen HØ, Liu J, et al. Structure, folding and flexibility of co-transcriptional RNA origami. *Nat Nanotechnol*, 2023, 18(7): 808-817
- [24] Vallina NS, McRae EKS, Geary C, et al. An RNA origami robot that traps and releases a fluorescent aptamer. *Sci Adv*, 2024, 10(12): eadk1250
- [25] Wang X, Yu S, Lou E, et al. RNA 3D structure prediction: progress and perspective. *Molecules*, 2023, 28(14): 5532
- [26] Ma H, Jia X, Zhang K, et al. Cryo-EM advances in RNA structure determination. *Sig Transduct Target Ther*, 2022, 7(1): 58
- [27] 李珊珊, 安林凤, 苏昭铭, 等. RNA三维结构的测定: 基于冷冻电子显微镜技术的新视角. *中国科学: 生命科学*, 2024, 54(4): 637-650
- [28] Xiong L, Kloss P, Douthwaite S, et al. Oxazolidinone resistance mutations in 23S rRNA of *Escherichia coli* reveal the central region of domain V as the primary site of drug action. *J Bacteriol*, 2000, 182(19): 5325-5331
- [29] Simpson AA, Tao Y, Leiman PG, et al. Structure of the bacteriophage φ29 DNA packaging motor. *Nature*, 2000, 408(6813): 745-750
- [30] Agrawal RK, Spahn CMT, Penczek P, et al. Visualization of tRNA movements on the *Escherichia coli* 70S ribosome during the elongation cycle. *J Cell Biol*, 2000, 150(3): 447-460
- [31] Guo Q, Goto S, Chen Y, et al. Dissecting the *in vivo* assembly of the 30S ribosomal subunit reveals the role of RimM and general features of the assembly process. *Nucleic Acids Res*, 2013, 41(4): 2609-2620
- [32] Liao M, Cao E, Julius D, et al. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*,

- 2013, 504(7478): 107-112
- [33] Cao E, Liao M, Cheng Y, et al. TRPV1 structures in distinct conformations reveal activation mechanisms. *Nature*, 2013, 504(7478): 113-118
- [34] Kimanius D, Dong L, Sharov G, et al. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J*, 2021, 478(24): 4169-4185
- [35] Punjani A, Rubinstein JL, Fleet DJ, et al. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*, 2017, 14(3): 290-296
- [36] DasGupta S, Piccirilli JA. The varkud satellite ribozyme: a thirty-year journey through biochemistry, crystallography, and computation. *Acc Chem Res*, 2021, 54(11): 2591-2602
- [37] Shechner DM. Architecture of an RNA polymerase ribozyme illuminates the RNA World. *Trends Genet*, 2024, 40(4): 291-292
- [38] Walter NG, Engelke DR. Ribozymes: catalytic RNAs that cut things, make things, and do odd and useful jobs. *Biologist (London)*, 2002, 49(5): 199
- [39] Qu G, Kaushal PS, Wang J, et al. Structure of a group II intron in complex with its reverse transcriptase. *Nat Struct Mol Biol*, 2016, 23(6): 549-557
- [40] Armache JP, Jarasch A, Anger AM, et al. Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. *Proc Natl Acad Sci USA*, 2010, 107(46): 19748-19753
- [41] Hashem Y, des Georges A, Dhote V, et al. Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. *Nature*, 2013, 503(7477): 539-543
- [42] Zhang K, Keane SC, Su Z, et al. Structure of the 30 kDa HIV-1 RNA dimerization signal by a hybrid cryo-EM, NMR, and molecular dynamics approach. *Structure*, 2018, 26(3): 490-498.e3
- [43] Xu J, Wang D, Gui M, et al. Structural assembly of the tailed bacteriophage ϕ 29. *Nat Commun*, 2019, 10(1): 2366
- [44] Kretsch RC, Xu L, Zheludev IN, et al. Tertiary folds of the SL5 RNA from the 5' proximal region of SARS-CoV-2 and related coronaviruses. *Proc Natl Acad Sci USA*, 2024, 121(10): e2320493121
- [45] Sampedro Vallina N, McRae EKS, Geary C, et al. An RNA paranemic crossover triangle as a 3D module for cotranscriptional nanoassembly. *Small*, 2023, 19(13): e2204651
- [46] Zhang K, Zheludev IN, Hagey RJ, et al. Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *Nat Struct Mol Biol*, 2021, 28(9): 747-754
- [47] Peterson JM, Becker ST, O'Leary CA, et al. Structure of the SARS-CoV-2 frameshift stimulatory element with an upstream multibranch loop. *Biochemistry*, 2024, 63(10): 1287-1296
- [48] Torabi SF, Chen YL, Zhang K, et al. Structural analyses of an RNA stability element interacting with poly(A). *Proc Natl Acad Sci USA*, 2021, 118(14): e2026656118
- [49] Krahn N, Zhang J, Melnikov SV, et al. tRNA shape is an identity element for an archaeal pyrrolysyl-tRNA synthetase from the human gut. *Nucleic Acids Res*, 2024, 52(2): 513-524
- [50] Bonilla SL, Sherlock ME, MacFadden A, et al. A viral RNA hijacks host machinery using dynamic conformational changes of a tRNA-like structure. *Science*, 2021, 374(6570): 955-960
- [51] de Moura TR, Purta E, Bernat A, et al. Conserved structures and dynamics in 5'-proximal regions of Betacoronavirus RNA genomes. *Nucleic Acids Res*, 2024, 52(6): 3419-3432
- [52] Liu ZX, Zhang S, Zhu HZ, et al. Hydrolytic endonucleolytic ribozyme (HYER) is programmable for sequence-specific DNA cleavage. *Science*, 2024, 383(6682): eadh4859
- [53] Bu F, Lin X, Liao W, et al. Ribocentre-switch: A database of riboswitches. *Nucleic Acids Res*, 2024, 52(D1): D265-D272
- [54] Fan X, Wang J, Zhang X, et al. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Å resolution. *Nat Commun*, 2019, 10(1): 2386
- [55] Prabhakar A, Krahn N, Zhang J, et al. Uncovering translation roadblocks during the development of a synthetic tRNA. *Nucleic Acids Res*, 2022, 50(18): 10201-10211
- [56] Keane SC, Van V, Frank HM, et al. NMR detection of intermolecular interaction sites in the dimeric 5'-leader of the HIV-1 genome. *Proc Natl Acad Sci USA*, 2016, 113(46): 13033-13038
- [57] Henderson R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quart Rev Biophys*, 1995, 28(2): 171-193
- [58] Liu Y, Huynh DT, Yeates TO. A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat Commun*, 2019, 10(1): 1864
- [59] Coscia F, Estrozi LF, Hans F, et al. Fusion to a homooligomeric scaffold allows cryo-EM analysis of a small protein. *Sci Rep*, 2016, 6(1): 30909
- [60] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021, 596(7873): 583-589
- [61] Varadi M, Bertoni D, Magana P, et al. AlphaFold Protein

- Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res*, 2024, 52(D1): D368-D375
- [62] Varadi M, Anyango S, Deshpande M, et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res*, 2022, 50(D1): D439-D444
- [63] Laing C, Schlick T. Computational approaches to RNA structure prediction, analysis, and design. *Curr Opin Struct Biol*, 2011, 21(3): 306-318
- [64] Zhao Y, Huang Y, Gong Z, et al. Automated and fast building of three-dimensional RNA structures. *Sci Rep*, 2012, 2(1): 734
- [65] Wang J, Zhao Y, Zhu C, et al. 3dRNAscore: A distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic Acids Res*, 2015, 43(10): e63
- [66] Wang J, Mao K, Zhao Y, et al. Optimization of RNA 3D structure prediction using evolutionary restraints of nucleotide-nucleotide interactions from direct coupling analysis. *Nucleic Acids Res*, 2017, 45(11): 6299-6309
- [67] Zhang Y, Wang J, Xiao Y. 3dRNA: Building RNA 3D structure with improved template library. *Comput Struct Biotechnol J*, 2020, 18: 2416-2423
- [68] Zhang Y, Wang J, Xiao Y. 3dRNA: 3D structure prediction from linear to circular RNAs. *J Mol Biol*, 2022, 434(11): 167452
- [69] Chen J, Hu Z, Sun S, et al. Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. arXiv preprint, arXiv: 2204.00300, 2022.
- [70] Shen T, Hu Z, Peng Z, et al. E2Efold-3D: end-to-end deep learning method for accurate de novo RNA 3D structure prediction. arXiv preprint, arXiv: 2207.01586, 2022.
- [71] Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 2024, 630(8016): 493-500
- [72] Shu D, Shu Y, Haque F, et al. Thermodynamically stable RNA three-way junction for constructing multifunctional nanoparticles for delivery of therapeutics. *Nat Nanotech*, 2011, 6(10): 658-667
- [73] Suresh V, Liu L, Adjeroh D, et al. RPI-Pred: Predicting ncRNA-protein interaction using sequence and structural information. *Nucleic Acids Res*, 2015, 43(3): 1370-1379
- [74] Pan X, Fang Y, Li X, et al. RBPsuite: RNA-protein binding sites prediction suite based on deep learning. *BMC Genomics*, 2020, 21(1): 884
- [75] Li X, Qu W, Yan J, et al. RPI-EDLCN: an ensemble deep learning framework based on capsule network for ncRNA-protein interaction prediction. *J Chem Inf Model*, 2023, 64(7): 2221-2235
- [76] Sun L, Xu K, Huang W, et al. Predicting dynamic cellular protein-RNA interactions by deep learning using *in vivo* RNA structures. *Cell Res*, 2021, 31(5): 495-516
- [77] Pan X, Yang Y, Xia C, et al. Recent methodology progress of deep learning for RNA-protein interaction prediction. *WIREs RNA*, 2019, 10(6): e1544
- [78] Liu X, Duan Y, Hong X, et al. Challenges in structural modeling of RNA-protein interactions. *Curr Opin Struct Biol*, 2023, 81: 102623
- [79] Yao H, Song Y, Chen Y, et al. Molecular architecture of the SARS-CoV-2 virus. *Cell*, 2020, 183(3): 730-738
- [80] Xia X, Sung PY, Martynowycz MW, et al. RNA genome packaging and capsid assembly of bluetongue virus visualized in host cells. *Cell*, 2024, 187(9): 2236-2249.e17
- [81] Park SJ, Kim YG, Park HJ. Identification of RNA pseudoknot-binding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening. *J Am Chem Soc*, 2011, 133(26): 10094-10100
- [82] Manfredonia I, Nithin C, Ponce-Salvatierra A, et al. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res*, 2020, 48(22): 12436-12452