



肖铭, 四川大学计算机学院(软件学院)副教授, 硕士研究生导师, 软件工程系副主任。主要研究方向为生物医学大数据智能分析与处理, 承担国家博士后面上基金、四川省科技厅重点研发项目等基金多项, 参与科技部重大专项、国家自然科学基金重点国际(地区)合作研究项目、自然科学基金面上项目等。近年来在 SCI 杂志和国际会议发表相关论文 20 余篇, 并任多个 SCI 杂志及国际会议的审稿人。相关研究成果曾获四川省科技进步二等奖、吴文俊人工智能科技进步三等奖、教育部“互联网+”大赛国家铜奖、“挑战杯”国家特等奖等。

人工智能在 RNA 三维结构预测与 RNA 相互作用中的应用

韦 睿, 章 乐, 肖 铭

(四川大学计算机学院, 成都 610065)

摘 要: RNA 通过折叠为特定三维结构与蛋白质、小分子等发生特异性互作, 在基因调控、免疫应答等关键生命过程中发挥核心作用。然而, 由于 RNA 构象自由度高、实验结构数据有限, 所以它的相关建模任务面临较大挑战。近年来, 人工智能(AI)技术为 RNA 三维结构与互作研究开辟了新的途径。为此, 本文系统综述 AI 在 RNA 三维结构预测, 与 RNA 相互作用建模中的研究进展。首先, 本文梳理 RNA 三维结构预测的三类方法演进脉络。随后, 本文不仅围绕 RNA-蛋白互作, 介绍了整体互作预测、结合位点与界面残基预测, 以及复合物三维结构建模三类任务中, AI 方法的发展, 还梳理了 AI 在 RNA-小分子互作中的应用。最后, 本文分析了该领域在高质量数据匮乏、模型可解释性有限等方面的挑战, 并从生成式分子设计、多模态联合建模、RNA 基础大模型发展, 及可解释 AI 方法等方面, 给出了未来研究方向的建议, 以期为推动 AI 与 RNA 生物学的交叉融合研究提供参考。

关键词: RNA 三维结构预测; RNA-蛋白互作; RNA-小分子互作; 深度学习; 预训练模型

中图分类号: TP18; Q75 **文献标志码:** A **DOI:** 10.19907/j.0490-6756.260108

Applications of artificial intelligence in RNA 3D structure prediction and RNA interaction modeling

WEI Rui, ZHANG Le, XIAO Ming

(College of Computer Science, Sichuan University, Chengdu 610065, China)

Abstract: RNA performs central roles in essential biological processes including gene regulation and immune responses by adopting specific three-dimensional (3D) architectures and mediating precise interactions with proteins, small molecules, and other biomolecules. However, modeling RNA remains challenging due to its high conformational flexibility and the paucity of experimentally determined structures. Recent breakthroughs

收稿日期: 2026-04-10 修回日期: 2026-05-12 接受日期: 2026-05-22

基金项目: 国家科技部重大专项(2024ZD0532900); 国家自然科学基金面上项目(62372316); 四川省科技计划项目省院省校合作项目(2025YFHZ0066)

作者简介: 韦睿(1999—), 女, 安徽池州人, 博士研究生, 研究方向为人工智能. E-mail: 2024323040020@stu.scu.edu.cn

通信作者: 肖铭. E-mail: xiaoming@scu.edu.cn

in artificial intelligence (AI) have opened new avenues for probing RNA 3D structures and their interaction modeling. In this review, we systematically survey recent advances in AI-driven methodologies for RNA 3D structure prediction and interaction modeling. We begin by outlining the evolution of three predominant categories of RNA 3D structure prediction algorithms. Subsequently, we examine progress in AI applications for RNA-protein interaction modeling, encompassing global interaction prediction, binding site and interface residue identification, and 3D complex structure prediction. We also highlight emerging AI approaches for RNA-small molecule interaction modeling. Finally, we address prevailing challenges such as the scarcity of high-quality datasets and limited model interpretability, and propose future directions, including generative molecular design, multimodal integrative modeling, RNA foundation models, and explainable AI. This review aims to provide a comprehensive reference to catalyze interdisciplinary research at the nexus of AI and RNA biology.

Keywords: RNA 3D structure prediction; RNA-protein interaction; RNA-small molecule interaction; deep learning; pre-trained models

RNA作为由核苷酸组成的重要生物大分子,长期以来主要被视为遗传信息传递的中间载体,但近些年研究表明,尽管RNA广泛参与催化、基因表达调控与免疫应答等关键生命过程^[1-2],然而其背后的分子机制仍有待深入。要阐明这些机制,核心在于解析RNA如何通过其三维结构实现特异性识别与调控,即RNA必须折叠成特定构象以形成功能界面,进而与蛋白质、核酸或小分子发生互作,通过组装成动态复合体,从而实现其调控功能^[3-7]。生物学层面这些研究有助于揭示RNA在催化以及转录后调控、染色质修饰与代谢感知等过程中的分子机制,进而发现相关调控网络的运作机制^[8-9]。同时在医学层面,RNA相关异常^[10-12]与疾病密切相关,而且RNA靶向小分子在传染病、肿瘤以及神经退行性疾病等多个疾病领域均展现出治疗应用前景^[13-15]。因此,通过解析RNA三维结构及其互作模式,不仅有助于从分子层面解释疾病相关RNA异常的病理机制,也为靶向RNA的小分子筛选、结合位点设计与选择性优化奠定方法基础,从而支撑RNA药物研发及个性化诊疗策略的推进^[16-21]。

近些年来,RNA结构与互作预测方法快速演化,其中人工智能(AI)技术已成为提升其性能的关键推动力。传统RNA结构预测主要依赖物理建模、二级结构约束与模板信息^[3],但如今AI方法通过从序列及相关数据中学习表征,从而显著提升结构预测的准确性。另外,蛋白质结构预测中AI的成功应用^[22-23]也为RNA的序列-结构-功能建模提供了方法学参照。在RNA互作建模方面,AI方法已覆盖RNA-蛋白与RNA-小分子等多类相互

作用,并正由结合位点识别进一步走向复合体结构与作用机制的可解释建模。一开始,AI可从蛋白表面理化特征中学习RNA识别偏好,从而为RNA实现特异性识别提供可靠依据^[24]。随着建模目标从界面推断进一步扩展到复合体构象,端到端方法已经可以直接预测RNA-蛋白的三维结构,并给出置信度,为互作的结构解释提供依据^[25]。更进一步地,面向多分子体系的联合建模已能在统一结构框架下描述RNA、蛋白与小分子的相互作用,为机制解析与靶向设计提供一致表征^[26]。

尽管近年来围绕RNA三维结构^[3, 27]、RNA-蛋白互作^[28-29]以及RNA-小分子互作^[16, 30]的综述不断涌现,但现有文献多分别聚焦于RNA三维结构预测或特定类型互作机制,对AI如何应用于RNA三维结构预测及RNA相互作用建模,仍缺乏系统梳理。事实上,RNA的三维结构直接影响其与蛋白质或小分子的识别方式,而相互作用建模也常依赖结构信息来研究结合位点和复合物构象。基于此,本文聚焦AI在RNA三维结构预测及RNA互作建模中的最新前沿进展,总结RNA-蛋白与RNA-小分子互作的代表性方法与评估体系,并探讨当前研究的挑战与未来方向。

1 AI在RNA三维结构预测的应用

RNA三维结构预测的目标,是在给定序列及可用辅助信息的条件下构建合理的空间结构。该任务的难点与RNA本身的结构特性密切相关。相较于蛋白质,RNA主链具有更高的构象自由度,因此其三维结构预测需要在更大的构象空间中进行搜索,其生成和筛选合理结构候选的难度也更

高。但是, RNA 的二级结构通常较为明确, 碱基配对关系决定了相当一部分残基间的相互作用, 并形成茎、发夹环、内环和多分支环等基本结构单元^[3]。因此, RNA 三维结构预测既不是无约束的构象搜索, 也不是简单的序列到坐标映射, 而是在多层结构约束下逐步确定整体拓扑结构和局部相互作用的过程。

因此, 在比较不同 RNA 三维结构预测方法时, 既需要关注整体折叠是否正确, 也需要关注局部相互作用是否得到准确刻画。除 $F1$ 分数这种常见分类指标外, RNA 三维结构预测中还常用均方根偏差(RMSD)、模板建模得分(TM-score)和相

互作用网络保真度(INF)等指标。RMSD 反映预测结构与实验结构之间的原子坐标偏差; TM-score 反映预测结构与实验结构的整体拓扑相似性; INF 反映碱基配对、碱基堆叠等局部相互作用与实验结构的一致性, 数值越高通常表示局部相互作用预测越准确。

如图 1 所示, 围绕这一任务, 现有算法大致可归纳为三类: 主要依赖二级结构约束以及模板调用或片段组装的传统算法, 先预测关键结构信息、再用于结构生成与优化的 AI 非端到端算法, 以及直接以 RNA 三维结构预测为目标进行建模的 AI 端到端算法。

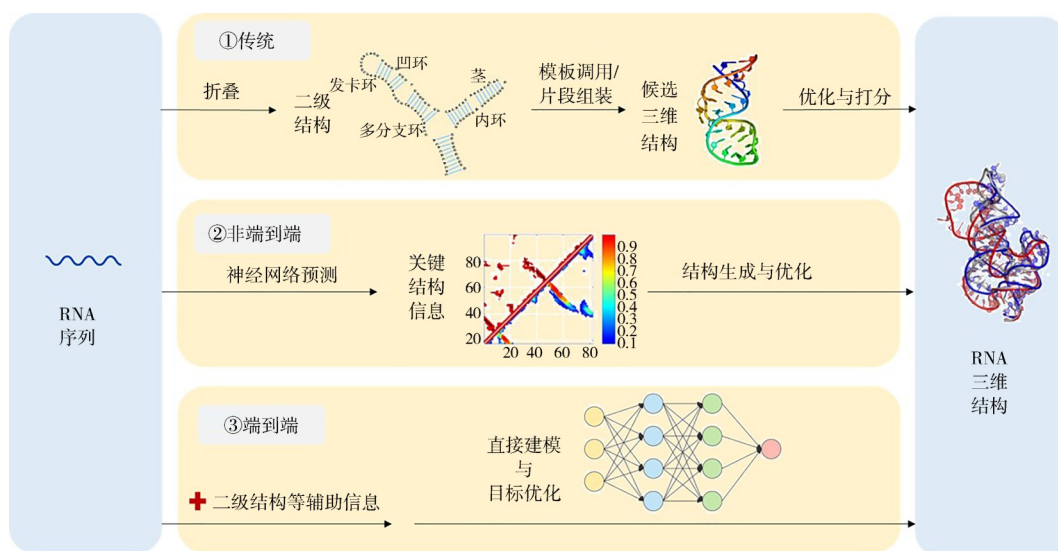


图 1 RNA 三维结构预测的三类主要算法框架

Fig. 1 Three major algorithmic frameworks for RNA 3D structure prediction

1.1 基于传统方法的 RNA 三维结构预测

传统方法的主流思路可以概括为: 先在二级结构等先验信息约束下生成候选结构, 再通过优化与打分排序筛选出合理结果。RNAComposer^[31]、Vfold^[32-33] 和 3dRNA v2.0^[34] 等代表性方法, 均以二级结构引导的模板调用或片段组装为核心。它们将茎、发夹环、内环和多分支环等局部结构单元映射到已有三维结构模块, 逐步构建 RNA 的初始三维骨架。已有研究表明^[33-35], 对于 Vfold 这类依赖约束的传统方法, 引入可靠的二级结构信息后, 模型对碱基配对等分子内相互作用的预测会更准确, 从而提升最终预测结果的质量。

在初始骨架构建完成后, 传统方法通常还需要通过结合能量最小化等函数对其予以优化, 从而改善局部几何结构并筛选最终模型^[36-37]。但这类方法的局限也较为明显, 一方面, 传统方法高度

依赖二级结构信息的准确性, 一旦二级结构预测出现偏差^[38], 后续的三维结构生成便会受到连锁影响; 另一方面, 候选结构的生成、优化和筛选仍主要依赖经验性能量函数或打分函数, 因此在复杂拓扑和长程相互作用的处理上仍然受限。尽管如此, 传统方法形成的“二级结构约束-模板调用或片段组装-优化与打分排序”框架, 依然为后续 AI 方法提供重要基础。

1.2 基于 AI 的非端到端 RNA 三维结构预测

与传统方法不同, 基于 AI 的非端到端算法不再主要依赖人工设定的建模规则, 而是直接从数据中学习 RNA 折叠相关的关键结构信息, 并在此基础上进行结构生成与优化。之所以首先沿着这一方向发展, 是因为 RNA 三维结构数据相对有限, 直接从序列预测完整三维结构仍然较为困难, 相比之下, 关键结构信息更容易学习, 也更适合作

为三维结构生成的约束。

1.2.1 关键结构信息预测

这一阶段的代表工作首先集中在关键结构信息预测上。例如,SPOT-RNA^[39]基于迁移学习技术,通过结合ResNet与LSTM来提升二级结构预测的稳定性,与其他方法相比,其在总体碱基对预测上的F1分数提升超过10%。SPOT-RNA-2D^[40]进一步融合配对概率矩阵与进化信息,输出距离型接触图以服务三维建模,不仅对长程残基接触的预测精确率达到66%~74%,还被证明是RNA三维结构预测的有效距离约束。总体来看,这类算法的主要作用在于通过提高三维结构生成所依赖的关键结构信息质量,从而在一定程度上降低对传统打分函数和大规模构象采样的依赖^[41]。

1.2.2 结构生成与优化

在关键结构信息预测能力提高之后,研究者开始进一步将这种与空间结构直接相关的信息用于RNA三维结构生成。这时,研究重点已从“预测关键结构信息”转向“如何利用这些信息生成三维结构”。虽然其整体流程仍保留传统方法的基本框架,但结构生成所依赖的关键约束,已逐步由经验规则和人工设定的约束,转向由模型学习得到的结构约束。

例如,epRNA^[42]采用卷积神经网络预测核苷酸间的距离矩阵,并将其作为三维结构重建的中间表征,随后通过离散分子动力学等后处理,将重建结果细化为全原子结构。DeepFoldRNA^[35]将学习得到的关键结构信息显式转化为结构优化中的约束项。该方法利用自注意力网络,预测残基间距离与取向分布,并将这些预测结果转化为结构优化中的约束项,再通过L-BFGS等方法生成RNA三维结构。因为DeepFoldRNA在均方根偏差(RMSD)和模板建模得分(TM-score)等指标上优于多种传统方法,所以这类方法在整体结构准确性和全局折叠相似性上具有明显优势。基于此,trRosettaRNA^[43]进一步发展了基于关键结构信息的建模思路。该算法使用多序列比对(MSA)结果和二级结构预测结果作为输入,并借鉴AlphaFold2^[44]的Evoformer思想,采用Transformer预测核苷酸间几何分布,并在能量最小化框架下生成三维结构。在15个独立测试样本上,trRosettaRNA的RMSD较DeepFoldRNA降低了3.3 Å。

总体来看,这一类AI算法的作用从提高关键结构信息预测精度,进一步扩展到指导RNA三维结构生成。因此,RNA三维结构预测对经验约束和大规模结构采样的依赖有所降低,模型在整体结构预测上的表现也得到明显改善。不过,这类方法仍需先预测关键结构信息,再通过后续优化得到最终结构,因此仍存在分步流程带来的信息损失问题。

1.3 基于AI的端到端RNA三维结构预测

为缓解非端到端方法中信息损失和误差传递等问题,RNA三维结构预测逐步发展出端到端建模方法。该类方法直接围绕RNA三维结构预测目标进行建模,通过让网络输出三维坐标或可用于三维结构重建的几何参数,从而使结构误差能够端到端反向传播。该方向之所以逐步可行,一方面是因为近年来AI技术的发展,提高了模型对序列长程依赖和空间关系的刻画能力;另一方面,自蒸馏和预训练模型等技术,也在一定程度上缓解了RNA三维结构数据有限的问题^[45-46]。

1.3.1 直接建模与目标优化

如图2所示,与非端到端方法相比,端到端方法通过强调围绕三维结构本身进行统一建模,从而更紧密地联合刻画全局拓扑与局部结构。例如,DRfold^[47]将序列与二级结构共同作为输入,通过Transformer学习RNA的序列表征和核苷酸对

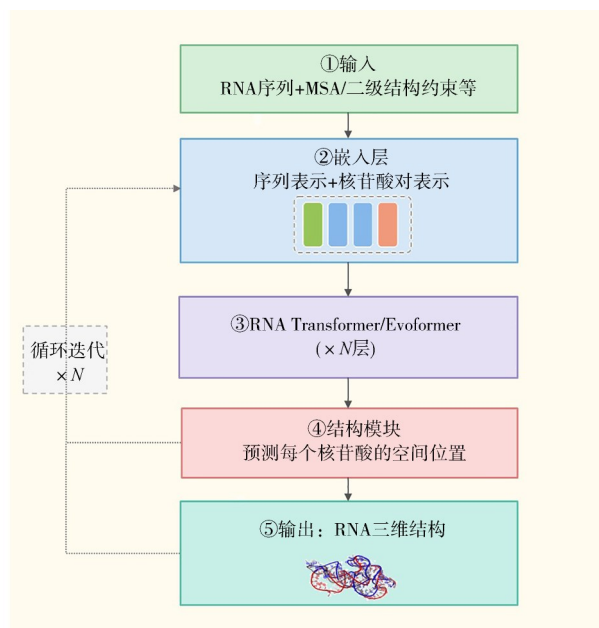


图2 端到端RNA三维结构预测通用框架

Fig. 2 General end-to-end framework for RNA 3D structure prediction

表示,并结合局部结构信息与几何约束完成三维结构重建。DRfold在非冗余测试集上的 TM-score 较其他算法提高了 73.3%,说明端到端建模能够明显提高预测结构与实验解析结构之间的整体相似性。

在 DRfold 的基础上, NuFold^[46]将 AlphaFold2 式端到端框架进一步适配到 RNA 三维结构预测中。该方法以 MSA 和预测的二级结构为输入,通过 Evoformer 学习核苷酸水平及核苷酸对之间的关系表示,并由结构模块直接生成 RNA 三维结构。NuFold 的一个关键改进是将灵活核苷酸中心表示法引入结构模块,从而更好地处理 RNA 三维结构。实验表明, NuFold 不仅降低了 RMSD, 而且其在相互作用网络保真度(INF)方面也优于其他方法。

1.3.2 预训练模型的引入

在端到端建模基础上,预训练模型的引入进一步缓解了 RNA 三维结构数据有限的问题。这类方法通过先在大规模 RNA 序列数据上学习通用表示,再将其迁移到三维结构预测任务中,从而提升模型的泛化能力与自动化建模能力。

例如, RhoFold+^[48]利用预训练模型 RNA-FM 提取 RNA 序列表征,并基于此,构建单链 RNA 三维结构自动化预测的端到端框架。RhoFold+ 在不同 RNA 家族和类型的数据上,均表现出较好的泛化能力,还可进一步预测 RNA 二级结构和螺旋间角度等结构特征。在 RNA-Puzzles 数据集^[49]上, RhoFold+ 实现了平均 RMSD 降低 4Å 的结果,并

最短可在 0.14 s 内完成准确建模,体现出较高的预测精度与计算效率。此外, RNAErnie^[50] 和 RiNALMo^[51]这类通用 RNA 预训练模型已在二级结构预测等下游任务中表现出较强能力, RNA 三维结构预测也是其后续的重要发展方向。

1.3.3 RNA 三维结构预测仍面临的挑战

如表 1 所示,尽管现有 RNA 三维结构预测方法已在整体结构预测和自动化建模方面取得明显进展,但公开盲测结果表明,该任务仍存在较大挑战。除了 RNA-Puzzles^[49], 结构预测关键评估(CASP)也是该领域具有代表性的盲测评估体系。CASP15^[52]是首次涉及 RNA 结构建模的 CASP 评测。其结果显示,现有方法在部分天然 RNA 上能够获得较合理的整体折叠,但在非经典碱基配对、氢键等局部结构细节方面仍有不足。CASP16^[53]进一步将核酸相关评测扩展到 RNA 单体、多聚体以及蛋白-核酸复合物等体系。评估结果显示,现有方法在具有相似三维结构模板的目标上通常能够取得较好表现。但对于缺乏可靠模板的核酸结构,尤其是涉及 RNA 与蛋白质等分子间相互作用的复杂体系,其预测准确性仍然受到明显限制。

与此同时,现有方法对长链 RNA、多链 RNA、环状 RNA 等复杂 RNA 类型的预测能力也仍然有限^[45]。因此, RNA 三维结构预测的进一步发展,不仅在于提高整体结构预测精度,也在于提升对复杂结构特征的建模能力,这也将为后续 RNA 与蛋白质、小分子等互作研究奠定更可靠的结构基础。

表 1 RNA 三维结构预测主要方法类型比较

Tab. 1 Comparison of major method types for RNA 3D structure prediction

方法类型	代表方法	主要优势	主要局限	适用场景
传统建模方法	RNAComposer、Vfold、3dRNA v2.0	流程成熟,可解释性较强,计算成本较低	依赖二级结构、模板或片段库质量,对复杂拓扑和长程作用刻画有限	小型 RNA 或模板信息较充分的结构预测
AI 非端到端方法	SPOT-RNA-2D、epRNA、DeepFoldRNA、trRosettaRNA	可学习接触、距离和取向等结构约束,提升结构生成质量	仍需后处理或能量优化,存在误差传递	结构约束的 RNA 三维结构建模
AI 端到端方法	DRfold、NuFold	自动化程度较高,可直接围绕三维结构目标建模	依赖训练数据覆盖范围,对复杂 RNA 和分布外样本泛化有限	无明显模板条件下的自动化结构预测
预训练表征方法	RhoFold+	可利用大规模序列信息,提升表征能力和预测效率	对多链 RNA、环状 RNA 及复合物建模仍有限	快速结构预测、结构特征预测及下游建模

2 AI在RNA-蛋白互作的应用

RNA 与蛋白质的相互作用广泛参与转录后调控、剪接、翻译与 RNA 稳定性维持等核心生命过

程。然而,实验方法已解析的 RNA-蛋白复合物三维结构数量,远少于已报道的 RNA-蛋白互作关系数量,大量已识别的互作关系在结合界面和复合物空间构象层面,仍缺乏直接的结构证据^[28]。因

此 AI 算法成为填补这一缺口的关键工具。但 RNA-蛋白互作的建模目标并不单一,从判断两个分子是否互作,到定位结合位点或关键残基,再到重建复合物的三维空间结构,不同任务在输入信息、建模策略、评价方式及适用场景上均存在显著

差异^[54]。

本节按 RNA-蛋白互作的建模目标,如图 3 所示,将现有 AI 算法划分为 3 类:整体互作预测、结合位点与界面残基预测以及复合物三维结构建模。

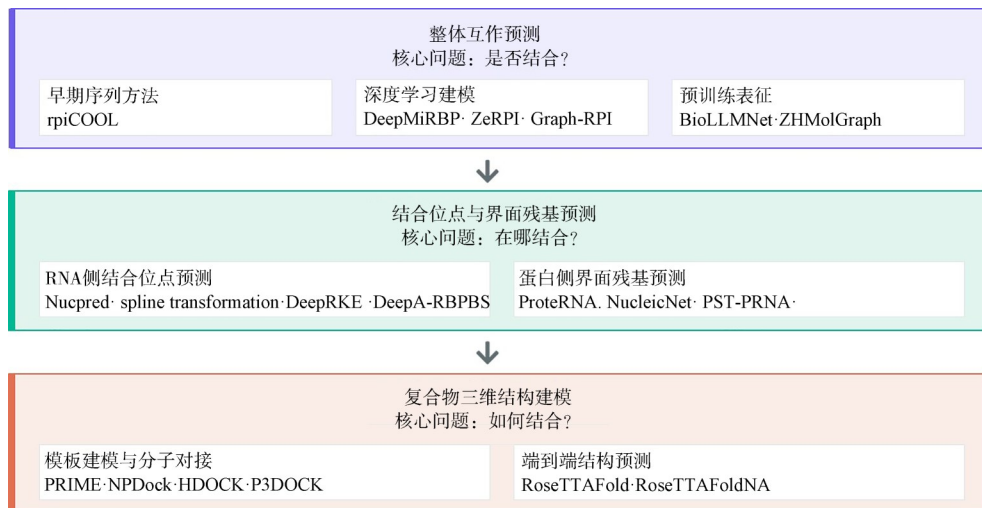


图 3 RNA-蛋白互作预测的三层递进建模框架

Fig. 3 Three-level progressive modeling framework for RNA-protein interaction prediction

2.1 整体互作预测

整体互作预测关注的核心问题是,给定一条 RNA 序列和一条蛋白质序列,它们之间是否发生物理结合。该任务通常被建模为分子对层面的二分类或互作概率预测,是后续结合位点定位与复合物结构建模的基础。

早期方法主要依赖手工设计的序列特征,如碱基或氨基酸组成、k-mer 频率和序列基序等,并结合支持向量机、随机森林等传统分类器完成预测^[55-56],输出给定 RNA-蛋白对的互作标签或互作概率分数。例如,rpiCOOL^[57]从已验证的 RNA-蛋白互作中提取序列基序与核苷酸组成特征,并通过特征选择降低冗余。由于不依赖复合物结构信息,这类方法适用于大规模初筛,但其性能通常受特征设计、负样本构造^[58]和训练集分布制约。

2.1.1 深度学习驱动的互作建模

随着深度学习的发展,整体互作预测逐渐从手工特征转向自动表征学习^[59]。例如,DeepMiRBP^[60]在 RNA 侧引入双向长短期记忆网络(Bi-LSTM)与注意力机制,在蛋白侧利用卷积神经网络(CNN)提取特征,实现了对 microRNA-蛋白互作的有效预测。相比传统方法,这类模型更容易捕获序列上下文、局部基序及复杂的非线性

互作模式。但这类方法通常依赖训练集与测试集分布相近的假设,对未见 RNA 或未见蛋白的泛化能力有限。为此,当前研究通过引入图学习与零样本学习来缓解这一问题。比如,ZeRPI^[61]通过采用图神经网络学习 RNA 和蛋白的联合表示,从而实现其对未见 RNA 结合蛋白(RBP)的零样本预测。Graph-RPI^[62]则通过统一的 RNA-蛋白互作图和自监督学习策略,从而降低对随机负样本的依赖,并提高了跨物种场景下的鲁棒性。整体来看,深度学习一方面可以提升基准数据集上的预测精度,另一方面也在未见 RNA、未见蛋白及跨物种场景中不断增强泛化能力。然而,因为这类方法的表征学习仍主要依赖任务内有限的标注数据,所以其表示能力和迁移潜力仍存在一定限制。

2.1.2 预训练表征

预训练大模型为整体互作预测提供了新的突破口。与依赖任务内标注数据的监督模型不同,预训练方法能够从大规模未标注序列中学习通用表示,从而减少人工特征工程依赖,并提升跨任务迁移能力^[63-66]。

例如,BioLLMNet^[63]仅以序列为输入,即通过在统一框架下处理 RNA-蛋白、RNA-RNA 和 RNA-小分子等多类互作预测任务,从而体现了预

训练表征的通用性。进一步地,ZHMolGraph^[67]将图结构表示与语言模型嵌入结合,在 RNA 和蛋白均未出现在训练集中的零样本条件下,保持稳定的预测性能。

总体而言,整体互作预测的发展路径,已由基于手工特征的传统机器学习方法,演进为深度表征学习,并进一步走向结合预训练模型与零样本泛化的建模范式。这类方法的核心价值在于,当

缺乏实验结构信息时,通过从海量 RNA-蛋白组合中,快速筛出最可能互作的候选对,从而为后续实验验证和机制研究提供优先级排序。为便于横向比较,表 2 总结了代表性方法的模型架构、输入信息、核心思路及泛化能力。然而,由于其主要回答的是“是否结合”,而非“在哪里结合”或“如何结合”,研究也自然进一步延伸到结合位点定位和复合物结构建模。

表 2 RNA-蛋白整体互作预测代表性算法比较

Tab. 2 Comparison of representative algorithms for global RNA-protein interaction prediction

方法	模型架构	输入信息	核心思路	泛化能力
rpiCOOL ^[57]	随机森林	序列信息	基于手工特征进行整体互作预测	有限
Cheng 等 ^[58]	支持向量机/随机森林/朴素贝叶斯	序列信息	通过高质量负样本提升分类效果	有限
DeepMiRBP ^[60]	Bi-LSTM+CNN 的混合深度学习框架	序列信息+结构信息	深度表征学习与迁移学习结合	有限
ZeRPI ^[61]	图神经网络(GNN)	序列信息+结构信息	面向未见蛋白的零样本互作预测	未见 RBP 的零样本预测
Graph-RPI ^[62]	GNN+自监督学习	序列信息	统一图建模并引入自监督学习	跨物种更稳健
BioLLMNet ^[63]	预训练语言模型融合框架	序列信息	统一多任务互作预测框架	跨任务迁移较强
ZHMolGraph ^[67]	图学习与语言模型融合框架	序列信息	图结构与语言模型联合建模	对未见 RNA 和蛋白均具有较强泛化能力

2.2 结合位点与界面残基预测

整体互作预测能够回答 RNA 与蛋白质是否互作,却无法揭示在哪里结合。结合位点与界面残基预测正是为解决这一问题。

2.2.1 RNA 侧结合位点预测

由于 RNA 侧结合位点预测,通常面向复合物三维结构缺乏或难以获得的场景,所以多依赖 RNA 序列及其有限的结构信息,来识别可能被 RNA 结合蛋白(RBP)识别的区域。

早期研究首先关注的是局部序列本身能否支持位点识别。一项系统性比较研究^[68]系统,评估了多种序列特征,与机器学习分类器在 RNA 侧结合位点预测中的表现,发现基于三联体/四联体局部序列模式的随机森林效果较优,并以此开发了 Nucpred 工具。该研究表明,即使缺乏复合物结构信息,RNA 局部序列模式本身,也包含能够区分结合位点的有效线索。

仅依赖局部序列模式,仍不足以解释 RBP 结合位点在转录本上的位置偏好,因为 RBP 的结合不仅取决于局部碱基组成,还与其到转录起始位

点、外显子边界和 polyA 位点等功能位置的相对距离密切相关。为此,spline transformation^[69]通过这些位置信息,作为连续特征融入神经网络模型,从而显著提高体内 RBP 结合位点的预测性能。

在此基础上,研究重点进一步转向序列-结构联合表征与上下文依赖建模。传统手工特征虽然能够提取局部模式,却难以同时融合 RNA 的结构信息,并有效描述远距离序列上下文之间的关联。为此,DeepRKE^[70]首先以词向量嵌入替代传统 one-hot 编码,对 RNA 序列及其二级结构信息予以联合表示,再结合 CNN 和 Bi-LSTM 联合提取局部模式与长程依赖,推动该任务由特征工程转向表征学习。随后,DeepA-RBPBS^[71]进一步使用 CNN,与双向门控循环单元网络(BiGRU)提取序列与结构特征,并引入注意力机制,以增强模型对关键特征的关注,同时也增强其对远距离上下文信息的捕捉能力。在多个数据集的测试中,DeepA-RBPBS 的性能均比同类算法有所提升。

在上述线性 RNA 位点预测框架基础上,相关方法进一步扩展到环状 RNA(circRNA)场景。由

于环状RNA具有闭环拓扑结构,其结合规律不能简单类比线性RNA,因此模型需要针对这一特殊分子类型进行适配。现有研究主要沿两条路径推进,一类通过融合多种序列编码与深度网络,从而提高对复杂序列模式的表征能力^[72-73];另一类通过显式引入碱基配对概率等二级结构特征,从而增强对局部构象的建模^[74]。当前,二者共同推动circRNA结合位点预测性能的提升。

2.2.2 蛋白侧界面残基预测

早期研究主要尝试在缺乏高质量复合物结构的条件下,利用蛋白自身信息,识别可能参与RNA接触的残基。2004年的神经网络方法^[75],已将预测目标定义为参与蛋白-RNA相互作用的残基的识别,并结合蛋白序列与二级结构信息进行建模。这是该方向较早的残基预测工作之一。随后,ProteRNA^[76]进一步将基于进化谱的SVM分类器,与序列保守性分析相结合,来识别与RNA结合的残基,预测准确率获得进一步提升。这类工作说明,即使没有完整复合物结构,序列和进化信息仍可用于缩小候选界面残基范围。但这类方法往往难以充分捕捉RNA结合界面的空间分布特征。

针对以上不足,后续研究开始引入蛋白结构信息,以提升界面残基预测能力。NucleicNet^[24]不再仅仅判断残基是否与RNA接触,而是进一步预测蛋白表面对RNA不同组分的结合偏好,将预测任务由接触识别,扩展到界面相互作用特征刻画。PST-PRNA^[77]则将蛋白三维表面编码为拓扑表示,并利用深度学习挖掘其中的空间模式,从而更有效地表征RNA结合界面的整体几何特征与局部环境。总体来看,蛋白侧界面残基预测的发展经历了由序列与保守性信息驱动,逐步走向结合结构信息进行界面表征的过程。

总体而言,结合位点与界面残基预测,通过将研究视角由“是否结合”推进到“在哪里结合”,从而显著提升其对蛋白-RNA相互作用机制的解析粒度。然而,这类方法仍难以从整体空间层面,解释蛋白质与RNA如何实现稳定结合等问题。因此,后续研究进一步转向复合物三维结构建模。

2.3 复合物三维结构建模

整体互作预测回答“是否结合”,位点预测回答“在哪结合”,而复合物三维结构建模旨在回答“如何结合”,即重建RNA与蛋白质在三维空间中

的相对构象及界面细节。其结果可用于RNA-蛋白互作机制解析、药物设计与蛋白工程。

2.3.1 模板建模与分子对接方法

最开始,RNA-蛋白复合物建模主要依赖两类思路:一类是利用已知相似复合物的结合模式来重建新复合物结构;另一类是以蛋白和RNA的已知或预先获得的单体结构为输入,再通过空间搜索生成二者可能的结合构象。前者通常被称为模板建模,后者则属于分子对接,也有一些方法将二者结合,形成多步骤建模流程。模板建模方面,PRIME^[78]表明,已知复合物中的结合模式在相似分子之间具有一定可迁移性,因此可利用模板复合物重建新的RNA-蛋白复合物结构。分子对接方面,NPDock^[79]构建了面向蛋白-核酸复合物的对接模型,可根据给定的蛋白和RNA结构生成候选复合物构象,并对结果进行排序。HDOCK^[80]和P3DOCK^[81]进一步结合模板信息与对接策略,利用两类方法的互补性,不仅提高了复合物结构预测的准确性,也拓展了复合物建模方法的适用范围。这类方法为复合物结构建模,提供了早期可行路径,但其性能往往较依赖已知模板的质量,以及输入蛋白和RNA单体结构的合理性。

2.3.2 端到端复合物结构预测

随着深度学习在结构预测中的发展,复合物建模开始由依赖单体结构输入的分步流程,逐步转向统一框架下的直接预测。RoseTTAFold^[23]采用三轨神经网络架构,通过在一维序列特征、二维成对关系和三维结构表示之间,进行联合迭代更新,从而以端到端方式预测蛋白质及其复合体结构。如图4所示,RoseTTAFoldNA^[25]在此基础上进一步纳入核酸信息,通过将统一建模框架扩展到蛋白-核酸复合物,从而直接从蛋白和核酸序列出发,输出复合物的三维结构及其置信度估计。相比传统的多阶段流程,这类方法推动复合物建模,通过由分步组合走向统一预测,从而减少了误差在单体建模、对接与重排序各环节之间的逐步累积,并增强了序列信息、分子间关系与三维结构特征的联合建模能力。

3 AI在RNA-小分子互作的应用

RNA-小分子相互作用建模的核心问题是给

定 RNA 和一个小分子,二者之间是否能够发生特异性结合。如图 5 所示,该任务通常包括 RNA 结合口袋与结合位点识别、候选小分子筛选、RNA-小分子结合模式预测与亲和力评估等子任务,是理解 RNA 分子识别机制并支持 RNA 靶向药物设计的重要基础。相较于蛋白质-小分子相互作用, RNA-小分子相互作用更难预测,主要原因在于

RNA 构象柔性更高、识别过程更易受金属离子与水分子等环境因素影响,且高分辨率复合物结构数据仍然有限^[30]。受此制约,传统分子对接方法,在 RNA 体系中,往往难以稳定刻画真实结合模式,因此 AI 逐渐被引入这一领域,用于提升各类子任务的建模效率与预测性能。

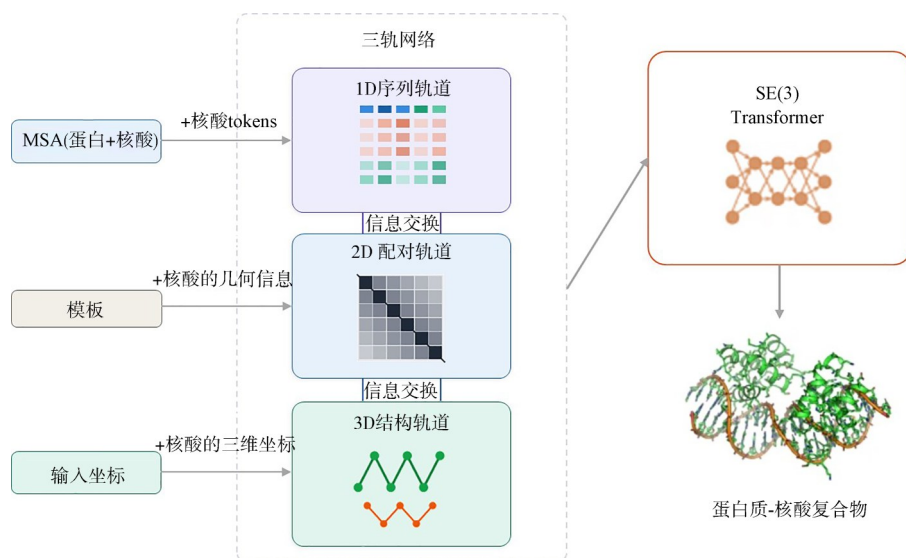


图 4 RoseTTAFoldNA 架构图

Fig. 4 Architecture of RoseTTAFoldNA

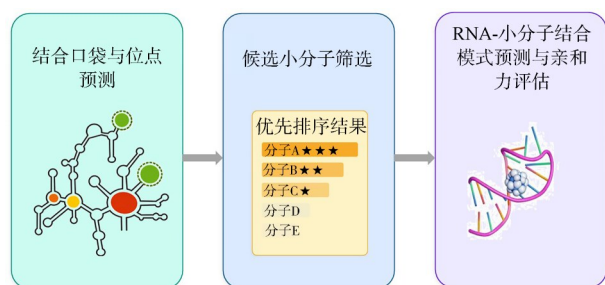


图 5 RNA-小分子相互作用建模研究框架

Fig. 5 Research framework for RNA-small molecule interaction modeling

3.1 RNA 可结合口袋与结合位点预测

在 RNA-小分子相互作用建模中,首先需要回答 RNA 上哪些区域可能结合小分子。相关研究往往先从 RNA 三维结构表面的结合口袋识别出发,进一步发展到核苷酸级别的结合位点预测。

早期的相关工作^[82]多从 RNA 表面的几何空腔出发识别候选口袋,再进一步判断这些区域是否具有配体可结合性。后续结合位点预测算法主要围绕如何更充分地表征 RNA 的多层次信息展开。MultiModRLBP^[83]较早体现了多模态联合建

模思路,将三维结构属性、关系图以及 RNA 语义信息共同纳入模型,以增强模型对微小结构差异的敏感性,从而提高结合位点预测的准确性。RL-site^[84]进一步引入预训练 RNA 语言模型,并结合图注意力网络融合序列上下文与结构信息,在公开测试集上的结果显著优于其他算法,表明预训练序列表征有助于提升结合位点预测性能。而 MVRBind^[85]则采用多视图卷积网络,从 RNA 的一级、二级和三级结构视角分别构图并进行特征融合,以获得更全面的核苷酸表征。另外,实验证明此算法在多构象条件下具有较好的预测鲁棒性。总体来看,这一阶段算法的关键,不在于单纯增加模型复杂度,而在于更有效地联合利用 RNA 的序列与结构信息,以提高结合位点预测的准确性与稳定性。

3.2 候选小分子的筛选

在 RNA 上的潜在结合位点被识别之后,下一步需要回答的是面对大量候选化合物,哪些小分子更有可能与 RNA 结合,即“哪些小分子更值得优先筛选”。

例如, RNAmigos^[86]将RNA位点的局部结构表示为增强碱基配对网络,并通过图神经网络学习该结构表示与小分子特征之间的对应关系,从而预测新位点可能偏好的小分子。该工作表明, RNA局部结构图中已经包含了,可用于小分子识别与初步筛选的重要信息。在此基础上, RNAmigos2^[87]进一步将这一思路,推进到基于结构的虚拟筛选任务中。该方法结合粗粒度三维建模、合成数据增强与RNA特异自监督学习,直接依据RNA口袋的结构表示对候选分子库进行快速排序,而非对每个候选分子逐一进行传统分子对接。该方法在显著降低计算成本的同时,取得了具有竞争力的筛选表现。

除此之外,近年的 RNAsmol^[88]则从更一般的RNA-小分子相互作用预测角度出发,通过联合建模RNA表征与小分子图特征,对RNA-小分子对进行相互作用打分,并进一步支持大规模候选分子的初步筛选。RNAsmol不需要结构输入即可产生可靠预测,并可适配多种药物设计场景。

3.3 RNA-小分子结合模式预测与亲和力评估

在候选小分子被筛选之后,研究进一步转向RNA-小分子结合模式预测,与亲和力评估这两个更具体的问题。

3.3.1 RNA-小分子结合模式预测

在结合模式预测中,传统分子对接通常会为同一RNA-小分子体系生成大量候选姿态(pose),但由于RNA骨架柔性高、局部化学环境复杂,经验打分函数往往难以稳定识别接近天然状态的正确姿态^[89]。因此,后续研究开始在对接后,利用AI对候选姿态,进行重新打分,以更有效地筛选出正确姿态。

这一方向的关键不在于是否生成候选姿态,而在于如何从这些候选结果中,识别更接近真实结合模式的正确姿态。RNAPosers^[89]的思路较为直接,它先把每个RNA-小分子候选姿态表示为一种“姿态指纹”,其中每个配体原子周围的局部RNA环境被编码为原子级特征,再组合成整个复合物的特征表示,最后利用机器学习分类器判断该姿态的天然性,即其是否更接近真实结合构象。SPRank^[90]则采用另一种思路。它不是单纯依赖一个经验打分函数,而是把知识库打分与机器学习重打分结合起来,并配合第三方对接软件对柔性配体和RNA构象集合进行采样。与仅依赖单一经验打分函数的做法相比,这一方法不再只对单

个候选构象进行静态评分,而是利用多构象采样结果,和统计学习模型对候选姿态进行重新排序,因此更适合用于RNA-小分子对接后的姿态判别与重打分。

除了提升姿态判别与重打分性能外,近年来也有研究开始关注模型的可解释性。例如, Szulc等^[91]将RNA-小分子复合物中的氢键、 π - π 堆叠等非共价相互作用,编码为结构互作指纹(SIFt),并结合SHAP、LIME等多种可解释AI方法分析模型决策依据。此类工作的重点并不在于提出新的结合模式预测算法,而在于解释模型为什么会作出这样的结合预测。

3.3.2 亲和力评估

在获得较可信的结合模式之后,研究焦点进一步转向定量问题,即一对RNA-小分子的结合强度如何。早期相关工作多依赖人工设计的特征和传统机器学习回归模型,通过RNA的序列组成、柔性以及小分子的结构与极性等信息预测结合亲和力^[92]。随后,这一方向逐步转向直接利用复合物三维结构进行建模。RLaffinity^[93]将RNA口袋与小分子表示为三维结构输入,利用3D-CNN预测结合亲和力,并结合对比学习驱动的自监督预训练,增强表征能力。实验结果表明,因为该模型在多项评价指标上均优于基线方法,所以三维卷积网络能够通过同时捕获RNA口袋的整体空间环境,与局部邻域核苷酸信息,从而提升亲和力预测性能。而如图6所示,DeepRSMA^[94]则从序列和图结构两个视角,联合表征RNA与小分子,并通过基于Transformer的cross-fusion模块,在核苷酸-原子层面建模二者的细粒度交互,从而提升亲和力预测性能。总体而言,这一阶段的重要变化,并不仅仅在于模型复杂度的不断提高,而在于建模对象,逐步从人工特征转向三维结构及细粒度相互作用信息。

总之, RNA-小分子相互作用建模,涵盖结合口袋与位点识别、候选分子筛选、结合模式预测与亲和力评估等关键任务,是RNA三维结构预测之后的重要下游问题。近年来,该方向的核心进展,主要体现在AI建模能力的提升上,即从依赖人工特征和经验规则的浅层方法,逐步转向联合利用序列、结构及相互作用信息的系统性建模框架。

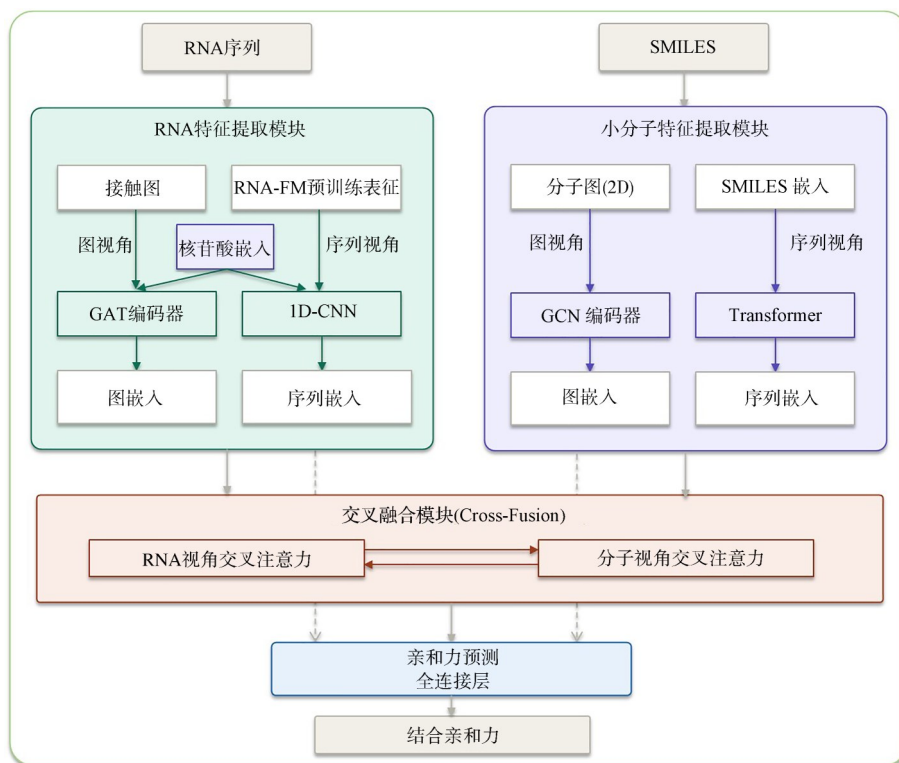


图 6 DeepRSMA 模型架构

Fig. 6 Overview of the DeepRSMA framework

4 挑战与展望

本文系统梳理了人工智能在 RNA 三维结构预测、RNA-蛋白互作以及 RNA-小分子互作等 3 个方向的研究进展。总体来看,相关研究已逐步由单一任务驱动建模,走向结构-互作-功能一体化建模。在 RNA 三维结构预测方面,算法经历了从基于模板调用与片段组装的传统方法,到借助 AI 预测关键结构,约束信息的非端到端方法,再到直接输出三维坐标的端到端方法的演进过程。预训练语言模型的引入,进一步缓解了 RNA 结构数据有限带来的制约,它们能通过在大规模序列数据上进行自监督学习,从而为结构预测提供了更丰富的特征表示。

在 RNA-蛋白互作建模方面,研究沿着“是否结合-在哪结合-如何结合”的思路不断深入,从基于手工特征的整体互作预测,逐步发展到面向复合物结构的端到端预测。同时,预训练表征与零样本学习的引入,显著提升了模型的性能。

在 RNA-小分子互作方面,AI 方法已被运用到结合口袋识别、候选分子筛选、结合模式预测及亲和力评估这三类建模任务,建模对象也由早期的人工特征,逐步转向三维结构及细粒度相互作用

信息。总体而言,AI 正在成为连接 RNA 序列、结构、分子识别与功能解析的重要技术纽带。

尽管上述研究已取得显著进展,该领域仍面临着诸多挑战。首先,高质量数据匮乏仍是制约模型性能提升的核心问题。相较于蛋白质体系, RNA 三维结构及 RNA-小分子复合物的实验解析数据仍较为有限。同时数据噪声、标注不一致及样本分布偏倚等问题,也进一步削弱了模型的泛化能力。其次,现有模型在处理长链 RNA、多链 RNA、环状 RNA 以及高柔性 RNA-小分子体系时,预测精度仍有待提高。并且, AI 模型普遍存在可解释性不足的问题,难以充分揭示预测结果背后的分子识别机制。为此,如何协调数据驱动建模,与物理机制约束之间的关系,仍是未来研究需要突破的重要方向。

针对上述问题,未来的研究方向可能以下几个:1) 生成式模型在 RNA 靶向小分子设计中的应用前景广阔。以扩散模型、变分自编码器等为代表的方法,可以在给定 RNA 的条件下,直接生成具有潜在结合能力的候选分子,从而推动研究由筛选分子,向设计分子转变。2) 多模态与多任务联合建模,可以在统一框架下同时处理 RNA 结构预测,及 RNA 与蛋白质、小分子的互作建模,实

现从序列、结构到复合物结构的协同推理。这方面的主要难点在于,不同任务的数据规模和标注粒度差异较大,统一模型容易出现训练不稳定。未来可通过共享表征、任务自适应模块和迁移学习,提升模型在小样本任务中的泛化能力。

3) RNA基础大模型的持续发展,将为下游任务提供更强的通用表征,结合自监督学习与迁移学习,可以进一步缓解标注数据不足的问题。值得注意的是,现有RNA语言模型多以序列为主要输入,对三维结构和相互作用信息的刻画仍然有限。

4) 后续可引入结构感知预训练目标,并在大规模序列预训练基础上融合结构与互作数据,以增强模型对RNA空间构象和分子识别的表征能力。

5) 可解释人工智能方法的引入,将有助于揭示模型决策与RNA分子识别机制之间的内在联系,增强模型输出的可信度,并为后续的实验验证与药物设计,提供更具现实意义的指导。随着人工智能方法与RNA生物学研究的结合不断加深, RNA结构解析、功能注释及RNA靶向药物发现等方向将获得进一步的发展。

参考文献:

- [1] Mandal M, Boese B, Barrick J E, *et al.* Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria [J]. *Cell*, 2003, 113(5): 577-586.
- [2] Wang Y, Ludwig J, Schuberth C, *et al.* Structural and functional insights into 5'-ppp RNA pattern recognition by the innate immune receptor RIG-I [J]. *Nat Struct Mol Biol*, 2010, 17(7): 781-787.
- [3] Ou X, Zhang Y, Xiong Y, *et al.* Advances in RNA 3D structure prediction [J]. *J Chem Inf Model*, 2022, 62(23): 5862-5874.
- [4] Hentze M W, Sommerkamp P, Ravi V, *et al.* Rethinking RNA-binding proteins: Riboregulation challenges prevailing views [J]. *Cell*, 2025, 188(18): 4811-4827.
- [5] Miao W, Porter D F, Lopez-Pajares V, *et al.* Regulation of RNA-binding proteins by small biomolecules [J]. *Nat Rev Mol Cell Biol*, 2026, 27(3): 213-233.
- [6] Mandal M, Breaker R R. Gene regulation by riboswitches [J]. *Nat Rev Mol Cell Biol*, 2004, 5(6): 451-463.
- [7] Serganov A, Nudler E. A decade of riboswitches [J]. *Cell*, 2013, 152(1/2): 17-24.
- [8] Mukherjee N, Wessels H-H, Lebedeva S, *et al.* Deciphering human ribonucleoprotein regulatory networks [J]. *Nucleic Acids Res*, 2019, 47(2): 570-581.
- [9] Yoneyama M, Kato H, Fujita T. Physiological functions of RIG-I-like receptors [J]. *Immunity*, 2024, 57(4): 731-751.
- [10] Scotti M M, Swanson M S. RNA mis-splicing in disease [J]. *Nat Rev Genet*, 2016, 17(1): 19-32.
- [11] Nemeth K, Bayraktar R, Ferracin M, *et al.* Non-coding RNAs in disease: From mechanisms to therapeutics [J]. *Nat Rev Genet*, 2024, 25(3): 211-232.
- [12] Gebauer F, Schwarzl T, Valcárcel J, *et al.* RNA-binding proteins in human genetic disease [J]. *Nat Rev Genet*, 2021, 22(3): 185-198.
- [13] Yu A M, Choi Y H, Tu M J. RNA drugs and RNA targets for small molecules: principles, progress, and challenges [J]. *Pharmacol Rev*, 2020, 72(4): 862-898.
- [14] Amiri A, Barreto G, Sathyapalan T, *et al.* siRNA therapeutics: Future promise for neurodegenerative diseases [J]. *Curr Neuropharmacol*, 2021, 19(11): 1896-1911.
- [15] Chawra H S, Agarwal M, Mishra A, *et al.* microRNA-21's role in PTEN suppression and PI3K/AKT activation: Implications for cancer biology [J]. *Pathol Res Pract*, 2024, 254: 155091.
- [16] Kovachka S, Panosetti M, Grimaldi B, *et al.* Small molecule approaches to targeting RNA [J]. *Nat Rev Chem*, 2024, 8(2): 120-135.
- [17] Zhu Y, Zhu L, Wang X, *et al.* RNA-based therapeutics: an overview and prospectus [J]. *Cell Death Dis*, 2022, 13(7): 644.
- [18] Childs-Disney J L, Yang X, Gibaut Q M R, *et al.* Targeting RNA structures with small molecules [J]. *Nat Rev Drug Discov*, 2022, 21(10): 736-762.
- [19] Lan W, Dong Y, Chen Q, *et al.* KGANCD: Predicting circRNA-disease associations based on knowledge graph attention network [J]. *Brief Bioinform*, 2022, 23: bbab494.
- [20] Lan W, Li C, Chen Q, *et al.* LGCD: Predicting CircRNA-disease association based on fusion of local and global features [J]. *IEEE/ACM Trans Comput Biol And Bioinf*, 2024, 21(5): 1413-1422.
- [21] Lan W, Dong Y, Chen Q, *et al.* IGNSCD: Predicting CircRNA-disease associations based on improved graph convolutional network and negative sampling [J]. *IEEE/ACM Trans Comput Biol And*

- Bioinf, 2022, 19(6): 3530-3538.
- [22] 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 [J]. Nucleic Acids Res, 2022, 50(D1): D439-D444.
- [23] Minkyung B, Frank D, Ivan A, *et al.* Accurate prediction of protein structures and interactions using a three-track neural network [J]. Science, 2021, 373(6557): 871-876.
- [24] Lam J H, Li Y, Zhu L, *et al.* A deep learning framework to predict binding preference of RNA constituents on protein surface [J]. Nat Commun, 2019, 10: 4941.
- [25] Baek M, McHugh R, Anishchenko I, *et al.* Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA [J]. Nat Methods, 2024, 21(1): 117-121.
- [26] Abramson J, Adler J, Dunger J, *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3 [J]. Nature, 2024, 630(8016): 493-500.
- [27] Yu H, Qi Y, Ding Y. Deep learning in RNA structure studies [J]. Front Mol Biosci, 2022, 9: 869601.
- [28] Liu X, Duan Y, Hong X, *et al.* Challenges in structural modeling of RNA-protein interactions [J]. Curr Opin Struct Biol, 2023, 81: 102623.
- [29] Basu S, Yu J, Kihara D, *et al.* Twenty years of advances in prediction of nucleic acid-binding residues in protein sequences [J]. Brief Bioinform, 2024, 26: bbaf016.
- [30] Cai Z, Ma H, Ye F, *et al.* Discovery of RNA-targeting small molecules: Challenges and future directions [J]. MedComm, 2025, 6(9): e70342.
- [31] Popenda M, Szachniuk M, Antczak M, *et al.* Automated 3D structure composition for large RNAs [J]. Nucleic Acids Res, 2012, 40(14): e112.
- [32] Xu X, Zhao P, Chen S J. Vfold: A web server for RNA structure and folding thermodynamics prediction [J]. PLoS One, 2014, 9(9): e107504.
- [33] Cheng Y, Zhang S, Xu X, *et al.* Vfold2D-MC: A physics-based hybrid model for predicting RNA secondary structure folding [J]. J Phys Chem B, 2021, 125(36): 10108-10118.
- [34] Wang J, Wang J, Huang Y, *et al.* 3dRNA v2.0: An updated web server for RNA 3D structure prediction [J]. Int J Mol Sci, 2019, 20(17): 4116.
- [35] Nithin C, Kmiecik S, Błaszczyk R, *et al.* Comparative analysis of RNA 3D structure prediction methods: Towards enhanced modeling of RNA-ligand interactions [J]. Nucleic Acids Res, 2024, 52(13): 7465-7486.
- [36] Zhang T, Hu G, Yang Y, *et al.* All-atom knowledge-based potential for RNA structure discrimination based on the distance-scaled finite ideal-gas reference state [J]. J Comput Biol, 2020, 27(6): 856-867.
- [37] Tan Y L, Wang X, Shi Y Z, *et al.* rsRNASP: a residue-separation-based statistical potential for RNA 3D structure evaluation [J]. Biophys J, 2022, 121(1): 142-156.
- [38] Zhao Y, Wang J, Zeng C, *et al.* Evaluation of RNA secondary structure prediction for both base-pairing and topology [J]. Biophys Rep, 2018, 4(3): 123-132.
- [39] Singh J, Hanson J, Paliwal K, *et al.* RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning [J]. Nat Commun, 2019, 10: 5407.
- [40] Singh J, Paliwal K, Litfin T, *et al.* Predicting RNA distance-based contact maps by integrated deep learning on physics-inferred secondary structure and evolutionary-derived mutational coupling [J]. Bioinformatics, 2022, 38(16): 3900-3910.
- [41] Mukherjee S, Moafinejad S N, Badeppally N G, *et al.* Advances in the field of RNA 3D structure prediction and modeling, with purely theoretical approaches, and with the use of experimental data [J]. Structure, 2024, 32(11): 1860-1876.
- [42] Sha C M, Wang J, Dokholyan N V. Predicting 3D RNA structure from the nucleotide sequence using Euclidean neural networks [J]. Biophys J, 2024, 123(17): 2671-2681.
- [43] Wang W, Feng C, Han R, *et al.* trRosettaRNA: Automated prediction of RNA 3D structure with transformer network [J]. Nat Commun, 2023, 14: 7266.
- [44] Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with AlphaFold [J]. Nature, 2021, 596(7873): 583-589.
- [45] Bernard C, Postic G, Ghannay S, *et al.* State-of-the-RNArt: Benchmarking current methods for RNA 3D structure prediction [J]. NAR Genom Bioinform, 2024, 6(2): lqae048.
- [46] Kagaya Y, Zhang Z, Ibtehaz N, *et al.* NuFold: End-to-end approach for RNA tertiary structure prediction

- with flexible nucleobase center representation [J]. Nat Commun, 2025, 16: 881.
- [47] Li Y, Zhang C, Feng C, *et al.* Integrating end-to-end learning with deep geometrical potentials for *ab initio* RNA structure prediction [J]. Nat Commun, 2023, 14: 5745.
- [48] Shen T, Hu Z, Sun S, *et al.* Accurate RNA 3D structure prediction using a language model-based deep learning approach [J]. Nat Methods, 2024, 21(12): 2287-2298.
- [49] Magnus M, Antczak M, Zok T, *et al.* RNA-Puzzles toolkit: A computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools [J]. Nucleic Acids Res, 2020, 48(2): 576-588.
- [50] Wang N, Bian J, Li Y, *et al.* Multi-purpose RNA language modelling with motif-aware pretraining and type-guided fine-tuning [J]. Nat Mach Intell, 2024, 6(5): 548-557.
- [51] Penić R J, Vlašić T, Huber R G, *et al.* RiNALMo: general-purpose RNA language models can generalize well on structure prediction tasks [J]. Nat Commun, 2025, 16: 5671.
- [52] Das R, Kretsch R C, Simpkin A J, *et al.* Assessment of three-dimensional RNA structure prediction in CASP15 [J]. Proteins Struct Funct Bioinform, 2023, 91(12): 1747-1770.
- [53] Kretsch R C, Hummer A M, He S, *et al.* Assessment of nucleic acid structure prediction in CASP16 [J]. Proteins: Structure, Function, and Bioinformatics, 2026, 94(1): 192-217.
- [54] Krautwurst S, Lamkiewicz K. RNA-protein interaction prediction without high-throughput data: An overview and benchmark of *in silico* tools [J]. Comput Struct Biotechnol J, 2024, 23: 4036-4046.
- [55] Muppirala U K, Honavar V G, Dobbs D. Predicting RNA-protein interactions using only sequence information [J]. BMC Bioinform, 2011, 12: 489.
- [56] Lu Q, Ren S, Lu M, *et al.* Computational prediction of associations between long non-coding RNAs and proteins [J]. BMC Genom, 2013, 14: 651.
- [57] Akbaripour-Elahabad M, Zahiri J, Rafeh R, *et al.* rpiCOOL: A tool for *in Silico* RNA-protein interaction detection using random forest [J]. J Theor Biol, 2016, 402: 1-8.
- [58] Cheng Z, Huang K, Wang Y, *et al.* Selecting high-quality negative samples for effectively predicting protein-RNA interactions [J]. BMC Syst Biol, 2017, 11(S2): 9.
- [59] Alipanahi B, Delong A, Weirauch M T, *et al.* Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning [J]. Nat Biotechnol, 2015, 33(8): 831-838.
- [60] Azizian S, Cui J. DeepMiRBP: A hybrid model for predicting microRNA-protein interactions based on transfer learning and cosine similarity [J]. BMC Bioinform, 2024, 25: 381.
- [61] Gao Y, Shi R, Yu G, *et al.* ZeRPI: A graph neural network model for zero-shot prediction of RNA-protein interactions [J]. Methods, 2025, 235: 45-52.
- [62] Guan J, Yao L, Xie P, *et al.* Graph-RPI: Predicting RNA-protein interactions *via* graph autoencoder and self-supervised learning strategies [J]. Brief Bioinform, 2025, 26(3): bbaf292.
- [63] Abir A R, Toki Tahmid M, Bayzid M S. BioLLM-Net: Enhancing RNA-interaction prediction with a specialized cross-LLM transformation network [J]. Brief Bioinform, 2025, 26(5): bbaf549.
- [64] Gan L, Wang X, Zhou Y, *et al.* Protein-binding RNA prediction based on integrated sequence-structure-function pre-training [J]. IEEE Trans Comput Biol Bioinform, 2025, 22(4): 1441-1452.
- [65] Li G, Jiang F, Zhu J, *et al.* HydraRNA: A hybrid architecture based full-length RNA language model [J]. Genome Biol, 2025, 26: 383.
- [66] Lan W, Tang Z, Liu M, *et al.* The large language models on biomedical data analysis: A survey [J]. IEEE J Biomed Health Inform, 2025, 29(6): 4486-4497.
- [67] Liu H, Jian Y, Zeng C, *et al.* RNA-protein interaction prediction using network-guided deep learning [J]. Commun Biol, 2025, 8: 247.
- [68] Agarwal A, Singh K, Kant S, *et al.* A comparative analysis of machine learning classifiers for predicting protein-binding nucleotides in RNA sequences [J]. Comput Struct Biotechnol J, 2022, 20: 3195-3207.
- [69] Avsec Ž, Barekatin M, Cheng J, *et al.* Modeling positional effects of regulatory sequences with spline transformations increases prediction accuracy of deep neural networks [J]. Bioinformatics, 2018, 34(8): 1261-1269.
- [70] Deng L, Liu Y, Shi Y, *et al.* Deep neural networks for inferring binding sites of RNA-binding proteins by using distributed representations of RNA primary sequence and secondary structure [J]. BMC Genomics, 2020, 21(Suppl. 13): 866.
- [71] Du Z, Xiao X, Uversky V N. DeepA-RBPBS: A

- hybrid convolution and recurrent neural network combined with attention mechanism for predicting RBP binding site[J]. *J Biomol Struct Dyn*, 2022, 40(9): 4250-4258.
- [72] Jia C, Bi Y, Chen J, *et al.* PASSION: An ensemble neural network approach for identifying the binding sites of RBPs on circRNAs [J]. *Bioinformatics*, 2020, 36(15): 4276-4282.
- [73] Lasantha D, Vidanagamachchi S, Nallaperuma S. CRIECNN: Ensemble convolutional neural network and advanced feature extraction methods for the precise forecasting of circRNA-RBP binding sites [J]. *Comput Biol Med*, 2024, 174: 108466.
- [74] Cao C, Wang C, Dai Q, *et al.* CRBPSA: CircRNA-RBP interaction sites identification using sequence structural attention model [J]. *BMC Biol*, 2024, 22: 260.
- [75] Jeong E, Chung I-F, Miyano S. A neural network method for identification of RNA-interacting residues in protein [J]. *Genome Inform*, 15(1): 105-116.
- [76] Huang Y F, Chiu L-Y, Huang C-C, *et al.* Predicting RNA-binding residues from evolutionary information and sequence conservation [J]. *BMC Genom*, 2010, 11(Suppl. 4): S2.
- [77] Li P, Liu Z P. PST-PRNA: Prediction of RNA-binding sites using protein surface topography and deep learning [J]. *Bioinformatics*, 2022, 38(8): 2162-2168.
- [78] Zheng J, Kundrotas P J, Vakser I A, *et al.* Template-based modeling of protein-RNA interactions [J]. *PLoS Comput Biol*, 2016, 12(9): e1005120.
- [79] Tuszyńska I, Magnus M, Jonak K, *et al.* NPDock: A web server for protein-nucleic acid docking [J]. *Nucleic Acids Res*, 2015, 43(W1): W425-W430.
- [80] Yan Y, Zhang D, Zhou P, *et al.* HDock: A web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy [J]. *Nucleic Acids Res*, 2017, 45(W1): W365-W373.
- [81] Zheng J, Hong X, Xie J, *et al.* P3DOCK: A protein-RNA docking webserver based on template-based and template-free docking [J]. *Bioinformatics*, 2020, 36(1): 96-103.
- [82] Xie J, Frank A T. Mining for ligandable cavities in RNA [J]. *ACS Med Chem Lett*, 2021, 12(6): 928-934.
- [83] Wang J, Quan L, Jin Z, *et al.* MultiModRLBP: A deep learning approach for multi-modal RNA-small molecule ligand binding sites prediction [J]. *IEEE J Biomed Health Inform*, 2024, 28(8): 4995-5006.
- [84] Sun S, Yang J, Gao L, *et al.* RNA language model and graph attention network for RNA and small molecule binding sites prediction [J]. *Bioinformatics*, 2025, 41(9): btaf447.
- [85] Chen S, Huang Z, Wang Y, *et al.* MVRBind: multi-view learning for RNA-small molecule binding site prediction [J]. *Brief Bioinform*, 2025, 26(5): bbaf489.
- [86] Oliver C, Mallet V, Gendron R S, *et al.* Augmented base pairing networks encode RNA-small molecule binding preferences [J]. *Nucleic Acids Res*, 2020, 48(14): 7690-7699.
- [87] Carvajal-Patino J G, Mallet V, Becerra D, *et al.* RNAmigos2: Accelerated structure-based RNA virtual screening with deep graph learning [J]. *Nat Commun*, 2025, 16: 2799.
- [88] Ma H, Gao L, Jin Y, *et al.* RNA-ligand interaction scoring *via* data perturbation and augmentation modeling [J]. *Nat Comput Sci*, 2025, 5(8): 648-660.
- [89] Chhabra S, Xie J, Frank A T. RNAPosers: Machine learning classifiers for ribonucleic acid-ligand poses [J]. *J Phys Chem B*, 2020, 124(22): 4436-4445.
- [90] Zhou Y, Jiang Y, Chen S J. SPRank—A knowledge-based scoring function for RNA-ligand pose prediction and virtual screening [J]. *J Chem Theory Comput*, 2024(12): 4c00681.
- [91] Szulc N A, Mackiewicz Z, Bujnicki J M, *et al.* Structural interaction fingerprints and machine learning for predicting and explaining binding of small molecule ligands to RNA [J]. *Brief Bioinform*, 2023, 24(4): bbad187.
- [92] Krishnan S R, Roy A, Gromiha M M. Reliable method for predicting the binding affinity of RNA-small molecule interactions using machine learning [J]. *Brief Bioinform*, 2024, 25(2): bbae002.
- [93] Sun S, Gao L. Contrastive pre-training and 3D convolution neural network for RNA and small molecule binding affinity prediction [J]. *Bioinformatics*, 2024, 40(4): btae155.
- [94] Huang Z, Wang Y, Chen S, *et al.* DeepRSMA: A cross-fusion-based deep learning method for RNA-small molecule binding affinity prediction [J]. *Bioinformatics*, 2024, 40(12): btae678.

(责任编辑: 伍少梅)