

植物体内 RNA 二级结构探测方法的研究进展

周晞雯 成柯 朱鸿亮

(中国农业大学食品科学与营养工程学院, 北京 100083)

摘要: 生物体内的 RNA 因氢键等作用配对折叠形成复杂的二级结构及三级结构, 并根据环境产生动态变化, 由此行使功能。植物中的基因表达调控及配体感应都与 RNA 的结构相关, RNA 结构也被认为是转录后调控的一项影响因素, RNA 的结构对与之互作的蛋白的结构和功能产生影响, 因此研究 RNA 功能过程中解析 RNA 结构尤为关键。目前, RNA 二级结构的探测实验方法从 X-射线经过酶切及化学探针标记, 发展至结合高通量测序即可建构模型。此外, 机器学习等方法也为 RNA 结构的探测提供了新的思路。本文综述了影响体内 RNA 结构的因素及近年来用于植物体内 RNA 结构预测的方法, 讨论了现有方法的应用限制下对 RNA 二级结构预测新方法开发的意义及存在的问题, 展望计算机预测方法与实验方法未来的发展趋势, 旨在为后续 RNA 结构相关研究提供方法参考。

关键词: RNA 结构; 二级结构; 探测方法; 植物生理

DOI:10.13560/j.cnki.biotech.bull.1985.2022-0518

Research Progress in the Approaches to *in vivo* RNA Secondary Structure Profiling in Plants

ZHOU Xi-wen CHENG Ke ZHU Hong-liang

(College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083)

Abstract: RNAs fold into complex secondary and tertiary structures via hydrogen bond pairing to enable their diverse functions in cells after dynamic changes according to environment. RNA structure is considered to be related to the regulation of gene expression and ligand sensing in plants, and is also considered to be an important factor in post-transcriptional regulation. Structured RNA can also affect the dimensional structure of RNA binding proteins and their functions. Thus, the analysis of RNA structure is particularly critical in the study of RNA function. At present, various detection methods for RNA secondary structure (RSS) has been developed from X-ray to enzyme digestion and chemical probe labeling, which can be combined with high-throughput sequencing to construct models. In addition, machine learning and other methods also provide new ideas for the detection of RNA structure. In this review, we summarize the factors affecting the RNA structure *in vivo* and recent reported methods for probing RNA structure *in vivo*, discuss the significances of developing novel prediction methods for RSS under the limitations of current methods as well as the issues, and prospect the future development trend of computer prediction methods and experimental methods, aiming to provide a methodological reference for subsequent research on RNA structure.

Key words: RNA structure; secondary structure; detection methods; plant physiology

生物体内的 RNA 是携带遗传信息的遗传物质, 具有催化及调控基本生物过程的功能。与 DNA 不同, ssRNA (single-strand RNA) 和 dsRNA (double-strand

RNA) 能够自发地进行碱基互补配对, 在结合或不结合蛋白质的情况下形成结构。生物体内的 RNA 也因环境差异而展示出与体外 RNA 结构上的差异, 因

收稿日期: 2022-04-27

基金项目: 国家自然科学基金项目 (31972472)

作者简介: 周晞雯, 女, 硕士研究生, 研究方向: 果蔬采后生理与分子生物学; E-mail: 18801059213@163.com

通讯作者: 朱鸿亮, 男, 博士, 教授, 研究方向: 果蔬采后生理与分子生物学; E-mail: hlzhu@cau.edu.cn

此, RNA 的体内结构解析是对有 RNA 结构参与的生命活动进行探究的必经之路。解析 RNA 二级结构(RNA secondary structure, RSS) 面临两个重要问题: 其一是体内、外 RNA 结构需要不同的探测方法; 其二是 RNA 结构在体内参与的生物过程中产生的动态变化。RNA 结构的研究方法经过数十年的研究, 从早期的 X 光和核磁共振技术发展至冷冻电镜, 到酶切标记与化学探针标记后利用反转录中引入的截断和错配搭配测序技术获得结构, 以及利用机器学习、深度学习根据碱基等进行预测, 逐渐趋向于安全化和精确化。现有的技术已经能够探知生物体内的 RNA 二级结构以及三级结构, 但是对于体内动态变化着的 RNA 仅能够提供较为稳定的构象, 如何更加精确地展示体内 RNA 结构的全貌以及关键性的生理调控中 RNA 的精确结构是接下来的研究重难点。

1 植物体内的 RNA 二级结构

RNA 的二级结构是以核酸序列为根据, 通过碱基互补配对以及分子间氢键形成的。生物体内的 RNA 存在形式包括双链(dsRNA) 和单链(ssRNA), 目前已广泛了解的结构包括发夹(hair-pins)、茎环(stem-loop)、假结(pseudoknot)、三链(triplex)、G- 四联体(G-quadruplexes) 等多种结构基序。这些结构是由不相邻的核苷酸相互作用形成, 包括在同一 RNA 分子内的顺式相互作用和需要其他生物分子(如 RNA、DNA 和蛋白质) 参与的分子间相互作用^[1]。

RNA 的类型和功能多样, 包括承担生命遗传密码的 mRNA, 编码转录组成部分的 tRNA 和 rRNA、snoRNA、siRNA、miRNA 以及 lncRNA 等。其中 rRNA 因参与翻译过程, 在蛋白质相互作用和生化功能区域是高度保守的, 但在其他区域是十分多变和快速进化的。tRNA 更是最早被阐明其保持“三叶草结构”是为了促进密码子-反密码子相互作用而相对稳定^[2]。snoRNA 长度较短, 是高度结构化的非编码 RNA, 这种 RNA 自身形成较为稳定的二级结构, 还会与其目标 RNA 配对形成分子间结构。siRNA 和 miRNA 都是真核 RNA 沉默机制的组成部分, 这些小 RNA 可以被用来直接沉默转录本和抑制翻译^[3]。在植物中, 这些小 RNA 同样通过 RNA 定向的 DNA 甲基化调节表观遗传标记^[4-5]。各种 RNA 独特的长

度和核苷酸序列决定了不同的 RNA 具有各自的结构特征, 而且相同种类的 RNA 在结构上具有一定的相似性, 因此在结构解析时就需要分别进行讨论。

1.1 RNA 二级结构的影响因素

RNA 二级结构参与调控多种体内 RNA 相关过程, 但是 RNA 的结构对环境因素的变化十分敏感。RNA 的序列决定了 RNA 形成的结构范围, 将已经具有结构的 RNA 加以变性条件后再将其复性, 在没有其他因素的情况下能够恢复最初的结构。但是生物体内的 RNA 结构往往是多重因素共同作用的结果。导致 RNA 结构变化的因素可以归纳为生物的体内因素和外界因素, 其中体内因素包含转录、空间位阻、RNA 结合蛋白(RNA binding protein, RBP) 和相互作用离子, 外界因素包括温度、渗透压等^[6-7]。

1.1.1 影响 RSS 的体内因素 RNA 的二级结构在生物体内受到转录反应、空间位阻、RBP 和相互作用离子的限制。RNA 在折叠之初受到具有 RNA 伴侣活性的蛋白质的引导, 避免仅仅在局部形成自由能最低的构象, 这种只关注局部的折叠就可能导致整体 RNA 形成的构象不符合能量最低条件, 难以产生稳定结构。当我们利用 RNA 序列预测结构时难以将所有导致结构变化的因素全部考虑以修正结构, 因此, RNA 真实的体内结构可能与体外自发形成的结构或计算机预测的最低自由能结构存在较大差异^[8]。

RBP 与 RNA 相互作用可以影响 RNA 结构。许多 RBP 具有 RNA 结合功能域, RNA 序列和结构特征使它们被 RBP 识别并定向结合。RNA 折叠形成三级结构过程中需要 RBP 参与, 更为复杂的三级结构使其更好地行使生物体内的功能。如 tRNA 具有的三叶草二级结构需要经过共轴堆叠形成倒 L 形的三维结构, 才能携带氨基酸进入核糖体参与蛋白质合成。对一组人类 MRPS21 核糖核酸等位基因 DMS-MaPseq 数据进行共同和单独分析时发现, 在 RBP 引起局部 RNA 结构重排, 即部分区域产生核糖核酸结构异质性时, RNA 折叠状态具有显著差异^[9]。

化学修饰同样可以改变 RNA 结构。RNA 修饰包括甲基化、羟甲基化、乙酰化等^[10]。RNA 分子上的修饰影响 RNA 与 RBP 结合和结构形成, 常见的 RNA 修饰位点包括 5- 甲基胞嘧啶(5-methyl-

cytosine, m⁵C)、肌苷、2'- 氧甲基化 (2'-O-methylation)、假尿嘧啶 (pseudouridine, ψ) 和 N⁶· 甲基腺嘌呤 (N⁶· methyladenine, m⁶A) 等^[11]。目前关于 mRNA 的甲基化修饰研究已有广泛报道, 其中 m⁶A 修饰是研究最为广泛的一种修饰, m⁶A 由甲基转移酶催化, 去甲基化酶去除, 并与 m⁶A 结合蛋白相互作用^[12], 这种修饰方式在许多疾病中起重要作用, 特别是肿瘤的形成。m⁶A 也被证明与干细胞分化和生物钟控制有关, 修饰后的 RNA 响应紫外线照射与热休克胁迫产生动态变化^[13]。许多试剂能够定向地修饰 RNA 骨架或者碱基, 为 RNA 的结构解析提供支持。

要解析 RNA 结构的生理功能, 就需要讨论 RBP 和化学修饰的影响。Tomezsko 等^[14] 基于 DMS-seq 发现酵母体内的结构化 mRNA 区域比体外的要少得多, 并且 RNA 结构变化与 ATP 含量相关, 当 ATP 浓度变化时 RSS 也会相应产生变化, 这说明 RNA 的结构是处于动态变化的。RNA 在体外难以获得与体内环境相同的 RBP 和离子条件, 且体外实验难以达到与胞内相同的活性氧浓度, 所以体外进行的 RNA 结构探测难以复原 RNA 在生物体内的真实构象。

1.1.2 影响 RSS 的外界因素 温度和离子渗透压都是影响 RNA 二级结构的外界因素。即使是耐热的 RNA 结构也经常在细胞中发生变性, 细胞环境改变对 RNA 结构产生重要影响^[15]。原核生物中存在一种典型的 RNA 温度计 (RNA thermometers), 这种 RNA 对温度变化敏感, 其结构随温度变化而发生改变。低温时其结构能抑制转录过程, 温度升高后结构变化不再具有抑制功能, 温度降低后则结构恢复, 由此使 RNA 结构与转录产生联系^[16]。此外, 在不同的温度下 RNA 采用不同的结构 (RNA switch) 在翻译调节中发挥作用。在植物中, 对黄化的拟南芥幼苗体内的 RNA 结构进行采样的研究表明, 与胁迫反应相关基因的 mRNA 在每个核苷酸中显示出较高的自由能或更长的最大环长度, 这两者都为构象变化提供了灵活性^[17]。热胁迫也被证实能够使 tRNA 的结构部分展开^[18]。

拟南芥中一类非编码转录本 *COOLAIR* 通过 smStructure-seq 被证实温暖环境和冷处理环境中, 具有不同的 RNA 构象。*COOLAIR* 参与 *FLC* (*FLOWERING LOCUS C*) 的转录调控, 以决定喜温

植物冬季年度或快速循环的繁殖策略, 也参与促进低温诱导的转录关闭^[19]。冷处理及冷处理后恢复的植物样本 RNA 结构经过可视化都阐明了 RNA 结构的差异, 但是植物中具体与温度相关的 RNA 结构变化机制还未得到完全解释, 那么系统地分析植物中 RNA 特征结构就可能为植物的环境胁迫传感器提供其机理解释。

渗透压对 RNA 二级结构造成的影响同样在植物生物学中具有重要意义。植株受到如洪水、干旱、过高的土壤盐分等渗透胁迫时, 细胞内渗透浓度就产生大规模变化。盐胁迫时拟南芥转录本丰度的变化与胁迫诱导的相对过饱和度的变化呈负相关, 在这项研究中, 盐胁迫诱导了 RNA 的重折叠, 尤其是与应激相关的 mRNA, 而 RNA 的重折叠增加了转录本丰度^[20-21]。

1.2 植物中 RSS 的功能

植物中的 mRNA 不仅作为遗传信息的载体, 还参与转录调控过程。植物体内的 lncRNA 和 miRNA 等与蛋白结合形成复杂的聚合体, 对植株的组织、器官甚至整个植株的生理功能产生重要的影响。核糖体 rRNA 被包裹在复杂的三级结构中, 只有形成合适的结构才能够在翻译过程中催化蛋白质合成, lncRNA 和小 RNA 也被证明参与相变、影响蛋白-蛋白互作等^[22]。随着探测技术的进步, 植物体内 RSS 具有的功能也被更加深入地研究, 下文对植物中 RSS 具有的功能进行介绍。

1.2.1 植物中 RSS 影响转录和翻译 植物中 RNA 二级结构与 mRNA 的转录加工过程紧密相关。多聚腺苷酸化 (Polyadenylation) 是 mRNA 加工的一个关键步骤, 这一过程切割 mRNA 的前体, 并在 mRNA 的 3' 端添加 poly (A) 尾^[23]。在拟南芥中, 约 60% 的 mRNA 具有多聚腺苷酸位点。一种广泛的认知是 pre-mRNA 的裂解和聚腺苷酸化依赖于 poly(A) 序列, 典型序列如 AAUAAA 或 AUUAAA。拟南芥的核 RNA 上 poly (A) 结构位点和备选位点上有两个单链区域, 分别位于 -28 nt 到 -17 nt 之间和 -4 nt 到 +1 nt 之间。而 -28 nt 到 -17 nt 的上游区域是传统 poly(A) 信号的富集区域, 这表明多聚 (A) 信号和多聚 (A) 位点都倾向于单链^[20]。哺乳动物 mRNA 的 3' 端区

域相较于其他区域更容易折叠。研究认为, mRNA 3' 区域的折叠将 poly (A) 信号和裂解位点固定在最佳距离, 促进了 mRNA 的多聚腺苷酸化^[24]。Liu 等^[20]在拟南芥中利用 SHAPE 标记并分析 mRNA 3' 端和 5' 端 RNA 结构特征发现, 核 mRNA 和胞质 mRNA 在翻译起始和停止位点上的折叠方式不同, 说明序列相同的核 mRNA 和胞质 mRNA 可能折叠形成不同的结构来执行各自的生物学功能。

RNA 二级结构可以调节翻译效率, 并且成熟的 mRNA 具有能够影响翻译的 RNA 结构^[25]。在拟南芥和水稻上的研究表明, mRNA 不具有复杂结构时往往具有高翻译能力, 这说明 mRNA 的结构可能在植物中具有翻译调节功能^[26]。在体外条件下, 翻译起始位点和核糖体暂停位点的结构特征是保守的。在哺乳动物和植物中, mRNA 起始密码子上游 5 nt 位置被证明是不具有复杂结构且易于靠近的, 这种简单的结构便于核糖体的靠近和翻译的起始^[27]。实际上, RNA 的结构以及修饰也被认为是一种具有普遍性的转录后调控的控制模式^[28]。

1.2.2 植物中 RSS 影响可变剪接 研究认为 RNA 结构变化影响 mRNA 的可变剪接。剪接是 mRNA 加工的关键步骤, 可变剪接位点的识别需要 RNA 结构开关参与。真核生物的基因是由外显子和内含子构成的, pre-mRNA 成为成熟 mRNA 的过程需要经过剪接去除内含子, 将外显子拼接成链。RNA 的剪接过程依赖剪接体、RBPs 及 mRNA 顺式作用元件的共同作用^[29]。其中, RNA 结构作为顺式调控元件招募蛋白质影响 RNA 剪接, 是转录后调控中不可缺少的部分。Ding 等^[30]利用 DMS-MaPseq 观察拟南芥发现在 5' 剪接位点上游约 40 nt 区域, 剪接转录本的结构明显比未剪接转录本的结构更加灵活, 这说明在 5' 剪接位点, 较强的 RSS 阻碍了拼接的第一步。

1.2.3 植物中 RSS 影响 RBP 的结合 RNA 二级结构影响 RNA-RBP 结合。如上所述, RBPs 参与 RNA 结构形成, 同时也受到 RNA 结构的调控。RNA 二级结构在体外和体内都能够影响 RBP 的结合。目前对于 RNA 二级结构影响 RNA-RBP 最充分的证据是哺乳动物中的 lncRNA, Xist, 它在雌性发育过程中表达后, 与关键的成核位点结合并扩散, 通过与大

量蛋白质的逐步相互作用引起全染色体沉默。而植物中 RBP-RNA 的研究也已经十分充分, 同一 mRNA 形成不同的二级结构后, 尽管能够结合相同的蛋白质, 但功能不同^[31]。更为广泛的表现是 RNA 结构因为生物体内的“开关”产生变化后, 与 RBP 结合的结构发生改变, 从而调控与 RBP 的结合^[13]。编码蛋白质结构域连接或无序区域的 RNA 结构会减弱蛋白质的延伸, 从而促进蛋白质的自然折叠^[6]。

高温、冷胁迫和盐胁迫等生物胁迫引起植物的应答反应被证实与 RBP 紧密相关, 而 RNA 结构通过结合或不结合 RBP 同样参与植物的抗逆和病原反应。盐胁迫等多种胁迫能够诱导拟南芥中 RNA, 特别是与胁迫相关的 mRNA 的再折叠, RNA 的再折叠改变了转录本的丰度, 从而调节植物的应激能力。不同于哺乳动物, 在拟南芥和水稻 RNA 结构研究中发现 UTR 区相较 CDS 区具有更少的结构, 表明 RNA 结构与翻译效率相关^[30]。在植物已有的研究中, 或许是由于调控结构元件的存在, 病原体反应性转录本往往是具有最高度折叠的。拟南芥中许多高度结构化的 mRNA 编码的 DEFENSIN-LIKE (DEFL) 蛋白、低分子量富含半胱氨酸蛋白、病程相关蛋白都在植物病原应答中发挥作用^[22]

2 基于实验的 RNA 结构预测方法

RNA 凭借其复杂的结构在生物体内发挥多种功能, 然而 RNA 结构在体内的动态变化给探测带来很大难度。近年来由于新一代测序技术的广泛应用, RNA 结构的探测方法也不断更新。高通量测序为大量的方法提供了技术支持, 它可以实现大规模的平行 DNA 测序, 同时检索数以千万计的 DNA 序列的信息。高通量测序足以在一个实验中覆盖全基因组 RNA 结构并表征数千种 RNA 的二级结构特征, 为 RNA 结构在体内外的功能及结合位点探究提供支持。通过实验直接或间接地获取 RNA 的结构是解析 RNA 的可信方法, 由于 RNA 的结构在体内体外具有差异, 因此通过实验等方式描述 RNA 的实际结构就是研究 RNA 及 RNA-RBP 复合体功能的必经之路, 而辨析 RNA 体内结构的方法研究也成为了研究 RNA 结构的热点之一。

2.1 体外分析RNA结构方法

2.1.1 物理法 最早使用在 RNA 结构探索上的技术是 X- 射线 (X-ray)。Chao 等^[32] 利用 X- 射线衍射初步确定了 tRNA 的结构。tRNA 三叶草结构的发现为 RNA 结构相关研究奠定了理论基础, 即 RNA 可能会形成特定的结构来执行生物学功能。此后较长的一段时间内, 检测 RNA 结构的实验方法主要为核磁共振技术 (nuclear magnetic resonance, NMR) 和 X- 射线晶体衍射技术。X- 射线和 NMR 最早应用于蛋白结构探测, 有趣的是尽管两者在 RNA 结构领域应用时间前后差别, X- 射线能够通过大量的图片分析补充核磁共振实验数据集不完整结构信息, 核磁共振获取的 RNA 结构比 X- 射线模型表现出更多的空间冲突和构象歧义^[33], 两者在实验中相辅相成^[34]。这两种技术涉及放射性标记, 对实验人员和 RNA 状态要求较高, 且分析单一 RNA 结构所需 X 光片和结构数量大所需成本高。近年来, 冷冻电镜为探测 RNA- 蛋白复合体的结构提供有力途径, 然而单一的无蛋白结合的 RNA 因其分子量小且结构不单一而难以仅利用冷冻电镜进行分析。即使有体外实验工作量大、难以展示 RNA 在生物体内的真实结构等限制, 采用物理方法能够直观地展示 RNA 的可视三维结构, 目前仍然是 RNA 结构的重要获得途径。

2.1.2 酶切探针标记法 酶法和化学法解构 RNA 现在十分依赖高通量测序技术 (next generation sequencing, NGS), 而 NGS 还未出现时, 广泛采用凝胶电泳的方式对探针位点进行解析^[35]。染料终止序列测序通常被称为 Sanger 测序, 它使用凝胶或毛细管电泳展示 RNA 探针位点, 一次只能提供有限数量的序列信息, 这决定了该方法只能用于较短的 RNA 序列。聚丙烯酰胺凝胶电泳 (polyacrylamide gel electrophoresis, PAGE) 读取修饰的方法逐渐被逆转录 (reverse transcription, RT) 并进行测序替代, 直接对 RNA 的结构特征进行解读。结构特异性核糖核酸酶常作为酶探针探测单链和双链区, 用于标记的酶探针包括 RNase V1, RNase S1, RNase T1 和 RNase T。其中 RNase S1 和 P1 对 ssRNA 切割有特异性, 而 RNase V1 针对 dsRNA^[36]。常见的两种酶探针分别切割 RNA 链产生 5' 磷酸和产生 3' 磷酸接

口, 酶切之后引入末端标记, 再进行测序就可以找到 cDNA 序列上的反转录断裂位点。对于成分较为复杂的目标 RNA, 单一的酶探针难以达到研究目的, 通常选用多种位点不同的酶共同切割 RNA, 再对 RNA 进行末端标记。然而, 核酸酶受空间位阻影响较大, 且难以透过细胞膜, 实验过程中使用的 Mg^{2+} 也可能影响 RNA 折叠, 因此体内真实的 RNA 结构就难以用酶法探究。

2.1.3 化学探针标记法 同样应用探针标记的化学探测方法因为试剂分子较小, 不受酶法局限性的拘束而被广泛应用。目前, 已经应用于 RNA 标记的试剂包括烷基化试剂如乙基亚硝基脒 (ENU)、硫酸二甲酯 (DMS)、碳化二亚胺 (CMTC)、1- (3- 二甲氨基丙基) -3- 乙基碳二亚胺 (EDC)、叠氮烟酸甲酰 (NAz) 和重亚硫酸盐等^[51]。化学探针又可以分为碱基特异性探针和非碱基特异性探针, 即利用试剂修饰碱基或核苷酸骨架。SHAPE 试剂 2- 甲基烟酰胺唑 (NAI) 或 1- 甲基 -7- 硝基异尿酸酐 (1M7) 将所有 4 个核苷酸共同的 2'- 羟基 (2'-OH) 酰化, 硫酸二甲酯 (DMS) 将腺嘌呤 (N^1 -A) 的无保护和非碱基配对 N^1 位置、胞嘧啶的 N^3 位置 (N^3 -C) 和鸟嘌呤 (N^7 -G) 的 $N7$ 位置烷基化 (表 1)。由于这些修饰阻碍了传统的逆转录酶通过核苷酸, 单链区域的修饰核苷酸可以通过 PAGE 或高通量测序检测到标记为逆转录停止。而这些探针往往难以结合已被 RBP 结合的 RNA, 因此可以通过修饰的读取判断 RNA 结合 RBP 的区域, 作为 RNA 与 RBP 结合基序的预测验证手段 (图 1)。由于生物膜通透性差, CMCT 等难以进入细胞中, 一般应用于体外研究 RNA 结构。

2.2 体内探测RNA结构方法

2.2.1 化学探针标记 RNA 方法 有一类小分子标记试剂能够进入生物膜在生物体内标记 RNA, 由此获得生物体内的 RNA 结构信息, 构建更贴近 RNA 在生物体内实际结构的模型。目前已被使用的标记探针包括 SHAPE、DMS 和 EDC 等。近年来, 一种被称为突变谱分析改进的方法被开发, 该方法通过在逆转录反应过程中在修饰位点引入错配而非停止逆转录来确定化学修饰的核苷酸, 从而高效率地利用测序来获得修饰位点信息^[52]。

表 1 用于 RNA 标记实验的酶和探针
Table 1 Probes and enzymes for RNA labeling

试剂 Reagents	体内或体外 <i>In vivo or in vitro</i>	标记位点 Targeted sites	方法 Methods	材料种类 Species
DMS	体内 & 体外 <i>In vivo & in vitro</i>	C N ³ , A N ¹ , U N ⁷	DMS-seq, Structure-seq, DMS-MaPseq, Dance-MaP	拟南芥 <i>Arabidopsis thaliana</i> ^[30] 大肠杆菌 <i>Escherichia coli</i> ^[37] 酵母细胞 <i>Saccharomyces</i> ^[38] 创伤弧菌 <i>Vibrio vulnificus</i> ^[39]
NAI, 1M7, NMIA, BzCN, _{NAI} -N3	体内 & 体外 <i>In vivo & in vitro</i>	2'-OH	SHAPE-seq, icSHAPE, SHAPE-MaPseq	斑马鱼 <i>Danio rerio</i> 小鼠 <i>Mus musculus</i> 拟南芥 <i>Arabidopsis thaliana</i> 水稻 <i>Oryza sativa</i> ^[26]
甲基乙二醛 Methylglyoxal	体内 & 体外 <i>In vivo & in vitro</i>	C N ³ , 4-NH ₂ , A N ¹ , 6-NH ₂ , G N ¹ , 2-NH ₂	RIC-seq	小鼠 <i>Mus musculus</i> ^[40] 枯草芽孢杆菌 <i>Bacillus subtilis</i> 大肠杆菌 <i>Escherichia coli</i> ^[41]
EDC	体内 & 体外 <i>In vivo & in vitro</i>	G N ¹ , U N ³	Structure-seq	小鼠 <i>Mus musculus</i> ^[42]
PDS, NAI	体外 <i>In vitro</i>		rG4-seq	HeLa 细胞 HeLa cells 大肠杆菌 <i>Escherichia coli</i> 拟南芥 <i>Arabidopsis thaliana</i> ^[43-44]
CMCT	体外 <i>In vitro</i>	U N ³ , G N ¹	CIRC-seq	小鼠 <i>Mus musculus</i> ^[45-46]
CMCT, DMS	体外 <i>In vitro</i>	C N ³ , A N ¹ , U N ⁷ , G N ¹	CIRS-seq	小鼠 <i>Mus musculus</i> ^[47]
RNase V1, RNase S	体外 <i>In vitro</i>		PARS	酵母 <i>Saccharomyces</i> ^[36] 人体淋巴细胞 Human lymphoblastoid cell [*] ^[9]
RNase P1	体外 <i>In vitro</i>		Fragseq	小鼠 <i>Mus musculus</i> ^[48]
RNase V1 and RNase One	体外 <i>In vitro</i>		ssRNA-seq, dsRNA-seq, PIP-seq	果蝇 <i>Drosophila melanogaster</i> 秀丽隐杆线虫 <i>Caenorhabditis elegans</i> ^[49] 拟南芥 <i>Arabidopsis thaliana</i> ^[50]

Note : DMS, dimethyl sulfate ; NAI, 2-methylnicotinic acid imidazolide; 1M7, 1-methyl-7-nitroisatoic anhydride; NMIA, N-methylisatoic anhydride; BzCN, benzoyl cyanide; NAI-N3, 2-methylnicotinic acid imidazolide azide; EDC, 1- (3-dimethyl aminopropyl) - 3-ethyl carbamide; PDS, pyridostatin; CMCT, 1-cyclohexyl-3- (2-morpholinoethyl) carbodiimide metho-p-toluenesulfonate. Methods : DMS-MaPseq, dimethyl sulfate mutational profiling with sequencing; Dance MaP, deconvolution and annotation of ribonucleic conformational ensembles; SHAPE-seq, selective 2'-hydroxyl acylation followed by primer extension; icSHAPE, *in vivo* click selective 2'-hydroxyl acylation and profiling experiment; RIC-seq, RNA *in situ* conformation sequencing technology; rG4-seq, RNA G-quadruplex sequencing; CIRS-seq, chemical inference of RNA structures followed by massive parallel sequencing; CIRC-seq, Circular RNA-Seq; PARS, parallel analysis of RNA structure; Fragseq, fragmentation sequencing; PIP-seq, protein interaction profile sequencing

利用 SHAPE 试剂标记并分析的方法 : SHAPE-seq (selective 2'-hydroxyl acylation followed by primer extension), icSHAPE (*in vivo* click selective 2'-hydroxyl acylation and profiling experiment), SHAPE-MaPseq 等 (表 1)。SHAPE 试剂能够对 RNA 上的 2'-OH 进行标记, 利用 SHAPE 方法就可以同时获得 4 种核苷酸的结构信息。在 icSHAPE 中, 细胞内的 RNA 被 NAI-N₃ 修饰, 使其包含生物素, 用链霉亲和素结合珠纯化。SHAPE 试剂加合物阻断 NAI 修饰碱基上的

逆转录延伸, 生成一个具有相应 cDNA 截断的文库。对于 SHAPE-MaPseq, 通过深度测序在逆转录过程中将修改后的核苷酸读取为不匹配, 通过 SHAPE 反应性计算得到修改后的核苷酸。

利用 DMS 试剂标记并进行分析的方法包括 : DMS-seq, Structure-seq, DMS-MaPseq 和 Dance MaP 等。DMS 的缺点是它只能提供腺嘌呤和胞嘧啶上的碱基配对数据, 缺少可靠的鸟嘌呤和尿嘧啶碱基配对探针, 预测获得的 RNA 结构不全面且可靠性较

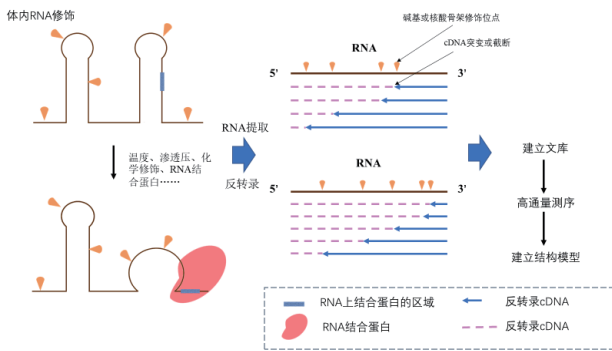


图1 RBP影响体内探针修饰位点

Fig. 1 RBP affects modification sites *in vivo*

差。DMS 早期仅作为体外标记探针使用，直到基于 DMS 在拟南芥、哺乳动物和酵母中体内的全基因组 RNA 结构分析的方法发表，而 DMS 在植物中应用条件的优化和方法的完善也为全局分析 RNA 结构提供帮助^[15]。在利用探针截断反转录的方法中，每个 RNA 分子只能观察到一个化学修饰的单一位点，因此低丰度的 RNA 不适用基于截断的 RNA 结构探测。

读取截断或者突变的方法是将 RNA 反转录后进行测序。解读标记的酶包括 Superscript III、SSII-Mn²⁺ 和 TGIRT (thermostable group II intron reverse transcriptases) 等^[52]。SSII-Mn²⁺ 在使用 Superscript II 时引入二价金属离子 Mn²⁺，以促进逆转录酶通读标记后的 RNA^[53]。MaPseq 是近年发展起来的一种替代 RT-stop 分析方法的方法，该方法使用的反转录酶工作时连续地反转录修饰后的 RNA 碱基，并在修饰处插入突变，由此可以对较长的序列进行测序。与 Structure-seq 一样，MaPseq 可以用于特定基因或全基因组的方式，检测到的突变相当于未配对的核苷酸。TGIRT 可以读取单个分子上的多个 DMS 修饰，增加给定修饰碱基上的测序深度。因其读取长度长、可靠性好，TGIRT 取代了更多传统的逆转录酶，如在 Structure-seq 和类似技术中使用的 Superscript III。DMS-MaPseq 即应用耐热逆转录酶通过在 DMS 修饰位点引入错配来分析修饰率。Zubradt 等^[54-55] 在酵母中利用 DMS-MaPseq 分析特定位点结构，并首次在动物体内运用该方法进行全基因组探测和特定基因结构分析，Wang 等^[56] 将 DMS-MaPseq 应用在拟南芥体内 RNA 结构全局分析和特定位点结构分析。

由于单独应用 DMS 不能提供 4 种碱基的信息，一系列化学修饰试剂也被开发用于弥补结构预测中的信息差距。水溶性碳二亚胺 1-乙基-3-(3-二甲氨基丙基)碳二亚胺 (EDC) 已成为一种有效的体内探针^[57]。EDC 与 EMCT 相似，更倾向于修饰 U 和 G，EDC 容易穿透细胞壁和细胞膜，并已被用于探测多个生命领域的 RNA，包括枯草芽孢杆菌、水稻和培养的 MEF 细胞。Mitchell 等^[58] 开发的 Structure-seq 就是综合 EDC 和 DMS 两种试剂进行读取。EDC 易于穿透细胞壁和细胞膜，已被用于包括枯草芽孢杆菌、水稻和 MEF 培养细胞在内的多种生物体内 RNA 标记，可在体内探测 4 种不同的 RNA 碱基。

LASER-seq (light activated structural examination of RNA, LASER) 能够修饰成对的核苷酸，使识别具有高级结构 RNA 的构象变化成为可能。烟酰胺叠氮化物 (NAz) 不标记碱基对，而是在激光照射下产生离子与胞嘧啶 C⁸ 和腺嘌呤 C⁸ 位置结合^[59]。但是该方法不能准确标记被蛋白结合或正在翻译的碱基，无法在 RNA-蛋白质相互作用方面提供帮助。此外乙二醛试剂也可作为修饰腺嘌呤、鸟嘌呤和胞嘧啶的探针，目前乙二醛体系已经在水稻、大肠杆菌、枯草芽孢杆菌和小鼠胚胎成纤维细胞 (MEF) 的体内 RNA 结构探测中得到广泛应用^[60]。

化学探针因其分子量小、易于结合 RNA 分子而广泛应用，目前还处于各具优缺点的状态。化学探针可以进入生物体内标记 RNA，而且可以同时运用多种探针在体内探测 4 种碱基的结构信息。然而不同探针所需的缓冲液和离子浓度不同，且难以解读有 RBP 结合或高级结构的 RNA，更加方便的标记探针还有待开发。

2.2.2 基于 RNA 交联的方法 小分子探针修饰方法能够在读取碱基信息后解析 RNA 结构，但不能明确 RNA 中碱基互补配对的情况和 RNA 分子之间的互作。基于交联的方法可以直接捕获分子内和分子间的 RNA-RNA 相互作用^[40]。在细胞内提供交联所需条件后，提取 RNA 后酶切消化，其中相互作用的片段仍连接在一起。由于连接产物产生的嵌合体结构，测序时不能进行读取，从而获得 RNA 相互作用的位点信息。许多基于交联和接近连接的技术已经被开发，在酵母、人类、小鼠和病毒等材料中均可获取

RNA-RNA 相互作用信息。

交联可以在体内和体外进行。PIP-seq (protein interaction profile sequencing) 是一项最近发展起来的技术,用于识别 RNA 在其原生状态下的二级结构。该技术可以应用于生物的组织或细胞中,首先使用甲醛或紫外光交联,随后进行的 ssRNA-seq 和 dsRNA-seq 同时对 RNA 二级结构和 RNA 蛋白相互作用进行全基因组识别,就可以得到 RNA-蛋白相互作用信息。CLASH (crosslinking ligation and sequencing of hybrids) 是一种基于紫外交联的方式,能捕捉与已知蛋白相关的 RNA-RNA 互作^[61]。补骨脂素交联可以识别直接的 RNA-RNA 相互作用,甲醛交联也被利用在识别有 RBP 参与的 RNA 互作中^[62]。MARIO (mapping RNA interactome *in vivo*) 方法被应用在小鼠体内进行 RNA-RNA 互作的识别验证,同时能够对 RNA 的二级及三级结构进行解析^[63]。该方法通过 RNase I 消化位点的密度来确定 RNA 单链区域,通过邻位连接技术捕获每个 RNA 的空间近端位点。尽管 RNA 交联的方法能够很好地在体内进行,但它仍然未解决交联位点覆盖面窄和实验中可能产生假阳性的问题。

3 RNA 结构研究的计算机方法

相较于实验方法需要花费大量时间和金钱,利用计算机和数学模型预测 RNA 二级结构以其经济、快速和安全的优势成为生物信息学领域的研究目标。RNA 二级结构预测常用方法为最小自由能法 (minimum free energy, MFE) 和序列比较分析法。即基于 RNA 碱基序列,参考或不参考试验获得的 RNA 结构数据,通过计算自由能和热力学分析获得结构。但是该方法不能保证在体内的准确性和真实性,很多现有的分析方法需要结合 RNA 探针标记获得的序列特征。现有 RNA 二级结构数据库中包含大量基于 NMR、X-ray 以及共变模型获得的 RNA 结构,且数据库中多为结构较为保守的非编码 RNA,如 rRNA、tRNA、Group I intron、Group II intron 以及核糖体开关等。基于已知 RNA 结构,机器学习和深度学习在 RNA 结构预测方面也展示出优越性。

3.1 基于实验数据构建二级结构

近年来,解析 RNA 二级结构有两种常用的方法,

其一是通过测序获得 RNA 的断裂或修饰信息,其二是通过碱基序列信息构建概率模型。RNA 序列信息的完整性对结构的预测具有很大影响, RNA 的碱基配对信息和单链部分的信息都有助于搭建更高分辨率的结构模型。因此将双链 RNA 测序和单链 RNA 测序数据结合,就可全面分析碱基配对情况并构造 RNA 结构模型^[22, 64]。依据碱基序列构建的模型和实验探测获得的结构信息指向的模型之间存在的交集就很可能是该 RNA 真实存在的结构之一。

依据核苷酸序列预测二级结构的方法始于 20 世纪 60 年代,将实验获得的碱基配对状态作为计算预测的限制条件,计算获得能量最低、被视作稳定状态的结构。此时如果实验数据不能完全展示碱基配对情况,或者同一序列同时存在多种结构,那么根据这种条件计算得到的结果就是不可信的。

在利用高通量测序获得 RNA 序列后,对 RNA 序列数据进行条件限制,读取信息并作结构预测的算法现在已有多种。最小自由能就是将实验数据转化为自由能项叠加到特定结构单元的自由能计算中,这种自由能项也被称为伪自由能 (pseudo-energies)。这种计算的前提是 RNA 碱基配对后自然形成自由能低的结构。在已知 RNA 序列的情况下,基于最小自由能可以对结构进行预测,但其并不能展示 RNA 在体内的真实构象,只能准确预测最低自由能结构中大约 70% 的已知碱基对状态^[65]。热力学模型中, Schrodner 等^[66]的最近邻模型将一个二级结构分解成几个特征子结构,标记为最近邻环,如发夹环、内环、凸起环、碱基对叠加,多分支回路和外部回路。每个最近邻回路的自由能可以通过把代表该回路的自由能参数相加计算。通过对分解的最近邻环自由能求和来计算整个 RNA 二级结构的自由能。利用 Zuker 算法可以有效地计算出具有最小自由能的最优二级结构^[67]。基于以上理论的方法包括 M-fold、UNAFold、RNAfold 和 RNA structure 等^[68]。随着 RNA 长度增加,预测所获得的可能二级结构数量呈指数增长,要在如此大量的预测结构中找到 RNA 的真实结构是自由能最小化法面临的巨大挑战。而序列比较法利用 NMR 和 X-射线等方法获得结构数据包,假定序列相似的 RNA 即使在不同物种中也具有相同或高度相似的结构,对已知序列和已知结构的

序列进行比对,从而获得该序列预测结构的方式,现在更多地用于 RNA 三级结构的预测中。

3.2 结合机器学习的混合方法

机器学习的核心难点是准确识别一组代表性特征,特征选择(feature selection)极大地影响模型质量。Kengo 等^[69]利用 MXfold2 算法发现,当使用深度神经网络学习的 RNA 折叠得分与 Turner 的最近邻自由能参数集成在一起时可使过拟合最小化。整合热力学信息有助于提高基于深度学习的 RNA 二级结构预测的可靠性,结合热力学模型和机器学习的混合方法能弥补彼此的缺点,如 SimFold 能够更准确地从训练数据中估算热力学参数,包括 RNA 序列、已知的部分二级结构和已知二级结构的自由能^[70]。

深度学习是机器学习的一个分支,在生物信息学、生命科学等广泛领域均有应用,近年已经在蛋白结构预测和 DNA 结构预测方面有所报道。利用深度学习预测 RNA 结构面临的主要问题是已知结构的 RNA 的可用于训练的数据非常有限,尤其缺乏复杂的高级 RNA 结构。由于只有少量高分辨率的 RNA 结构可用,深度学习模型只能先通过比较分析注释的大型 RNA 二级结构数据库进行训练,然后再将学习转移到从 3D 结构衍生出来的精确二级结构上。除了预测单条序列的 RNA 二级结构外,概率模型还可应用于预测一组结构相似的 RNA 序列的共同结构。目前深度学习方法已经被应用于探查人体内 RBP-RNA 结合位点。有报道对 riboSNitches 利用 CLIP-seq 等数据库进行解析发现 riboSNitches 引导的 RBP-RNA 结合后结构改变可能导致等位基因特异性结合的增加^[71]。

近年来不断有对 RNA 测序数据进行分析获得结构信息的软件被开发,目前应用较广的包括 MC-Fold/MC-Sym^[72]、RNA structure^[73]、ViennaRNA package^[74]、StructureFold^[26]等。智能地筛选有效限制条件对于 RNA 分析技术的发展而言是很有必要的。其中 StructureFold 可以将实验获取的高通量 RNA 结构分析数据输入后作为限制条件进行结构预测。通过依次使用迭代映射、获取 RT 信号和反应模块,该方法可以得出每个核苷酸的 FSR (final structural reactivities),这提供了 RNA 以单链存在的

可能性。同样,对于植物体内 RNA 结构的预测,机器学习方法也有着广阔的发展空间与应用场景。然而,不论是传统机器学习方法还是结合神经网络的深度学习方法,对 RNA 结构的训练数据集都有着较强的依赖性,当植物体内 RNA 结构特征越能被准确的识别和探测,建立质量可靠的高分辨率 RNA 结构数据库的可行性也将不断提高。只有基于准确可靠的探测数据集进行训练和检验,机器学习方法才能在植物 RNA 结构预测中更好的延伸和应用。

4 总结与展望

RNA 不仅是遗传中 DNA 和蛋白之间用于信息传递的桥梁,其丰富的结构信息足以成为生物遗传和生理调控中极为重要的研究方向。数十年以来,对于 RNA 结构的研究从未停止, RNA 结构分析仍面临问题与挑战,最关键的一点就是 RNA 体内状态的改变。尽管现有高通量测序和分析方法能对特定条件如蛋白结合下的 RNA 结构进行解析,并且评价 RNA 结构的打分系统也在不断进化,但评分的标准仍然在不断变化,如何制定出使 RNA 结构最贴近体内广泛存在的结构的评分标准仍有待解决。

现有分析 RNA 结构的测算方法大多需要实验数据的支撑,真实可靠的数据和测算出可信的 RNA 结构是同源序列分析和比对的基石。应用广泛的化学标记法的多种标记试剂在植物中已有报道,但仍存在植株大小和部位的限制,而标记试剂成分对植物生理或 RNA 结构的影响还不能完全排除,随着 RNA 结构研究的不断深入,效率更高、更易进入植物且无负面影响的标记探针亟待开发。利用标记 RNA 的方式不仅能够对 RNA 的二级结构进行测算,还能够进一步推算 RNA 的三级结构。尽管 RNA 二级结构还未完全阐明,但是生物体内的生理活动也依赖于 RNA 的三维结构,探究 RNA 三级结构行之有效的办法还在探索中。利用计算机基于 RNA 结构数据库预测 RNA 三维结构的方法层出不穷,Vfold3D、3DRNA、FARFAR 以及基于物理模型的 iFold、NAST 和 SimRNA 等均展示出当下研究的趋势^[75]。

以往基于来源于 PDB 数据库中存储的结构所进行的计算已有成效,最近发展起来的基于神经网络

的模型如 AlphaFold 和 RoseTTAFold 也已经被证明是高效且有力的预测及评估蛋白质 3D 结构的方法。而蛋白的结构研究和 RNA 的结构研究往往是相互促进,互相学习的,我们完全有理由相信 Alpha Fold 及更多的后续研究方法将会为植物体内的 RNA 结构研究增加新的方向。

参考文献

- [1] Qian XY, Zhao JY, Yeung PY, et al. Revealing lncRNA structures and interactions by sequencing-based approaches [J]. *Trends Biochem Sci*, 2019, 44 (1): 33-52.
- [2] Sherwood AV, Henkin TM. Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses [J]. *Annu Rev Microbiol*, 2016, 70: 361-374.
- [3] Song XW, Li Y, Cao XF, et al. MicroRNAs and their regulatory roles in plant-environment interactions [J]. *Annu Rev Plant Biol*, 2019, 70: 489-525.
- [4] Mukherjee S, Murata A, Ishida R, et al. HT-SELEX-based identification of binding pre-miRNA hairpin-motif for small molecules [J]. *Mol Ther Nucleic Acids*, 2021, 27: 165-174.
- [5] Jang B, Jang H, Kim H, et al. Protein-RNA interaction guided chemical modification of Dicer substrate RNA nanostructures for superior *in vivo* gene silencing [J]. *J Control Release*, 2022, 343: 57-65.
- [6] Tang Y, Assmann SM, Bevilacqua PC. Protein structure is related to RNA structural reactivity *in vivo* [J]. *J Mol Biol*, 2016, 428 (5 Pt A): 758-766.
- [7] Ruiz-Ferrer V, Voinnet O. Roles of plant small RNAs in biotic stress responses [J]. *Annu Rev Plant Biol*, 2009, 60: 485-510.
- [8] Huston NC, Wan H, Strine MS, et al. Comprehensive *in vivo* secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms [J]. *Mol Cell*, 2021, 81 (3): 584-598.e5.
- [9] Wan Y, Qu K, Zhang QC, et al. Landscape and variation of RNA secondary structure across the human transcriptome [J]. *Nature*, 2014, 505 (7485): 706-709.
- [10] Han B, Xi W, Hong YT, et al. Mutual regulation of noncoding RNAs and RNA modifications in psychopathology: potential therapeutic targets for psychiatric disorders? [J]. *Pharmacol Ther*, 2022, 237: 108254.
- [11] Rajendren S, Karijovich J. The impact of RNA modifications on the biology of DNA virus infection [J]. *Eur J Cell Biol*, 2022, 101 (3): 151239.
- [12] 李金珂, 袁焰. N6-甲基腺嘌呤 RNA 修饰在泌尿系统肿瘤中的研究进展 [J]. *分子诊断与治疗杂志*, 2021, 13 (10): 1726-1730.
Li JK, Yuan Y. The advance of N6-methyladenosine RNA modification in urological cancers [J]. *J Mol Diagn Ther*, 2021, 13 (10): 1726-1730.
- [13] Lewis CJT, Pan T, Kalsotra A. RNA modifications and structures cooperate to guide RNA-protein interactions [J]. *Nat Rev Mol Cell Biol*, 2017, 18 (3): 202-210.
- [14] Tomezsko P, Swaminathan H, Rouskin S. Viral RNA structure analysis using DMS-MaPseq [J]. *Methods*, 2020, 183: 68-75.
- [15] Rouskin S, Zubradt M, Washietl S, et al. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo* [J]. *Nature*, 2014, 505 (7485): 701-705.
- [16] Nshogozabahizi JC, Aubrey KL, Ross JA, et al. Applications and limitations of regulatory RNA elements in synthetic biology and biotechnology [J]. *J Appl Microbiol*, 2019, 127 (4): 968-984.
- [17] Vu LD, Gevaert K, de Smet I. Feeling the heat: searching for plant thermosensors [J]. *Trends Plant Sci*, 2019, 24 (3): 210-219.
- [18] Yamagami R, Siegel JP, Assmann SM, et al. Genome-wide analysis of the *in vivo* tRNA structurome reveals RNA structural and modification dynamics under heat stress [J]. *Proc Natl Acad Sci USA*, 2022, 119 (25): e2201237119.
- [19] Yang ML, Zhu P, Cheema J, et al. *In vivo* single-molecule analysis reveals COOLAIR RNA structural diversity [J]. *Nature*, 2022, 609 (7926): 394-399.
- [20] Liu ZS, Liu Q, Yang XF, et al. *In vivo* nuclear RNA structurome reveals RNA-structure regulation of mRNA processing in plants [J]. *Genome Biol*, 2021, 22 (1): 11.
- [21] Xue Y, Gracia B, Herschlag D, et al. Visualizing the formation of an RNA folding intermediate through a fast highly modular secondary structure switch [J]. *Nat Commun*, 2016, 7: ncomms11768.
- [22] Li F, Zheng Q, Vandivier LE, et al. Regulatory impact of RNA secondary structure across the *Arabidopsis* transcriptome [J]. *Plant Cell*, 2012, 24 (11): 4346-4359.
- [23] Neve J, Patel R, Wang ZQ, et al. Cleavage and polyadenylation: ending the message expands gene regulation [J]. *RNA Biol*,

- 2017, 14 (7) : 865-890.
- [24] Wu XB, Bartel DP. Widespread influence of 3'-end structures on mammalian mRNA processing and stability [J] . Cell, 2017, 169 (5) : 905-917.e11.
- [25] Zydowicz-Machtel P, Swiatkowska A, Popenda Ł, et al. Variants of the 5'-terminal region of p53 mRNA influence the ribosomal scanning and translation efficiency [J] . Sci Rep, 2018, 8 (1) : 1533.
- [26] Yang XF, Yu HP, Sun WQ, et al. Wheat *in vivo* RNA structure landscape reveals a prevalent role of RNA structure in modulating translational subgenome expression asymmetry [J] . Genome Biol, 2021, 22 (1) : 326.
- [27] Spitale RC, Flynn RA, Zhang QC, et al. Structural imprints *in vivo* decode RNA regulatory mechanisms [J] . Nature, 2015, 519 (7544) : 486-490.
- [28] Jia GF, Fu Y, He C. Reversible RNA adenosine methylation in biological regulation [J] . Trends Genet, 2013, 29 (2) : 108-115.
- [29] 邹永新, 龚瑶琴. 影响 RNA 剪接的基因变异 [J] . 遗传, 2017, 39 (3) : 200-207.
- Zou YX, Gong YQ. Aberrant RNA splicing as the molecular basis of some pathogenic variants [J] . Hereditas, 2017, 39 (3) : 200-207.
- [30] Ding YL, Tang Y, Kwok CK, et al. *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features [J] . Nature, 2014, 505 (7485) : 696-700.
- [31] Van Treeck B, Parker R. Emerging roles for intermolecular RNA-RNA interactions in RNP assemblies [J] . Cell, 2018, 174 (4) : 791-802.
- [32] Chao JA, Williamson JR. Joint X-ray and NMR refinement of the yeast L30e-mRNA complex [J] . Structure, 2004, 12 (7) : 1165-1176.
- [33] Bermejo GA, Clore GM, Schwieters CD. Improving NMR structures of RNA [J] . Structure, 2016, 24 (5) : 806-815.
- [34] Bermejo GA, Clore GM, Schwieters CD. Smooth statistical torsion angle potential derived from a large conformational database via adaptive kernel density estimation improves the quality of NMR protein structures [J] . Protein Sci, 2012, 21 (12) : 1824-1836.
- [35] Rouda SS, Skordalakes E. Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding [J] . Structure, 2007, 15 (11) : 1403-1412.
- [36] Kertesz M, Wan Y, Mazor E, et al. Genome-wide measurement of RNA secondary structure in yeast [J] . Nature, 2010, 467 (7311) : 103-107.
- [37] Hulscher RM, Bohon J, Rappé MC, et al. Probing the structure of ribosome assembly intermediates *in vivo* using DMS and hydroxyl radical footprinting [J] . Methods, 2016, 103: 49-56.
- [38] Umeyama T, Ito T. DMS-seq for *in vivo* genome-wide mapping of protein-DNA interactions and nucleosome centers [J] . Cell Rep, 2017, 21 (1) : 289-300.
- [39] Olson SW, Turner AMW, Arney JW, et al. Discovery of a large-scale, cell-state-responsive allosteric switch in the 7SK RNA using DANCE-MaP [J] . Mol Cell, 2022, 82 (9) : 1708-1723.e10.
- [40] Sas-Chen A, Thomas JM, Matzov D, et al. Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping [J] . Nature, 2020, 583 (7817) : 638-643.
- [41] Mitchell D, Ritchey L E. Glyoxals as *in vivo* RNA structural probes of guanine base-pairing [J] . RNA, 2018, 24 (1) : 114-24.
- [42] Wang PY, Sexton AN, Culligan WJ, et al. Carbodiimide reagents for the chemical probing of RNA structure in cells [J] . RNA, 2019, 25 (1) : 135-146.
- [43] Dumetz F, Chow EYC, Harris LM, et al. G-quadruplex RNA motifs influence gene expression in the malaria parasite *Plasmodium falciparum* [J] . Nucleic Acids Res, 2021, 49 (21) : 12486-12501.
- [44] Yang XF, Cheema J, Zhang YY, et al. RNA G-quadruplex structures exist and function *in vivo* in plants [J] . Genome Biol, 2020, 21 (1) : 226.
- [45] Andrade JM, dos Santos RF, Arraiano CM. RNA structure analysis by chemical probing with DMS and CMCT [J] . Methods Mol Biol, 2020, 2106: 209-223.
- [46] Zhao XK, Xing PW, et al. Focal amplifications are associated with chromothripsis events and diverse prognoses in gastric cardia adenocarcinoma [J] . Nat Commun, 2021, 12 (1) : 6489.
- [47] Incarnato D, Neri F, et al. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome [J] . Genome Biol, 2014, 15 (10) : 491.
- [48] Underwood JG, Uzilov AV, Katzman S, et al. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing [J] . Nat Methods, 2010, 7 (12) : 995-1001.

- [49] Foley SW, Gregory BD. Protein interaction profile sequencing (PIP-seq) [J]. *Curr Protoc Mol Biol*, 2016, 116: 27.5.1-27. 5.15.
- [50] Anderson SJ, Willmann MR, Gregory BD. Protein interaction profile sequencing (PIP-seq) in plants [J]. *Curr Protoc Plant Biol*, 2016, 1 (1): 163-183.
- [51] Cordero P, Kladwang W, VanLang CC, et al. Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference [J]. *Biochemistry*, 2012, 51 (36): 7037-7039.
- [52] Siegfried NA, Busan S, Rice GM, et al. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP) [J]. *Nat Methods*, 2014, 11 (9): 959-965.
- [53] Homan PJ, Favorov OV, Lavender CA, et al. Single-molecule correlated chemical probing of RNA [J]. *Proc Natl Acad Sci USA*, 2014, 111 (38): 13858-13863.
- [54] Zubradt M, et al. Target-specific DMS-MaPseq for *in vivo* RNA structure determination [J]. *Protoc Exch*, 2016, 155:30-40.
- [55] Zubradt M, Gupta P, Persad S, et al. DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo* [J]. *Nat Methods*, 2017, 14 (1): 75-82.
- [56] Wang ZY, Wang MY, Wang T, et al. Genome-wide probing RNA structure with the modified DMS-MaPseq in *Arabidopsis* [J]. *Methods*, 2019, 155: 30-40.
- [57] Kubota M, Tran C, Spitale RC. Progress and challenges for chemical probing of RNA structure inside living cells [J]. *Nat Chem Biol*, 2015, 11 (12): 933-941.
- [58] Mitchell D, Assmann SM, Bevilacqua PC. Probing RNA structure *in vivo* [J]. *Curr Opin Struct Biol*, 2019, 59: 151-158.
- [59] Feng C, Chan D, Joseph J, et al. Light-based RNA structure probing [J]. *Nat Methods*, 2018, 15 (3): 163.
- [60] Li YY, et al. The structural modification of DNA nucleosides by nonenzymatic glycation: an *in vitro* study based on the reactions of glyoxal and methylglyoxal with 2'-deoxyguanosine [J]. *Anal Bioanal Chem*, 2008, 390 (2): 679-688.
- [61] Kudla G, Wan Y, Helwak A. RNA conformation capture by proximity ligation [J]. *Annu Rev Genomics Hum Genet*, 2020, 21: 81-100.
- [62] Cai ZK, et al. RIC-seq for global *in situ* profiling of RNA-RNA spatial interactions [J]. *Nature*, 2020, 582 (7812): 432-437.
- [63] Nguyen TC, Cao XY, Yu PF, et al. Mapping RNA-RNA interactome and RNA structure *in vivo* by MARIO [J]. *Nat Commun*, 2016, 7: 12023.
- [64] Jansson MD, Häfner SJ, Altinel K, et al. Regulation of translation by site-specific ribosomal RNA methylation [J]. *Nat Struct Mol Biol*, 2021, 28 (11): 889-899.
- [65] Mathews DH, Disney MD, Childs JL, et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure [J]. *Proc Natl Acad Sci USA*, 2004, 101 (19): 7287-7292.
- [66] Schroeder SJ, Turner DH. Optical melting measurements of nucleic acid thermodynamics [J]. *Methods Enzymol*, 2009, 468: 371-387.
- [67] Turner DH, Mathews DH. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure [J]. *Nucleic Acids Res*, 2010, 38: D280-D282.
- [68] 蔡兆奎, 薛愿超. RNA 结构解析技术研究进展 [J]. *生命科学*, 2021, 33 (3): 281-291.
- Cai ZK, Xue YC. Research progresses of RNA structure probing technologies [J]. *Chin Bull Life Sci*, 2021, 33 (3): 281-291.
- [69] Sato K, Akiyama M, Sakakibara Y. RNA secondary structure prediction using deep learning with thermodynamic integration [J]. *Nat Commun*, 2021, 12 (1): 941.
- [70] Andronescu M, Condon A, Hoos HH, et al. Computational approaches for RNA energy parameter estimation [J]. *RNA*, 2010, 16 (12): 2304-2318.
- [71] Sun L, Xu K, Huang WZ, et al. Predicting dynamic cellular protein-RNA interactions by deep learning using *in vivo* RNA structures [J]. *Cell Res*, 2021, 31 (5): 495-516.
- [72] Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data [J]. *Nature*, 2008, 452(7183): 51-55.
- [73] Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis [J]. *BMC Bioinformatics*, 2010, 11: 129.
- [74] Lorenz R, Bernhart SH, et al. ViennaRNA package 2.0 [J]. *Algorithms Mol Biol*, 2011, 6: 26.
- [75] Tan YL, Wang XX, Shi YZ, et al. rsRNASP: a residue-separation-based statistical potential for RNA 3D structure evaluation [J]. *Biophys J*, 2022, 121 (1): 142-156.