Genomics Communications

REVIEW

Open Access

https://doi.org/10.48130/gcomm-0026-0022

Genomics Communications **2026**, 3: e023

From structural probing to predictive modeling: high-throughput strategies for resolving higher-order RNA structures *in vivo*

Jinyuan Tao[1#*], Shichang Li[1#] and Minglei Yang[1,2*]

[1] *Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230027, China*
[2] *Institute of Health and Medicine, Hefei Comprehensive National Science Center, Hefei 230027, China*
[#] Authors contributed equally: Jinyuan Tao, Shichang Li
[*] Correspondence: jinyuan@ustc.edu.cn (Tao J); mlyangrna@ustc.edu.cn (Yang M)

Abstract

RNA functions are governed not only by primary sequence and local base pairing (RNA secondary structure) but also by long-range tertiary interactions, spatial proximities, ribonucleoprotein organization, and dynamic conformational ensembles that collectively constitute higher-order RNA architecture. Conventional structural biology approaches, while providing atomic resolution insights, are inherently low-throughput and often disconnected from native cellular contexts. Researchers have achieved remarkable progress in developing high-throughput experimental strategies for probing RNA higher-order structures *in vivo*, including psoralen-mediated duplex capture (PARIS, SPLASH, LIGR-seq, COMRADES), DMS-based tertiary interaction mapping, and chemical proximity crosslinking (SHARC, KARR-seq, RIC-seq, MARIO). These technologies enable transcriptome-scale interrogation of RNA structural constraints under native conditions. Meanwhile, computational modeling has evolved from physics-based *ab initio* folding and template-based assembly to deep learning frameworks capable of predicting three-dimensional (3D) RNA structures and conformational ensembles directly from sequence. This review systematically summarizes current experimental and computational strategies for resolving RNA higher-order structures, outlines their advantages and limitations, and highlights emerging applications across animal and plant systems. We further discuss how the integration of *in vivo* structural probing with artificial intelligence (AI)-driven modeling is transforming RNA structural biology from descriptive observation into predictive science, with profound implications for RNA-targeted therapeutics and molecular breeding.

Keywords: RNA higher-order structure, *in vivo* structural probing, High-throughput crosslinking, RNA 3D modeling, AI-driven prediction

Citation: Tao J, Li S, Yang M. 2026. From structural probing to predictive modeling: high-throughput strategies for resolving higher-order RNA structures *in vivo*. *Genomics Communications* 3: e023 https://doi.org/10.48130/gcomm-0026-0022

Introduction

RNA has long transcended its historical definition as a passive intermediate messenger of genetic information alongside the advancement of modern molecular biology. Beyond the linear primary sequence that encodes genetic information and orchestrates complex biological processes across different species[1–4], RNA secondary structure formed by local canonical base pairing has increasingly emerged as a central regulatory layer that governs nearly all facets of RNA functionality. More importantly, the full biological function of RNA is ultimately executed through sophisticated higher-order spatial architectures built upon secondary structural scaffolds. These encompass long-distance intramolecular base contacts, non-canonical hydrogen-bond interactions, three-dimensional spatial proximities, ribonucleoprotein (RNP) assemblies, and heterogeneous dynamic conformational ensembles. Such layered structural hierarchies act as the critical molecular bridge linking linear RNA sequences to their regulatory role outputs[5].

Classically, ribozymes execute their catalytic activities exclusively by forming intricate higher-order RNA architectures, which precisely coordinate metal ions and RNA substrates to carry out sequence-specific catalytic reactions for RNA cleavage, ligation, and splicing without protein cofactors[6–8]. Beyond ribozymes, other RNAs, including messenger RNAs (mRNAs) and long non-coding RNAs (lncRNAs), fold into intricate tertiary and higher-order architectures that orchestrate diverse biological processes across animal and plant kingdoms, from directing molecular regulatory mechanisms including epigenetic silencing and post-transcriptional RNA processing, to governing developmental programs such as embryonic cell fate specification, tissue morphogenesis and developmental timing, and further enabling environmental stress responses[9–11]. In animals, UTR-located RNA conformational switches undergo reversible conformational rearrangements to modulate gene expression at the transcriptional and translational levels [12,13]. For instance, a native riboswitch located in human vascular endothelial growth factor-A (*VEGFA*) mRNA 3′UTR undergoes reversible two-state conformational rearrangement to regulate *VEGFA* translation by integrating signals from interferon (IFN)-γ and hypoxia in myeloid cells[14,15]. The well-characterized lncRNA *Xist* folds into conserved multi-helical tertiary domains that recruit epigenetic repressors to drive X-chromosome inactivation during mammalian embryonic development[16], neural lncRNA *Pnky* adopts a compact, highly structured conformation to bind splicing factors and govern neural stem cell differentiation[17], and *C. elegans* developmental timing lncRNA *lep-5* (*lepidote-2*) containing conserved stem-loops may function as an RNA scaffold to assemble LEP-2 and LIN-28 (LINEAGE-28) complexes and control juvenile-to-adult transition[18,19]. Meanwhile, in sessile plants, RNA higher-order structures, particularly RNA G-quadruplexes (RG4s), serve as critical regulators of growth, development, and stress adaptation. The first plant RG4 was identified in the 5′-UTR of *Arabidopsis ATR* (ataxia-telangiectasia and Rad3-related protein) mRNA and strongly represses translation[20]. Transcriptome-wide structure profiling subsequently confirmed that hundreds of RG4s are folded in living cells, and genetic and biochemical analyses further demonstrated their regulatory roles in plant growth[21]. RG4s within the 5′-UTRs of

© The Author(s)

https://doi.org/10.48130/gcomm-0026-0022

SUPPRESSOR OF MAX2 1-LIKE4/5 (*SMXL4/5*) mRNAs are specifically recognized and stabilized by JULGI, a RanBP2-type zinc-finger RNA-binding protein, thereby repressing their translation and controlling phloem differentiation[22], while an RG4 in *SHORT ROOT* (*SHR*) mRNA triggers liquid–liquid phase separation essential for root tissue patterning[23]. In addition, environmental signals also act through RG4s, as cold enhances 3′-UTR RG4 folding to stabilize mRNAs and regulate cold-responsive growth[24]. Beyond that, our previous *in vivo* single-molecule studies revealed that *COOLAIR*, an antisense lncRNA transcribed from the *Arabidopsis* floral repressor locus *FLOWERING LOCUS C* (*FLC*), possesses a cold-sensitive hypervariable structural motif complementary to the *FLC* transcription start site to modulate *FLC* transcription and control flowering time[25]. Lastly, in both animals and plants, the characteristic secondary folds of C/D and H/ACA box snoRNAs serve as dual structural roles, which are indispensable for assembling core snoRNP complexes and forming antisense pairing pockets to target rRNA/snRNA residues for conserved chemical modifications required for ribosome biogenesis; therefore, disruption of intact snoRNA structural motifs triggers aberrant cell fate and disease progression in mammals, as well as severe developmental defects in plant organ development[26–29]. Accordingly, resolving RNA higher-order structures *in vivo* provides higher-order structural blueprints for designing RNA-targeted therapeutic agents and crop improvement strategies.

During the decades of methodological advancement, early classical *in vitro* higher-order structural biology approaches, including X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and single-particle cryo-electron microscopy (cryo-EM), deliver atomic-resolution static RNA conformations by relying on highly purified, homogeneous RNA or ribonucleoprotein samples, yet they suffer from severe throughput limitations and fail to capture the native, dynamic conformational states of RNAs inside living animal and plant cells[30–34]. To overcome these limitations, the experimental resolution of RNA higher-order structures has gradually transitioned toward sequencing-coupled high-throughput experimental probing *in vivo*[35].

In this review, we mainly focus on the higher-order structures of RNA, which refer to spatial organizational forms beyond the level of local base pairing and encompass tertiary and quaternary organization, including long-range base pairing, non-canonical interactions, spatial proximities, three-dimensional (3D) topological conformations, and other conformational ensembles, and then systematically summarize recent high-throughput experimental methods *in vivo* and computational modeling strategies for resolving RNA higher-order structures, as well as their potential application value for investigating RNA regulatory mechanisms in both animal and plant systems. At the end, we also discuss future prospects for artificial intelligence (AI)-driven single-molecule structural detection and multimodal integrated modeling. Such technical innovations establish a solid foundation to unravel RNA structure–function regulatory networks and accelerate transformative applications, including the design of targeted RNA therapeutics and the breeding of stress-resistant crops.

Experimental approaches for resolving RNA higher-order structures by probing *in vivo*

Distinct from secondary structures defined by local pairings, RNA higher-order structures highlight structural coupling between distant sequence regions and the global spatial folding configurations, and consist of intramolecular tertiary architectures and supramolecular quaternary assemblies[5,35]. Tertiary structures arise from long-range intramolecular interactions within individual RNA transcripts, mainly including non-nested long-range base-pairing motifs (represented by pseudoknots), helical stacking, special higher-order nucleic acid folds (represented by RNA triple helices and G-quadruplexes), and other non-canonical tertiary interactions (represented by distal loop-receptor and ribose zipper interactions)[36–38]. RNA quaternary structures represent supramolecular assemblies, such as intermolecular RNA–RNA duplexes and ribonucleoprotein complexes composed of RNAs and RNA-binding proteins (RBP)[39–41]. To date, a series of probing technologies has been developed, such as chemical crosslinking and spatial proximity capture, enabling the inference of long-range interactions, spatial proximities, and local structural constraints within RNA molecules under native conditions. These high-throughput structural probing methodologies have significantly expanded the scale and dimensionality of RNA structural research.

According to the fundamental principles of structural probing, distinct chemical groups on RNA exhibit differential chemical reactivities governed by chemical probes depending on their local conformational surroundings, including base-pairing status of nucleobases, spatial constraint of the RNA 2'-hydroxyl (2'-OH) group[32,42] (Fig. 1a). Therefore, the experimental basis for resolving RNA higher-order structures essentially relies on translating the differential reactivities of nucleotides and riboses within specific structural contexts into measurable molecular signals. High-throughput approaches for capturing RNA higher-order structures can be broadly categorized into two major groups based on their probing principles and obtained structural information. The first group, denoted as duplex capture approaches, in which psoralen derivatives are used as photo-reversible intercalating crosslinkers to form covalent bonds between pyrimidine residues (predominantly U) on opposing strands of RNA duplexes, consists of PARIS/PARIS2 (Psoralen Analysis of RNA Interactions and Structures), SPLASH (Sequencing of Psoralen crosslinked, Ligated and Selected Hybrids), LIGR-seq (Ligation of Interacting RNA followed by High-throughput Sequencing), and COMRADES (Cross-linking of Matched RNAs and Deep Sequencing)[43–47] (Fig. 1b). The second group, referred to as spatial proximity capture approaches, are primarily used to record direct 3D spatial contacts between linearly distant RNA fragments or nucleotides with or without endogenous RBPs participation, thereby providing constraints for RNA 3D folding topology (Fig. 1c–e). These methodologies are mainly represented by SHARC (Spatial 2′-Hydroxyl Acylation Reversible Crosslinking), RIC-seq (RNA *in situ* conformation sequencing), and KARR-seq (Kethoxal-assisted RNA-RNA Interaction Sequencing)[48–50]. In addition, another technology called MARIO (Mapping RNA interactome *in vivo*) first uses UV crosslinking to capture direct RNA–protein contacts, and then relies on formaldehyde and EGS (ethylene glycol bis [succinimidyl succinate])-mediated dual crosslinking to capture protein-assisted intramolecular RNA spatial contacts in living cells[51] (Fig. 1e). All these experimental approaches will be described in detail below.

RNA tertiary structures capture *in vivo*

Among these methodologies developed to resolve RNA tertiary structures, PARIS/PARIS2, SPLASH, LIGR-seq, and COMRADES share a common feature in employing psoralen derivatives as photo-reversible intercalating crosslinkers. Upon irradiation with 365 nm ultraviolet (UV) light, their furan or pyrone rings undergo photochemical reactions with adjacent pyrimidine residues, forming stable covalent cyclobutane bridges between pyrimidine residues on opposing

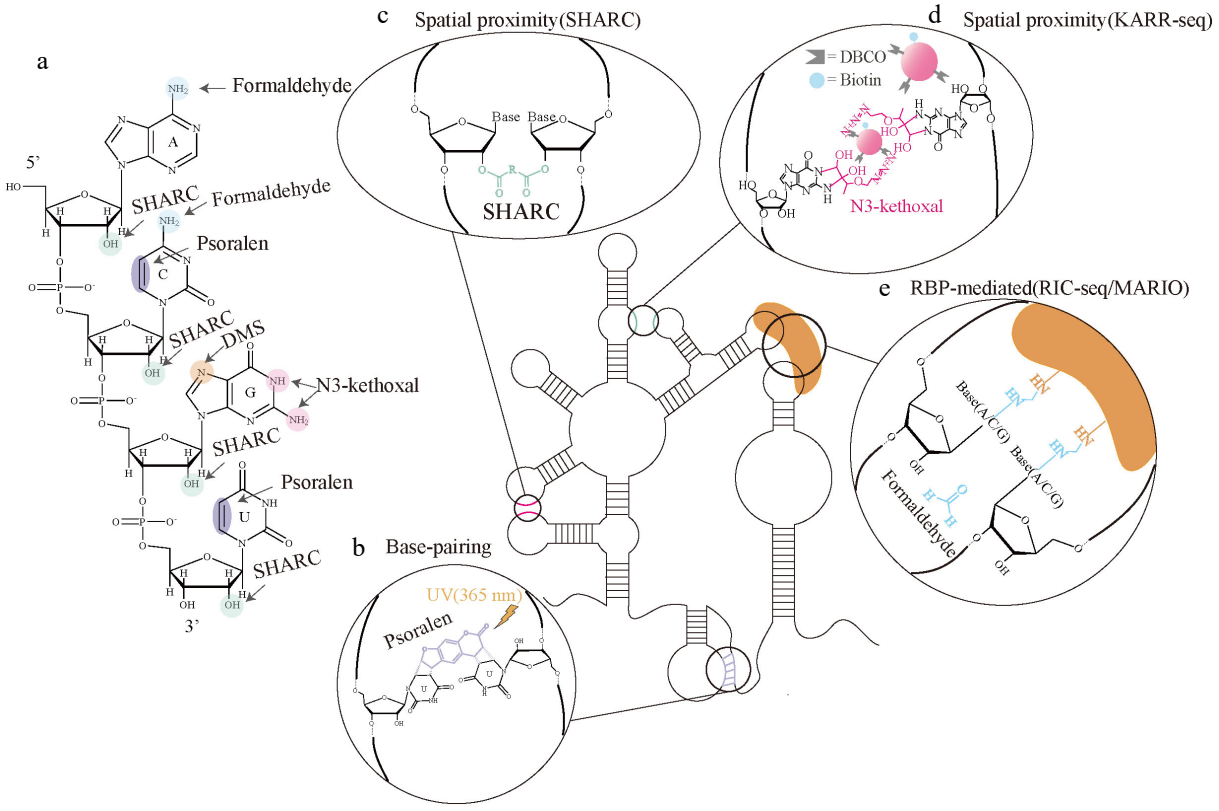


Fig. 1 Experimental approaches to structural information in RNA higher-order structures. (a) Formaldehyde reacts with amino groups of nucleic acid bases (A, C, G) to form cross-links (blue)[49–51]. The SHARC reagent mediates reversible acylation of the RNA 2'-OH (cyan)[48]. Psoralen intercalates into double-stranded regions and induces covalent cross-linking of pyrimidines (U, C) (purple)[43–47]. Dimethyl sulfate (DMS) methylates the N7 position of guanines (G)[52,53]. KARR technology exploits N_3-kethoxal to selectively methylate the N1 and exocyclic N2 positions of unpaired guanine (G)[50]. (b) Psoralen preferentially intercalates into RNA duplex regions and forms covalent cross-links between opposing strands of RNA upon UV (365 nm) activation[43–47]. (c) SHARC reagents react with the 2'-OH groups of ribose on spatially proximal nucleotides, generating reversible cross-links to capture higher-order structural proximity[48]. (d) KARR-seq method labels unpaired G with N3-kethoxal in living cells, then uses DBCO-PAMAM dendrimers and copper-free SPAAC click chemistry to covalently connect spatially close RNA segments[50]. (e) Formaldehyde crosslinking or dual crosslinking (formaldehyde and EGS) stabilizes RBP-bridged spatial RNA–RNA contacts[49,51].

strands of base-paired RNA duplexes; the crosslinked RNA is extracted and subjected to controlled fragmentation, and proximity ligation is then performed under crosslinking constraints. Subsequently, 254 nm UV photoreversal is applied to specifically break psoralen-mediated crosslinks, releasing individual RNA strands while retaining the ligated chimeric junctions. Finally, high-throughput sequencing is conducted, and intramolecular long-range paired segments corresponding to RNA tertiary structures are identified from chimeric reads[43–47] (Fig. 2a).

PARIS/PARIS2

Despite shared core crosslinking chemistry, these techniques differ substantially in RNA enrichment strategies, workflow complexity, and analytical capability. The original PARIS employs 4′-aminomethyltrioxsalen (AMT) psoralen as a photo-reversible intercalating crosslinker. After RNA extraction and controlled fragmentation, crosslinked RNA hybrids are separated from uncrosslinked RNA via native–denatured two-dimensional (ND2D) gel electrophoresis. Subsequently, the purified crosslinked fragments undergo proximity ligation, crosslink reversal, and library construction. However, the workflow suffers from low crosslink recovery, severe sample loss during laborious ND2D gel purification, and relatively low library yield. Later, the updated version PARIS2 was developed; it adopts amotosalen, a more soluble and efficient psoralen derivative instead of AMT, to enhance crosslinking efficiency. Additional technical advancements cover optimized protocols for full recovery of crosslinked RNA, simplified RNase III-mediated RNA fragmentation, denatured–denatured two-dimensional (DD2D) gel purification for crosslinked RNA hybrids, optimized adapter ligation, and reduced photochemical RNA damage. Collectively, these improvements yield over 4,000 times higher experimental efficiency compared with the original PARIS[43–44].

SPLASH

Unlike PARIS/PARIS2, which require laborious two-dimensional gel purification, SPLASH adopts biotin-tagged psoralen combined with streptavidin bead enrichment, which substantially simplifies experimental workflows, reducing manual operation and sample loss. Briefly, biotin-conjugated psoralen (e.g., psoralen-PEG_3-biotin) is delivered into live cells with a mild permeabilizing agent such as digitonin and irradiated at 365 nm, so that psoralen intercalates at helical regions and forms covalent cyclobutane adducts with pyrimidines on opposite strands, capturing both intramolecular and intermolecular base-paired duplexes at their site of contact. The biotin tag then allows reacted fragments to be recovered directly on streptavidin magnetic beads, circumventing the size-based gel fractionation that limits the throughput and recovery of PARIS. After RNase digestion and proximity ligation, the chimeric products are reverse-crosslinked and converted into sequencing libraries, yielding genome-wide maps of RNA duplexes[45]. This methodology also supports unbiased transcriptome-wide profiling.

https://doi.org/10.48130/gcomm-0026-0022

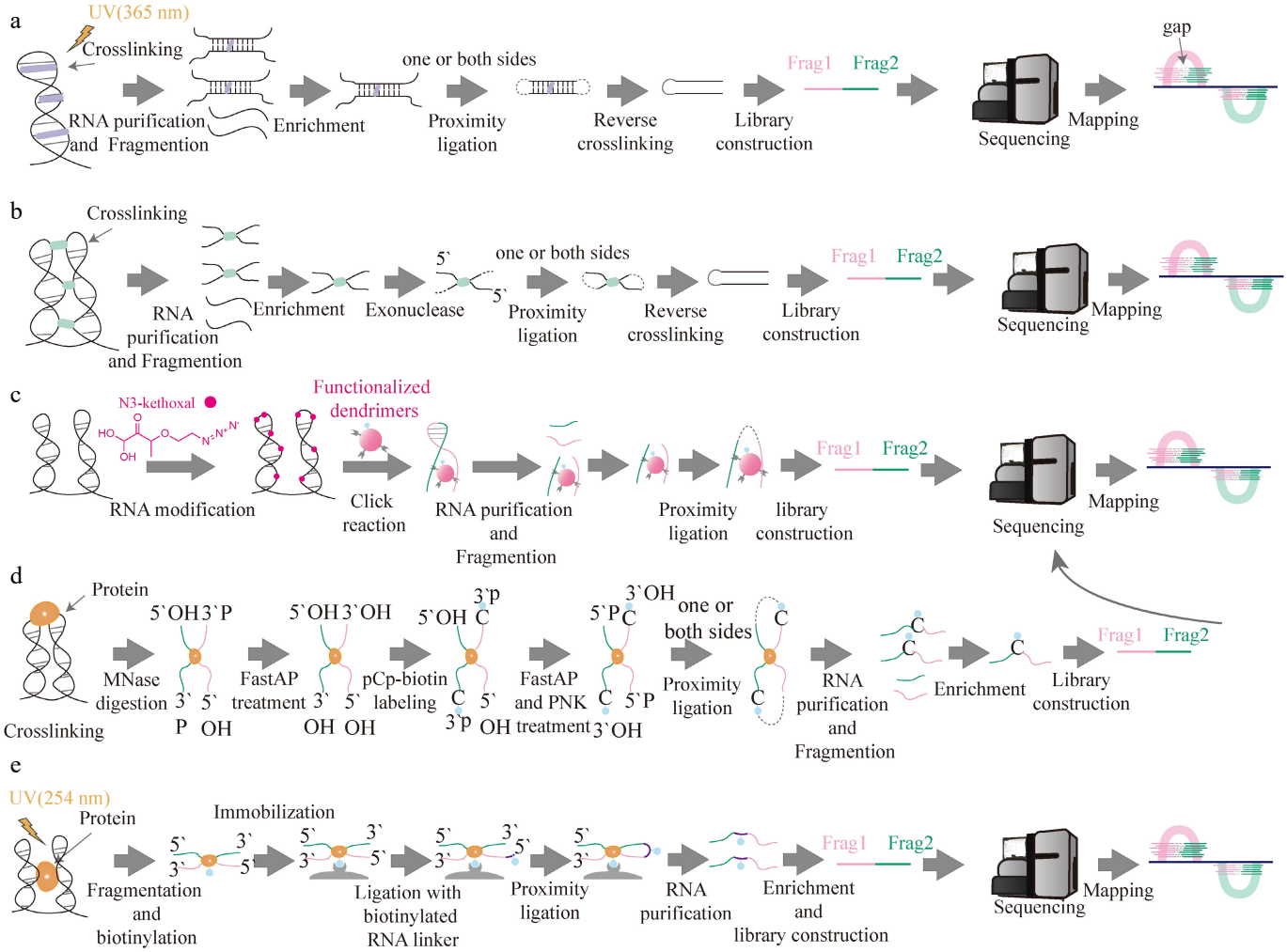


Fig. 2 High-throughput experimental methods for resolving RNA higher-order structures. (a) Approaches for capturing RNA tertiary structures[43–47]. RNA molecules are crosslinked *in vivo* using 365 nm UV-activated psoralen. After fragmentation and enrichment of crosslinked RNA fragments, proximity ligation is performed, followed by crosslink reversal. Ligated RNA fragments are constructed into sequencing libraries to map base-paired duplex regions. (b) SHARC[48]: cellular RNA is subjected to bifunctional reversible acylating crosslinking. After crosslink formation, RNA undergoes purification and enrichment, followed by exonuclease trimming of RNA termini. Proximity ligation is carried out prior to alkaline-triggered crosslink hydrolysis. Libraries built from chimeric fragments identify spatially proximal RNA regions independent of canonical base pairing. (c) KARR-seq[50]: cells are incubated with membrane-permeable N_3-kethoxal to label unpaired G residues. Spatially proximal RNA segments are crosslinked via DBCO-conjugated PAMAM dendrimers through SPAAC click chemistry. After RNA fragmentation, proximity ligation and purification, chimeric RNA templates are enriched for library construction to detect protein-independent spatial contacts between RNA regions. (d) RIC-seq[49]: RNA–protein complexes are fixed via formaldehyde crosslinking in situ, and cells are permeabilized. Accessible RNA regions are mildly digested with MNase. Samples are sequentially treated with FastAP (Thermosensitive Alkaline Phosphatase) for dephosphorylation, pCp-biotin labeling of RNA 3' termini, and additional FastAP and PNK (T4 polynucleotide kinase) treatments for end repair. Proximity ligation proceeds inside fixed cells. Purification, fragmentation, and enrichment of chimeric RNA are implemented prior to library construction to capture RBP-mediated RNA–RNA spatial contacts. (e) First, UV crosslinking is used to capture direct RNA-protein contacts; MARIO[51] then relies on dual crosslinking (formaldehyde and EGS) to trap RBP-bridged RNA spatial contacts, including intramolecular or intermolecular RNA interactions. Distinctly, proximity ligation is carried out post-cell lysis. Subsequent enrichment and library construction follow analogous workflows to decipher RNA tertiary and quaternary interactions.

LIGR-seq

LIGR-seq uses the psoralen derivative AMT, which intercalates into RNA duplexes and, upon 365 nm UV irradiation, generates reversible interstrand adducts between juxtaposed pyrimidine bases, thereby fixing base-paired regions in their native context within living cells. After cell lysis, the RNA is subjected to limited single-strand fragmentation by S1 endonuclease followed by proximity ligation with circRNA ligase, so that the ends of duplexes held in spatial proximity are joined into chimeric molecules. The key enrichment step is digestion with RNase R, a 3' to 5' exoribonuclease that degrades linear and structured but not crosslinked RNA, thereby removing uncrosslinked RNAs and markedly reducing background. Crosslinks are then reversed by 254 nm irradiation, and the recovered chimeras are converted into sequencing libraries, yielding genome-wide maps of RNA–RNA duplex interactions. Similar to PARIS/PARIS2 and SPLASH, LIGR-seq supports unbiased transcriptome-wide profiling[46].

COMRADES

Distinctively, COMRADES employs azide-modified psoralen derivatives as photo-reversible intercalators and relies on antisense probe pull-down to enrich individual RNA transcripts of interest rather than capturing all cellular RNA duplexes globally, followed by click-chemistry-based secondary enrichment of crosslinked hybrids. Benefiting from paired experimental and control libraries, COMRADES

effectively eliminates artifacts arising from *in vitro* RNA annealing and ligation, making it uniquely suitable for dissecting intramolecular long-range base-pairing architectures and alternative conformations within long RNA molecules[47].

Dimethyl sulfate (DMS) enables identification of RNA tertiary interactions

DMS has long served as a classic chemical probe exclusively used to map RNA secondary structures. Its canonical reactivity targets the Watson–Crick N1 site of unpaired adenines and the N3 site of unpaired cytosines, revealing canonical base-pairing status to resolve stems, loops, and other secondary folds. Recent technical advances demonstrate that DMS can also methylate the N7 position of guanines, which lies on the Hoogsteen face of guanine bases. N7-G modification signals precisely report noncanonical Hoogsteen hydrogen bonds, which are the core intermolecular forces sustaining complex RNA tertiary architectures such as RNA G-quadruplexes, pseudoknots, and multi-helical junctions. Because the methyl group of m^7G is positioned on the Hoogsteen rather than the Watson–Crick face, it only minimally perturbs reverse transcription, and N7-G is therefore invisible to conventional DMS-MaP. Two strategies have been developed to overcome this limitation. BASH MaP (Borohydride Assisted Structure determination of N7G Hoogsteen interactions through Misincorporation Profiling) retains DMS modification as the starting point but renders N7-G detectable through chemical conversion. m^7G is reduced with potassium borohydride and then converted into an apurinic/apyrimidinic (AP) site by heating in acidic buffer, and the resulting AP site is detected as a reverse-transcription stop, thereby amplifying an otherwise silent signal, coupled with the DAGGER (Deconvolution Assisted by N7-G Gaps Enabled by Reduction and depurination) pipeline; this readout resolves alternative tertiary conformations, including G-quadruplex folding states and dynamics, at the single-molecule level[52]. More recently, msDMS-MaP (multi-site DMS-MaP) achieves the same goal with a simpler workflow, in which optimized reverse transcription decodes the typically invisible N7-G modification directly through a tautomer-induced mutational signature, concurrently with conventional N1 and N3 modifications, requiring only minor changes to established MaP protocols. In this way, a single DMS probing experiment now yields both N1/N3 reactivity profiles and N7 protection profiles, furnishing single-nucleotide constraints to reconstruct diverse RNA tertiary structures in living cells[53].

Spatial proximity capture for RNA quaternary structure *in vivo*

Spatial proximity capture technologies designed to resolve RNA quaternary structures employ crosslinking strategies to trap RNA molecules that are spatially adjacent within cells. These methods capture intermolecular RNA contacts stabilized either by synthetic bifunctional crosslinkers (SHARC, KARR-seq) or endogenous RBP scaffolds (RIC-seq, MARIO)[48–51]. Following fragmentation, spatially proximal RNA regions undergo proximity ligation to produce chimeric RNA templates for high-throughput sequencing. The resulting chimeric reads uncover intermolecular RNA–RNA interactions and higher-order ribonucleoprotein complexes that define RNA quaternary structures. Unlike RNA tertiary structure capture technologies restricted to intramolecular base-paired tertiary helices, these methods capture supramolecular assemblies formed between discrete RNA molecules and associated RBPs. Nevertheless, they differ substantially in crosslinking chemistry and dependence on RBP scaffolds.

SHARC

SHARC employs dipicolinic acid imidazolide (DPI), a bifunctional reversible acylating crosslinker, to crosslink spatially adjacent 2′-OH groups of RNA ribose unassisted by proteins in living cells. After crosslink formation, RNA undergoes fragmentation and exonuclease trimming, which stops digestion at crosslink sites to pinpoint interacting nucleotides. Proximity ligation is performed while crosslinks remain intact, followed by mild alkaline hydrolysis to reverse ester crosslinks without degrading RNA. High-throughput sequencing of chimeric reads captures intermolecular spatial contacts, non-canonical interactions, and conformational heterogeneity. Furthermore, the defined linker length of SHARC crosslinkers provides spatial distance constraints, supporting high-resolution computational modeling of RNA quaternary structure (Fig. 2b). Notably, it has been successfully used for identifying compact folding of the 7SK RNA, a critical regulator of transcriptional elongation[42,48].

KARR-seq

KARR-seq technology exploits N_3-kethoxal, a cell-permeable azide-modified kethoxal derivative, to selectively label unpaired guanine residues of RNA in living cells. Subsequent crosslinking is achieved by dibenzocyclooctyne (DBCO)-conjugated poly(amidoamine) dendrimers (PAMAM) through strain-promoted azide–alkyne cycloaddition (SPAAC), a copper-free bioorthogonal click reaction that covalently bridges spatially proximal RNA segments independent of canonical base pairing or RBP scaffolds. After RNA fragmentation, crosslinked RNA assemblies are enriched, followed by proximity ligation. High-throughput sequencing of chimeric reads identifies intermolecular RNA–RNA assemblies constituting RNA supramolecular assemblies. Overall, distinct from SHARC, KARR-seq labels single-stranded guanines (G) without protein scaffolds and adopts tunable-size dendrimer crosslinkers to capture RNA contacts across variable spatial distances, offering flexible detection for diverse RNA supramolecular assemblies[50] (Fig. 2c).

RIC-seq

RIC-seq employs formaldehyde as a crosslinker to stabilize RBP-tethered spatial RNA contacts *in vivo*. Following cell permeabilization and mild RNA fragmentation treatment by micrococcal nuclease (MNase), biotinylated cytidine bisphosphate (pCp-biotin) is used to label the 3′ termini of RNA fragments. A key advantage of this technique is that proximity ligation is conducted *in situ* inside fixed cells rather than after cell lysis. This process produces chimeric RNA molecules derived from spatially adjacent RNA segments bridged by shared RBPs. Subsequently, after enrichment and purification of biotin-tagged chimeric RNAs, strand-specific sequencing libraries are constructed for high-throughput sequencing. Analysis of chimeric reads identifies intermolecular RNA assemblies corresponding to RNA quaternary architectures[49] (Fig. 2d).

Unlike SHARC and KARR-seq, which rely on synthetic chemical crosslinkers to trap RNA contacts without RBP scaffolds, the capture capacity of RIC-seq strictly relies on endogenous RBPs. Hence, RIC-seq generally reflects functional RNA organizational states more closely, rather than merely random spatial proximity. However, RIC-seq cannot distinguish direct RNA–RNA spatial contacts from indirect interactions separated by intervening RBPs.

MARIO

In addition to UV crosslinking, which captures direct RNA–protein contacts, MARIO employs dual crosslinking with formaldehyde and EGS to stabilize RBP-bridged spatial contacts between RNA fragments inside cells. After cell lysis and mild RNA fragmentation, crosslinked RBP complexes are biotin-labeled and enriched on streptavidin magnetic

https://doi.org/10.48130/gcomm-0026-0022

beads, an enrichment strategy that distinguishes MARIO from RIC-seq, which instead marks RNA 3' termini. Proximity ligation is then performed on the immobilized RNP complexes to generate chimeric RNA molecules, in which RNA fragments brought into close 3D proximity by shared RBPs are joined. High-throughput sequencing of these chimeric reads identifies intramolecular spatially adjacent RNA segments, which correspond to protein-stabilized, non-base-paired RNA tertiary structures. Because MARIO reports the intramolecular tertiary structure of individual transcripts alongside intermolecular RNA–RNA complexes, it bridges the tertiary- and quaternary-structure categories. However, MARIO cannot discriminate direct spatial interactions from indirect contacts mediated by protein scaffolds, and because ligation is performed after cell lysis, the native spatial context may be disrupted, increasing the risk of capturing spurious RNA contacts[51] (Fig. 2e).

Computational modeling empowers the reconstruction of RNA higher-order structures

Over the past decades, with the advancements of experimental RNA higher-order structure determination technologies and high-throughput sequencing, computational modeling of RNA higher-order structures has gradually evolved from physical simulation approaches to data-driven predictive models. First, the earliest *ab initio* RNA folding algorithms simulated RNA conformational folding based on thermodynamic rules and atomic interaction physics. Later, as experimentally solved RNA structures accumulated, template-based modeling strategies gradually emerged, bridging physics-based and data-driven workflows. Mostly recently, the rise of deep learning has fundamentally transformed the RNA structure modeling paradigm and established the core data-driven technical system[54–57] (Fig. 3).

Ab initio folding strategies

Classical *ab initio* folding refers to template-free computational pipelines that reconstruct RNA tertiary and quaternary architectures purely from primary RNA sequences, guided by physical force fields or statistical energy potentials, without relying on solved homologous RNA structural backbones. *Ab initio* methods simulate RNA folding following physical laws and can capture dynamic folding features such as energy landscapes. These models describe RNA molecules at the atomic scale and apply intermolecular forces to mimic physiological environments. To search all possible conformational space, researchers employ sampling algorithms including Monte Carlo (MC) and molecular dynamics sampling (Fig. 3a). We summarize some representative coarse-grained *ab initio* simulators and their core characteristics, as well as their compatibility with the *in vivo* crosslinking sequencing experimental methods introduced above.

NAST (Nucleic Acid Simulation Tool) is the simplest single-particle *ab initio* simulator. Each nucleotide is represented by only one particle marking the backbone position, omitting fine atomic details of base stacking and base pairing. It extracts statistical potentials from ribosome crystal structures to perform MC conformational sampling. Starting from fully linear, unfolded RNA chains, NAST rapidly sketches the global backbone layout and eliminates sterically clashing conformations via simple distance-based energy penalties[58]. However, it only delivers rough backbone outlines and is designed for large RNA molecules, making it tractable for RNAs that are computationally prohibitive for all-atom methods. In the context of structural restraint integration, it is mainly compatible with duplex capture restraints from PARIS and SPLASH and cannot process high-precision spatial proximity data from SHARC, RIC-seq, or KARR-seq.

iFoldRNA/iFoldRNA v2 relies on discrete molecular dynamics (DMD) replica-exchange sampling and adopts a three-bead coarse-grained system, with independent energy terms for backbone torsion, canonical base pairing, base stacking, and hydrophobic interactions. The original iFoldRNA only supports RNA sequences shorter than 50 nt. The upgraded iFoldRNA v2 adds Root Mean Square Deviation (RMSD)-based trajectory clustering after simulation to extract representative low-energy conformations[59,60]. Its computational cost surges sharply for RNAs longer than 100 nt, making it unsuitable for long RNAs. It is compatible with duplex capture restraints from psoralen-mediated technologies to constrain intramolecular long-range base-paired helices.

Ernwin adopts Markov Chain Monte Carlo (MCMC) sampling and includes steric repulsion potentials plus knowledge-based energy terms for RNA secondary structural motifs. Instead of sampling single nucleotides individually, it first arranges the spatial positions of complete helical segments and fills intervening single-stranded loops with flexible distance restraints. A multi-scale pipeline can be coupled with SPlit-and-conQueR (SPQR) to convert low-resolution coarse skeletons into full-atom refined structures. This tool outputs only low-resolution global topology and struggles to accurately capture complex non-canonical tertiary motifs such as pseudoknots and A-minor interactions[61].

BARNACLE (Bayesian network model of RNA using circular distributions and maximum likelihood estimation) is based on dynamic Bayesian networks. It models seven key backbone dihedral angles of RNA via mixed circular probability distributions and employs hidden Markov chains to capture sequential conformational correlations along RNA strands. It conducts continuous conformational sampling without prebuilt discrete fragment libraries and excels at depicting flexible torsion landscapes of single-stranded loops[62]. Combined with SHARC's nucleotide-level distance constraints, it can analyze heterogeneous dynamic loop conformations captured by *in vivo* sequencing, yet it lacks complete tertiary interaction potentials and cannot independently complete full-length RNA 3D reconstruction[62].

OxRNA is a coarse-grained model in which each nucleotide is treated as a single rigid body with five interaction sites (backbone, hydrogen-bonding, cross-stacking, and 3'-and 5'-stacking) specifically optimized for RNA thermodynamic folding landscapes. It integrates Virtual Move Monte Carlo (VCMC) and umbrella sampling algorithms, equipped with temperature-dependent free energy potentials that separately quantify short-range nearest-neighbor interactions and long-distance inter-nucleotide spatial forces. It can fully simulate RNA thermal unfolding and refolding and output complete thermodynamic folding landscapes, making it ideal for studying temperature-sensitive RNA elements such as RNA thermoswitches and G-quadruplexes. It supports distance restraints from PARIS and SHARC to calibrate thermal folding pathways, but raw contact matrices generated by RIC-seq and MARIO must be manually converted into standardized topology files before input[55,63].

SimRNA and SimRNAweb v2.0 share the same five-bead coarse-grained simulation core and employ asymmetric Metropolis Monte Carlo sampling to search thermodynamically favorable RNA 3D conformations. Their multi-layer scoring function constrains intrinsic RNA geometric features, including bond lengths and backbone twists, while describing long-range spatial contacts between bases and RNA backbones. After simulation, all generated structures are

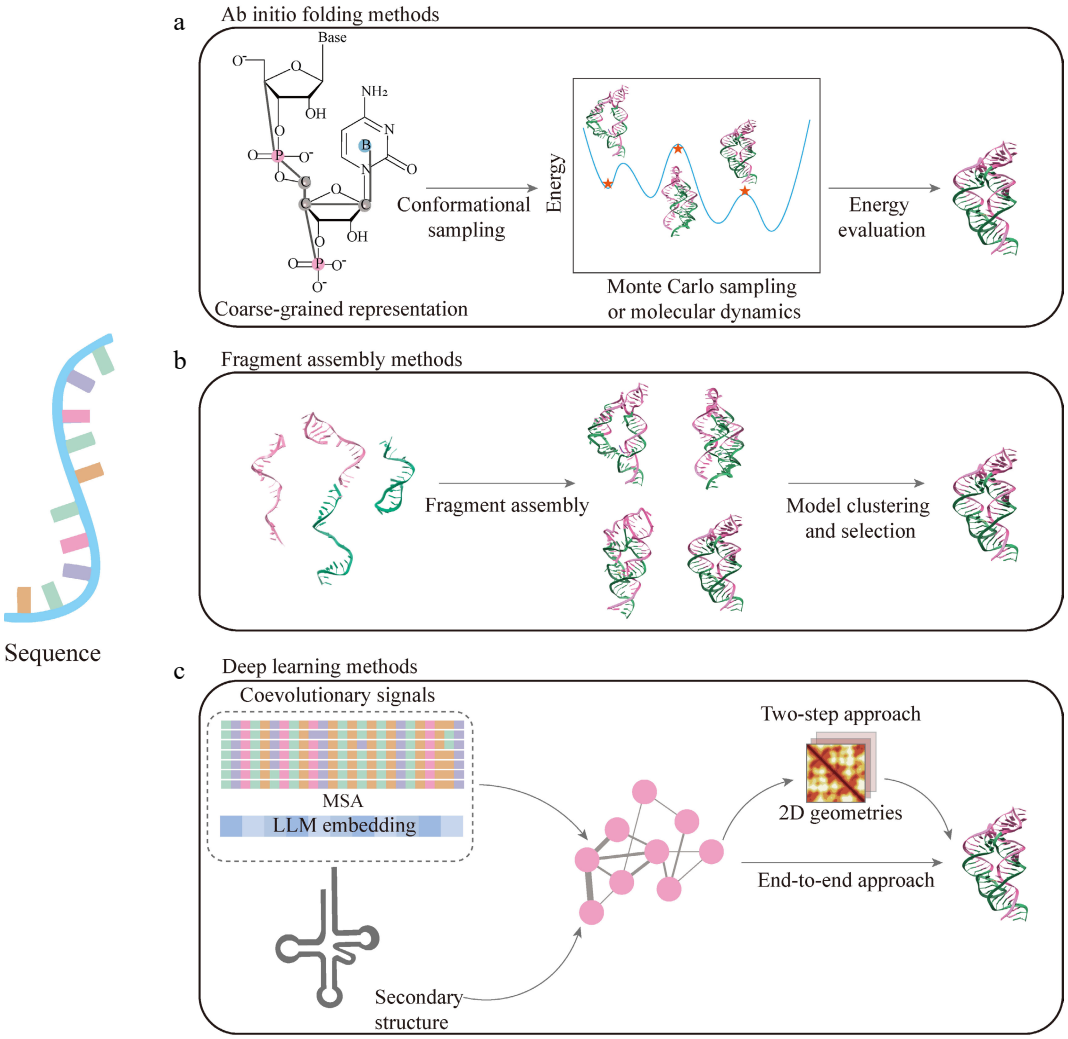


Fig. 3 Computational methods for RNA tertiary structure modeling. (a) *Ab initio* folding methods[58–72]. RNA molecules are represented using coarse-grained models, followed by exploration of the conformational space through sampling of the energy landscape, typically using Monte Carlo or molecular dynamics approaches. Candidate conformations are evaluated using energy functions to identify low-energy RNA structures. (b) Fragment assembly methods[75–81]. Structural fragments are selected from libraries derived from known RNA structures and assembled into full-length models. The resulting candidate structures are subsequently clustered and evaluated to identify representative conformations. (c) Deep learning methods[85–90]. Deep learning approaches integrate co-evolutionary signals and secondary structure information as inputs, and use neural network architectures to predict RNA structures. These methods generally follow either a two-step strategy, in which intermediate geometric constraints are first predicted and then used to reconstruct three-dimensional structures, or an end-to-end strategy that directly predicts RNA three-dimensional conformations from sequence-derived features.

clustered by RMSD, and the central conformation of the lowest energy cluster is output as the most native like candidate structure. SimRNAweb v2.0 is an accessible online graphical server built upon the SimRNA core, with greatly improved compatibility with *in vivo* crosslinking sequencing data[64,65]. It directly imports chimeric contact matrices from RIC-seq and MARIO without extra format conversion, and parallelized MC sampling drastically cuts the runtime for long RNA transcripts. A dedicated multi-conformation ensemble output module is integrated to resolve heterogeneous dynamic RNA folds captured by all spatial proximity capture methods (SHARC, KARR-seq, RIC-seq, and MARIO), with built-in interactive trajectory visualization and energy landscape analysis tools for intuitive interpretation of RNA folding heterogeneity. It is a widely used platform for integrating multi-type *in vivo* structural restraints for RNA modeling.

IsRNA, IsRNA1/2 (Iterative Simulated Reference state coarse-grained RNA model), and **IsRNAcirc** (Iterative Simulated Reference state model for circular RNA) were developed by the same research group. All of which use a coarse-grained framework with five beads per nucleotide and adopt Replica-Exchange Molecular Dynamics (REMD) sampling, with high capacity to model diverse non-canonical base-pairing tertiary interactions. The foundational IsRNA established the iterative simulated reference state theory for correlated intra-RNA interactions, underpinning IsRNA1 and IsRNA2. IsRNA1 only includes potentials for canonical Watson–Crick pairs and fits simple stem-loop RNAs, while IsRNA2 extends the force field with terms for A-minor, Hoogsteen, and multi-base tertiary contacts, greatly improving predictions of complex junctions and pseudoknots. Both initiate simulations from fully linear unfolded RNA to produce stable tertiary conformational ensembles. Meanwhile, IsRNAcirc is a specialized *ab initio* folding tool exclusively designed for circular RNAs, which adds an extra cyclic backbone continuity energy penalty to eliminate artificial ring-opening or structural overlapping artifacts during REMD sampling. It is fully compatible with

https://doi.org/10.48130/gcomm-0026-0022

distance restraints derived from PARIS, MARIO, and RIC-seq to constrain internal tertiary loops of circular long non-coding RNAs and directly outputs closed-loop, full-length circRNA tertiary ensembles[66–69].

HiRE-RNA is REMD-based sampling with geometry parameters trained on hundreds of high-resolution RNA crystal structures. Its energy function specifically optimizes multi-base stacking and non-Watson–Crick pairing potentials, ideal for ribozymes and riboswitches with intricate tertiary motifs. When paired with high-precision distance restraints from SHARC, it accurately reconstructs subtle local non-canonical interactions[70].

RNAJP (RNA Junction Prediction dual-resolution coarse-grained model) combines fine per-nucleotide sampling and global helix topology search. Its energy function contains 12 independent terms covering covalent geometric constraints, three classes of base pairing (Watson–Crick, Hoogsteen, sugar-edge), and long-range helical stacking. It is specially optimized to predict tertiary structures of multi-branched RNA junctions, which are hard to resolve via other tools. Restraints from PARIS and MARIO are commonly imported to correct long-range spatial contacts within junction regions and compensate for missing tertiary topological information derived purely from RNA sequences[71].

TiRNA (Temperature and Ion-dependent coarse-grained RNA folding model) represents the first *ab initio* RNA simulator that integrates electrostatic potentials dependent on physiological Mg^{2+}/Na^{+} ion concentrations and temperature into its force field; it combines VCMC and umbrella sampling to accurately simulate ion-stabilized pseudoknots and G-quadruplex tertiary folds, and natively supports pairing-agnostic universal 2′-OH crosslink restraints acquired from SHARC and KARR-seq to calibrate long-range spatial proximities of RNAs inside living cells[72].

Template-based strategies

Template-based modeling reconstructs RNA tertiary structures by reusing experimentally resolved homologous RNA backbones or conserved structural motifs, relying on evolutionary and spatial structural conservation across RNA families. Unlike *ab initio* folding that starts purely from raw sequences, template-based methods significantly narrow the conformational search space, reducing computational cost and improving modeling robustness. The general workflow consists of three core steps: sequence or structural motif alignment, homologous fragment replacement and assembly, followed by geometric energy refinement to eliminate steric clashes and irrational torsion angles (Fig. 3b).

Template-based tools can readily integrate long-range spatial restraints captured by *in vivo* crosslink sequencing (psoralen-mediated duplex capture methods, COMRADES, MARIO, RIC-seq, SHARC, KARR-seq) as auxiliary energy penalties during structural optimization. These experimental contact constraints correct deviations caused by incomplete template homology, fix misfolded long-distance helical arrangements, and refine the spatial topology of multiway junctions, pseudoknots, and flexible loop regions that are poorly resolved by template information alone. Representative template-dependent modeling tools are detailed as follows.

ModeRNA is a classic comparative homology modeling tool specialized for conserved RNA families. It first performs global and local sequence alignment between the target RNA and homologous template structures, then replaces divergent sequence fragments while preserving the conserved backbone scaffold of the template. Geometric constraints and energy minimization are applied to maintain continuous, stereochemically reasonable RNA backbones. It mainly accepts duplex pairing restraints from psoralen-based duplex capture technologies (PARIS, SPLASH) to correct global helical layouts, yet lacks native support for protein-mediated spatial contacts detected by MARIO or RIC-seq. It performs best for highly conserved ribosomal RNAs, snoRNAs, and small structured regulatory RNAs[73].

RNABuilder leverages multi-resolution modelling (MRM) and multibody dynamics simulation for RNA construction. Based on template-residue alignment and spatial constraint assignment, its force field incorporates tertiary base-pair interactions. Distinct from motif library fragment assembly pipelines, it models template-missing segments like short linkers by combining coarse-grained conformational sampling and full-atom refinement. The tool accepts psoralen-derived duplex restraints from PARIS, SPLASH, and COMRADES to restrict global helical topology. Distance matrices from SHARC, KARR-seq MARIO, and RIC-seq can also be imported as tertiary restraints after standardized format conversion[74].

Vfold series (Vfold3D, VfoldLA, and Vfold-Pipeline). The Vfold toolkit integrates template fragment assembly with statistical physical potentials. Vfold3D builds coarse-grained RNA scaffolds based on predicted secondary structures, retrieving matching helix and junction fragments from pre-built template libraries to assemble full-length atomic models. VfoldLA expands loop and multiway junction template libraries to handle complex topologies containing pseudoknots. The complete Vfold-Pipeline automates end-to-end secondary structure prediction, fragment matching, and 3D refinement[75].

These series fully support all classes of *in vivo* crosslink restraints. PARIS duplex pairing constraints stabilize long-range intramolecular helices, while MARIO/RIC-seq protein-bridged spatial contacts and SHARC/KARR-seq protein-independent 2′-OH proximity signals are incorporated as distance penalties during energy optimization, greatly improving modeling accuracy for lncRNAs with intricate tertiary junctions.

FebRNA is an independent fragment ensemble-based template assembly tool for RNA 3D-structure reconstruction. Starting from predicted or user-provided secondary structure topology, it assembles coarse-grained fragment ensembles and rebuilds all-atom models[76]. It performs well for pseudoknots and multiway junctions, and can incorporate experimental distance restraints from PARIS, MARIO, RIC-seq, SHARC, and KARR-seq during structural optimization.

RNAComposer maps each predicted RNA secondary structural element (stems, loops, and junctions) to pre-solved 3D motif fragments stored in the FRABASE database, then assembles individual fragments via fragment-based alignment and global geometric optimization. It delivers fast automated all-atom models when reliable secondary structures and matching template motifs are available[77]. It can import distance restraints derived from PARIS and SHARC to adjust the spatial arrangement of helical domains, yet struggles to accurately model large conformational rearrangements captured by RIC-seq and MARIO due to its rigid fragment assembly logic.

3dRNA adopts modular fragment assembly coupled with a dedicated scoring function, 3dRNAScore, which natively incorporates base-pairing and experimental contact restraints during structural assembly and refinement. It supports modeling both linear and circular RNAs by extending cyclic backbone geometric limits[78]. It is compatible with PARIS, MARIO, and RIC-seq crosslink matrices to correct the tertiary topology of circular lncRNA internal loops, compensating for missing template information in non-conserved loop regions.

FARNA and its upgraded version **FARFAR2** belong to Rosetta-based template-assisted fragment assembly tools. They extract short structural fragments from known RNA crystal libraries to assemble full-length conformations, with multi-stage all-atom energy minimization for structural refinement. FARFAR2 optimizes sampling strategies for complex global folds including pseudoknots and multi-helical junctions[79,80]. Both tools accept duplex and spatial proximity restraints from all mainstream *in vivo* probing methods. SHARC high-resolution 2′-OH distance constraints are particularly effective for refining subtle non-canonical tertiary contacts unrepresented in template fragment libraries.

MC-Sym assembles RNA 3D architectures from a comprehensive library of precomputed nucleotide and helical fragments, guided by input secondary structure topologies. It relies on MC fragment rearrangement to optimize assembled conformations and resolve steric conflicts[81]. It can integrate PARIS base-pair restraints to fix global duplex frameworks, but requires manual format conversion to load contact matrices from RIC-seq and MARIO, limiting its convenience for large-scale transcriptome-wide modeling workflows.

RNAthreader is a recently developed threading tool for RNA template identification and target template alignment. It combines secondary structure topological diagrams with sequence profiles to improve template matching, especially for RNAs containing pseudoknots[82]. It generates residue-level alignment outputs for downstream tertiary structure refinement. Although it does not perform all-atom reconstruction internally, alignment results can be combined with distance restraints from *in vivo* cross-link sequencing after format conversion.

Deep learning-based algorithms

In recent years, artificial intelligence (AI) has delivered transformative breakthroughs. As a core subset of modern AI, deep learning-based algorithms utilize large language model (LLM) embeddings and evolutionary multiple sequence alignment (MSA) signals to infer RNA tertiary and quaternary structures without handcrafted physical potentials required by classical computational tools. Covering two-step restraint reconstruction, end-to-end coordinate prediction, and generative ensemble modeling, such methods enable fast, template-free RNA monomer and complex structure prediction, whose outputs can be further refined or filtered with experimental RNA interaction restraints and classical energy minimization workflows[56] (Fig. 3c).

Restraint reconstruction RNA prediction architectures

Inspired by AlphaFold2's[83] transformer-based distance prediction and trRosetta-style energy minimization[84], early AI-driven RNA 3D-structure prediction adopted a two-step, restraint-first paradigm. trRosettaRNA[85], DeepFoldRNA[86], and epRNA[87] are typical representatives of this workflow. Specifically, trRosettaRNA employs transformer architectures to predict inter-residue distances and spatial orientations from MSAs and predicted secondary structures, followed by Rosetta energy minimization to assemble complete all-atom RNA conformations. DeepFoldRNA utilizes deep self-attention networks to generate *de novo* RNA geometric potentials, adopts a three-bead coarse-grained nucleotide representation, and applies iterative gradient-based energy optimization to reconstruct full tertiary structures from predicted restraint maps; it also introduces self-training strategies to compensate for the scarcity of experimentally resolved RNA structures during model training. By contrast, epRNA builds rotation- and translation-invariant convolutional networks based on Euclidean parametrization, which directly predict global distance matrices merely from raw RNA sequences to provide fundamental restraints for downstream structural folding.

End-to-end direct atomic coordinate prediction

Two-step restraint-reconstruction approaches suffer from inherent limitations, despite their wide applicability. Simplified one-dimensional (1D) and 2D training objectives often lead to suboptimal structural optimization, while the iterative energy minimization process substantially slows down inference speed. To address these bottlenecks, the RNA structure prediction field has progressively adopted full end-to-end architectures derived from AlphaFold2, which directly predict all atom 3D coordinates without explicit restraint generation and post-processing optimization. Representative end-to-end models include RhoFold+[88] and NuFold[89]. RhoFold+ further incorporates pre-trained RNA language models to compensate for sparse coevolutionary signals in MSAs, enabling robust prediction with or without homologous sequence information. NuFold follows a classic Evoformer-based architecture with nucleobase-centered nucleotide representation, effectively capturing fine-scale local RNA geometric features. Different from pure end-to-end pipelines that may produce structurally unreasonable conformations, DRfold and its upgraded version DRfold2[90] adopt a hybrid modeling strategy. These methods integrate end-to-end learned geometric potentials with auxiliary restraint information, balancing prediction accuracy, inference efficiency, and physicochemical plausibility of the final RNA tertiary structures.

Generative ensemble prediction model

Beyond those discriminative end-to-end frameworks, RNAbpFlow[91] advances this field as a novel SE (3) (Special Euclidean group)-equivariant flow matching generative model that learns the full distribution of feasible RNA conformations. It only requires raw sequences and base-pairing information as conditional inputs, completely eliminating the dependence on MSA or homologous templates. By adding base-pairing guidance during training, it ensures that both canonical and non-canonical base pairs form properly at every step of structure building, and its intrinsic generative workflow eliminates the requirement for computationally expensive subsequent geometric refinement. Additionally, DynaRNA[92] is a diffusion-based generative computational model for generating dynamic RNA tertiary conformational ensembles. Powered by a denoising diffusion probabilistic model coupled with equivariant graph neural networks, it samples diverse 3D RNA conformations without MSA input and captures rare excited conformational states.

Macromolecular quaternary complex predictors

Beyond RNA tertiary structure prediction, several deep-learning tools enable modeling of quaternary assemblies, including RNA–RNA dimers and RNA–protein complexes. As an early specialized framework, RoseTTAFoldNA[93] extends the RoseTTAFold architecture to handle nucleic acid-containing assemblies; it leverages multiple sequence alignment information to jointly predict RNA tertiary folds and protein–RNA complex structures, producing all-atom models with built-in confidence metrics. AlphaFold 3[94] further expands complex structure prediction into a unified general-purpose framework, supporting joint modeling of proteins, RNA, DNA, small-molecule ligands, and ions within one system. Despite its broad molecular coverage, its performance on RNA-containing quaternary assemblies remains limited by the scarcity of experimental RNA-complex training data. To mitigate the reliance on evolutionary covariation signals required by the above two tools, ProRNA3D-single[95] is developed for protein–RNA complex prediction using only individual input sequences. It pairs pre-trained biological language models via geometric attention modules to infer intermolecular interaction patterns and optimizes the relative orientation between protein and

https://doi.org/10.48130/gcomm-0026-0022

RNA chains, delivering competitive predictions even in the absence of paired protein-RNA MSAs.

Benchmarking and integration of experimental restraints across modeling strategies

RNA 3D prediction has long been evaluated on community-wide benchmarks, for example, CASP15 (Critical Assessment of Structure Prediction, the first CASP round to include RNA targets)[96] and RNA-Puzzles (a community-wide blind RNA structure prediction challenge)[97], on which all three strategy classes are assessed on identical, unpublished targets using shared metrics such as RMSD, TM-score (a measure of global topological similarity)[98] and lDDT (local Distance Difference Test, which assesses local interatomic distances independently of superposition)[99]. On these benchmarks, *ab initio* methods recover correct global topology but have limited local accuracy; template-based methods achieve the lowest RMSD when suitable templates exist yet degrade sharply when they do not, and deep-learning methods have recently overtaken both, albeit unevenly. Three caveats qualify the apparent deep learning advantage: (1) accuracy remains strongly target-dependent (compact single-domain RNAs with abundant homologs do best, however, large multi-domain, homolog-poor, and flexible RNAs do not); (2) near-atomic accuracy has not been achieved for RNA as it has for proteins, reflecting the scarcity of high-resolution RNA structures and the dependence of folding on ions, ligands and cellular context; (3) cross-study comparability is limited, because test sets and metrics differ and training-set leakage is a persistent concern given the small RNA structural databank.

For the integration of probing restraints, because all three strategy classes are ultimately constrained by the same experimental data, how restraints are coupled to modeling is the crux of experimental-computational integration. Current practice falls into three modes: (1) restraints as additional input features, paired-site or contact maps from psoralen duplex-capture methods or spatial-proximity methods are supplied to the model alongside sequence and MSA; (2) restraints as constraints in the loss function or in post-prediction refinement, where structures that violate experimental distances are penalized or subsequently refined toward the measured contacts; and (3) restraints as selection or validation criteria, used only to rank candidate structures. Two cross-cutting issues follow. Restraint resolution and geometry differ by experimental class; duplex-capture methods report base-paired partners whereas proximity-capture methods report 3D contacts, so restraints are not interchangeable without appropriate transformation. In addition, most deep learning models were trained without restraint information and therefore require retraining or fine-tuning to exploit it. In our view, developing models natively designed to consume heterogeneous, error-prone restraints is a central challenge for the field.

Discussion and conclusions

As advances in probing and computational tools gradually unravel the full RNA structural landscape, these intricate folding architectures concealed within linear RNA sequences have emerged as indispensable regulators operating at the post-transcriptional layer. In this review, we summarize nine different experimental approaches (Table 1) for probing RNA tertiary and quaternary structures *in vivo*. PARIS, PARIS2, and COMRADES have enabled researchers to characterize multiple representative RNA higher-order structures, including the unique duplex motif within *XIST* lncRNA[16], the full-genome RNA topology and dynamic conformational shifts of enterovirus D68, and the global RNA architecture of the Zika virus[43,44,47]. LIGR-seq and SPLASH were applied to resolve an extensive landscape of RNA–RNA interactions that involve every major category of mRNA and non-coding RNA[45,46]. Moreover, recently refined DMS-MaPseq pipelines enable the capture of Hoogsteen interactions *in vivo* to reconstruct various RNA tertiary conformations, enriching methodologies for RNA tertiary structure determination[53]. Furthermore, those spatial proximity capture approaches have enabled systematic interrogation of RNA quaternary structures at the transcriptome scale. SHARC resolves protein-independent intramolecular tertiary folds and intermolecular quaternary assemblies while providing distance restraints for modeling heterogeneous conformational ensembles[48]. KARR-seq captures unpaired guanine-mediated long-range tertiary contacts and protein-free quaternary RNA assemblies that are orthogonal to canonical duplex base pairing[50]. RIC-seq maps RBP-tethered tertiary contacts and *in situ* quaternary supramolecular architectures, including RNP complexes, lncRNA–mRNA *trans* interactions, and RNA condensates, thereby delineating the cellular 3D RNA interactome topology[49]. MARIO complements these methods by profiling protein-assisted RNA–RNA interactions through targeted capture of specific RNAs and their interacting partners, revealing principles of higher-order organization and regulation across the eukaryotic transcriptome[51]. Collectively, these methods span the full spectrum from intramolecular to intermolecular, protein-independent to protein-dependent, and local to genome-wide scales, providing a multi-layered experimental framework for dissecting RNA higher-order structural organization in living cells.

To date, these high-throughput methods have been extensively deployed to map RNA higher-order structures and interactomes in human cell lines, mammalian tissues, and viral systems, providing a powerful framework for understanding how RNA conformations regulate gene expression, cellular identity, and disease progression. These findings also highlight the translational value of RNA structural information, as pathogenic RNA conformations and RNA–protein interaction interfaces may serve as targets for small molecules, antisense oligonucleotides, and other RNA-directed therapeutic strategies[34,100,101]. However, their application in plants remains virtually unexplored. A notable exception is RIC-seq, which was recently adapted for plant use. Researchers developed pRIC-seq (plant-adapted RIC-seq) and successfully applied this proximity-capturing pipeline to diploid and tetraploid cotton, demonstrating the feasibility of transcriptome-scale RNA spatial interactome mapping in plants[102]. Adapting additional high-throughput structural probing pipelines to plant systems, including overcoming challenges posed by cell walls, abundant secondary metabolites, and organellar RNA complexity, would enable transcriptome-wide interrogation of RNA folding dynamics during developmental transitions, diurnal cycles, and stress acclimation. Such efforts hold the potential to uncover plant-specific RNA structural regulatory codes and to identify novel RNA elements as targets for crop improvement.

High-throughput computational methods and deep learning-based algorithms have fundamentally reshaped the landscape of RNA higher-order structure research by bridging the widening gap between sequence abundance and experimentally resolved structures (Table 2). *Ab initio* folding and template-based approaches established the foundational frameworks for inferring 3D RNA conformations from thermodynamic principles and known structural motifs, enabling predictions from sequence to structure[58–82]. Recently, deep learning has promoted great progress in the field of RNA structural biology. Early methods predicted pairwise nucleotide constraints and then assembled 3D structures from them. Modern models generate atomic coordinates directly from sequence, and generative approaches can capture the

Table 1. Comparison of high-throughput experimental approaches for probing RNA higher-order structures *in vivo*.

Category	Method	Probe/principle	Structural information	Resolution/scale	Main strengths	Main limitations
Psoralen duplex capture	PARIS[43]	AMT, 365 nm UV, ND2D enrichment and proximity ligation	Direct intra-/intermolecular RNA duplexes and long-range base pairing	Near-base-pair, transcriptome-wide	Direct evidence of base pairing, unbiased toward specific RNAs or RBPs, captures long-range and alternative duplexes	Labor-intensive 2D-gel purification, relatively low recovery, psoralen sequence/duplex bias
	PARIS2[44]	Amotosalen, UV, optimized extraction, RNase III fragmentation and DD2D purification	Direct RNA duplexes and alternative pairing conformations	Near-base-pair, transcriptome-wide or target-enriched	More than 4,000-fold more efficient than PARIS, improved recovery and reduced photodamage	Retains psoralen bias, still requires UV and 2D-gel purification
	SPLASH[45]	Biotinylated psoralen, UV, streptavidin enrichment	Intra/intermolecular base-paired RNA interactions	Duplex-region, transcriptome-wide	Simple affinity enrichment; avoids 2D-gel purification, genome-wide mapping	Limited cellular penetration of biotinylated psoralen, monoadduct/background and fragment-length biases
	LIGR-seq[46]	AMT, UV, proximity ligation followed by RNase R enrichment	Intra-/intermolecular RNA duplexes	Duplex-region, transcriptome-wide	Relatively simple enrichment; matched controls facilitate background estimation	RNase R may retain highly structured uncrosslinked RNAs, background chimeras, psoralen bias
	COMRADES[47]	Azide-psoralen, UV, antisense capture and click enrichment	Base-paired interactions and alternative conformations of selected RNAs	Duplex-region, target-enriched	Very high depth for defined RNAs, suitable for long RNAs and structural ensembles	Requires predefined targets and antisense probes; not unbiased transcriptome-wide, retains psoralen bias
Protein-independent spatial proximity capture	SHARC[48]	Reversible bifunctional 2′-OH acylation, exonuclease trimming	Pairing-independent spatial proximity and approximate 3D-distance constraints	Near-nucleotide after trimming, transcriptome-wide	Provides spatial restraints for 3D modeling, detects noncanonical contacts and alternative conformations	Accessibility bias toward flexible/solvent-exposed 2′-OH groups, complex purification, resolution depends on trimming
	KARR-seq[50]	N3-kethoxal, DBCO-PAMAM SPAAC crosslinking	Stable/transient, non-duplex intra-/intermolecular spatial contacts	Contact/fragment level, transcriptome-wide	RBP-independent capture, tunable spatial range, captures transient and non-duplex interactions	Requires accessible G residues, contact frequency does not directly equal physical distance, proximity may not indicate direct interaction
RBP-mediated spatial proximity capture	RIC-seq[49]	Formaldehyde fixation, MNase fragmentation, pCp-biotin 3'-end labeling and *in situ* proximity ligation	RBP-associated intra-/intermolecular RNA contacts and RNP topology	High-resolution contact map, transcriptome-wide	Preserves native cellular organization, global mapping of RNP-mediated RNA networks	Cannot generally distinguish direct RNA–RNA contacts from protein-bridged proximity; low-abundance RNAs require deeper sequencing
	MARIO[51]	Formaldehyde/EGS dual crosslinking (with a UV route for direct RNA-protein contacts), protein biotinylation and streptavidin enrichment	RBP-assisted intra-/intermolecular RNA proximity	Interaction-site/fragment level, transcriptome-wide	No specific antibody, tagged RBP or anchor RNA required, applicable to tissues, linker aids chimera identification	Does not directly demonstrate base pairing; contacts may be protein-mediated, post-lysis ligation may introduce non-native proximity
DMS-based N7-G probing	BASH MaP[52]	DMS, KBH4 reduction, acid depurination converts m^7G into RT-detectable AP sites	N7-G accessibility/protection, Hoogsteen interactions and tertiary contacts	Single-nucleotide and single-molecule, mainly targeted	Converts otherwise RT-invisible m^7G into robust mutation signals, integrates Watson–Crick- and Hoogsteen-face information	Requires extra chemical conversion steps, N7-G-centered; protection does not identify the interaction partner or 3D distance
	msDMS-MaP[53]	Optimized MaP RT directly detects N7-methyl-G together with N1/N3 modifications	Secondary, tertiary, and RNP/quaternary-state information	Single-nucleotide and single-molecule, targeted/rRNA	Multiple structural layers from one DMS experiment; no m7G chemical conversion required	N7-G protection has multiple possible causes; interpretation requires controls; G-centered and not a direct pairwise contact map

full conformational ensemble of an RNA rather than a single static structure[56,84–92]. Beyond individual RNAs, deep learning-driven computational workflows enable joint modeling of RNA together with proteins, DNA, and small-molecule ligands in larger assemblies[93–95]. Collectively, these advances are transforming RNA structural biology from descriptive observation into predictive science.

Challenges and future perspectives

In the future, the integration of high-throughput experimental probing data with AI-driven modeling will serve as the major driving force for progress in this field. Experimental techniques including PARIS, SHARC, RIC-seq, and DMS-MaPseq generate *in vivo* spatial constraints, which can both guide computational structure predictions and independently verify their outputs. Conversely, computational models help select promising RNA targets and resolve structural ambiguities that pure experimental approaches cannot resolve. This reciprocal feedback loop is indispensable for addressing several long-standing bottlenecks in RNA structural biology, such as predicting long RNAs, which remains the most difficult case; conformational space grows combinatorially with sequence length, multi-domain RNAs have poorly constrained inter-domain contacts, such RNAs are frequently homolog-poor and underrepresented in structural databases, and computational cost scales unfavorably with length. A second bottleneck is the limited coverage of low-abundance transcripts, since crosslinking- and sequencing-based methods generate signal proportional to abundance and are therefore biased toward highly expressed RNAs; low-abundance regulatory and non-coding transcripts remain largely uncharacterized. Most fundamentally,

Table 2. Comparison of representative computational approaches for RNA higher-order structure modeling and their compatibility with experimental restraints.

Category	Representative method	Input/representation	Modeling/sampling	Experimental restraints	Strengths	Limitations
Physics-/knowledge-based CG	NAST[58]	Sequence, SS, 1 bead/nt	Knowledge-based CG sampling	Direct	Large-RNA modeling integrates sparse structural data	Coarse structural resolution
	iFoldRNA v2[59,60]	Sequence, 3 beads/nt	Replica-exchange DMD	Direct	Native base-pairing/solvent-accessibility restraints, all-atom reconstruction	Sampling and force-field accuracy limit long unconstrained RNAs
	Ernwin[61]	Sequence and SS, helix-based CG representation	MCMC sampling of helix arrangements	None	Efficient global-shape sampling for large RNAs	Produces coarse-grained topology with limited local detail
	BARNACLE[62]	RNA backbone/local structured dihedral-angle representation	Dynamic Bayesian network sampling	None	Efficient sampling of realistic local backbone conformations	Long-range interactions require a separate global energy function
	oxRNA[63]	Sequence, rigid-body nucleotide	MC/MD, thermodynamic model	Generic	RNA thermodynamics, folding pathways, and free-energy landscapes	Experimental contacts require conversion into generic external forces/restraints
	SimRNA[64, 65]	Sequence, optional SS, 5 interaction centers/nt	MC/REMC	Direct	Supports pseudoknots and native SS/distance restraints	Conformational sampling becomes difficult for long RNAs
	IsRNA family[66–69]	Sequence or SS, 5 beads/nt	REMD with statistical potentials	Limited	Models canonical and noncanonical base pairs applicable to large-RNA refinement	Accuracy remains limited by CG force fields and conformational sampling
	HiRE-RNA[70]	Sequence, high-resolution CG representation	MD/REMD	Limited	Detailed noncanonical and multiple base-pair interactions, folding dynamics	Validation in the original study focused mainly on 1276 nt systems
	RNAJP[71]	Sequence, SS, CG representation	Global topology sampling, refinement	Direct	Multiway junctions, noncanonical interactions, and global topology	Requires an input 2D structure primarily optimized for junction-rich RNAs
	TiRNA[72]	Sequence, optional SS, 3 beads/nt	MC, temperature-/ion-dependent force field	None	Explicit temperature/ion effects, predicts pseudoknots, multiway junctions and thermal stability	Native experimental-probing restraint integration was not reported
Template-/fragment-based	ModeRNA[73]	Target–template alignment, homologous template	Comparative modeling	None	Efficient and accurate when close homologous templates are available	Strongly dependent on template availability and alignment quality
	RNABuilder[74]	Sequence, structural constraints	Multibody dynamics	Direct	Flexible integration of user-defined experimental constraints scales to large RNAs	Requires substantial prior structural/contact information
	Vfold series[75]	Sequence, SS, motif/template library	Template/motif assembly, refinement	Direct	Automated modeling of pseudoknots, loops, and multiway junctions	Accuracy is limited by availability of suitable structural templates
	FebRNA[76]	Sequence SSCG fragment ensembles	Fragment-ensemble assembly all-atom reconstruction	Limited	Reliable modeling of pseudoknots and multiway junctions	Fragment-library gaps; tertiary motifs not explicitly modelled, validation mainly ≤ ~390 nt
	RNAComposer[77]	Sequence, SS, motif library	Automated fragment/motif assembly	Direct	Fast automated modeling; accepts experimental atom-distance restraints	Depends on accurate SS and availability of appropriate structural motifs
	3dRNA[78]	Sequence, SS, structural fragments	Fragment assembly, scoring/refinement	Direct	Can refine models using experimental/evolutionary contact restraints	Fragment/template coverage limits accurate modeling of novel global folds
	FARFAR2[80]	Sequence, optional SS, atomic fragments	Rosetta MC, all-atom refinement	Generic	High-resolution all-atom modeling of complex RNA folds	*De novo* sampling becomes a major bottleneck above ~80 nt; scoring-function limitations remain
	MC-Sym[81]	Sequence, SS/NCM constraints, motif library	Motif assembly stochastic/MC search	Limited	Explicit representation of canonical and noncanonical nucleotide motifs	Limited motif-library coverage; best suited to relatively short RNAs
	RNAthreader[82]	Sequence profile, SS topology	Template identification and threading	None	Fast and accurate remote-template detection, including pseudoknotted RNAs	Performs template alignment rather than independent full 3D reconstruction

(to be continued)

https://doi.org/10.48130/gcomm-0026-0022

https://doi.org/10.48130/gcomm-0026-0022

Genomics Communications

Table 2. (continued)

Category	Representative method	Input/representation	Modeling/sampling	Experimental restraints	Strengths	Limitations
Deep learning, restraint-first	trRosettaRNA[85]	Sequence, MSA, predicted SS	Pairwise geometry prediction and optimization	None	Automated prediction with competitive global-topology accuracy	Sensitive to SS quality and structural homology; performance decreases on novel RNAs
	DeepFoldRNA[86]	Sequence, MSA-derived features, predicted SS	Geometric-potential prediction and gradient-based optimization	None	Fast atomic-level prediction compared with MC-based approaches	Relies on homologous-sequence/MSA-derived evolutionary information
	epRNA[87]	RNA sequence	Euclidean-invariant neural network and distance-matrix prediction	None	Rapid sequence-only prediction of RNA 3D backbone geometry	Limited availability of experimental RNA structures constrains training
	RhoFold+[88]	Sequence, RNA language-model features, MSA when available	End-to-end 3D prediction	None	Strong automated single-chain prediction, RNA language-model integration	Accuracy remains sensitive to MSA quality and depth
	NuFold[89]	Sequence, MSA, predicted SS	Evoformer-like end-to-end prediction	None	Direct full-atom output, accurate local nucleotide geometry	Sensitive to MSA/SS quality; limited RNA structural training data
	DRfold2[90]	Single sequence, pretrained RNA language-model features	End-to-end denoising, post-optimization	Limited	MSA-independent single-sequence; prediction-efficient inference	Performance decreases for larger RNAs (> ~200 nt); benchmarks focus on monomeric RNAs
Generative modeling	RNAbpFlow[91]	Sequence, base-pair map	SE(3)-equivariant flow matching	Limited	MSA/template-independent all-atom conformational-ensemble generation	Strongly dependent on input base-pair accuracy, not optimized for very large RNAs, probing restraints are not yet implemented
	DynaRNA[92]	RNA sequence, no MSA required	Diffusion model, equivariant GNN	None	Rapid generation of dynamic RNA conformational ensembles	PDB-biased training data underrepresent extended/disordered conformations
Complex modeling	RoseTTAFoldNA[93]	Protein/nucleic-acid sequences, MSA	Three-track/SE(3)-equivariant complex prediction	None	Direct prediction of protein–RNA and other nucleic-acid-containing complexes	RNA-only accuracy is unevenlarge/synthetic RNA targets remain difficult
	AlphaFold 3[94]	Protein, RNA, DNA, ligand and ion components	Diffusion-based all-atom complex prediction	None	Broad multimolecular modeling scope in a unified framework	Non-Watson–Crick interactions, orphan RNAs, and alternative conformations remain challenging
	ProRNA3D-single[95]	Single protein with RNA sequences pretrained language-model features	Interaction-map prediction, geometric restraints and optimization	None	ProteinRNA complex prediction without paired MSA/co-evolutionary input	Specialized for proteinRNA complexes rather than general RNA monomer prediction

Direct: native support for experimental restraints. **Generic**: generic distance, contact, base-pairing, or structural restraints can be incorporated, but experimental data must first be converted into the format accepted by the modeling framework. **Indirect**: experimental information is not accepted as a native structural restraint during prediction but can be incorporated through secondary-structure conditioning, downstream filtering, model selection, or refinement. **None**: no publicly documented mechanism for incorporating experimental probing restraints. **Custom**: experimental restraints may be incorporated through user-defined or manually implemented restraint terms rather than a standardized native interface. **Limited**: only restricted forms of structural information or specific restraint types are supported. **SS**, secondary structure; **MSA**, multiple sequence alignment; **CG**, coarse-grained; **MC,** Monte Carlo; **MCMC**, Markov chain Monte Carlo; **MD**, molecular dynamics; **DMD**, discrete molecular dynamics; **REMC**, replica-exchange Monte Carlo; **REMD**, replica-exchange molecular dynamics; **GNN**, graph neural network; **NCM**, nucleotide cyclic motif; **nt**, nucleotide.

https://doi.org/10.48130/gcomm-0026-0022

current probing methods report a population-averaged signal, so RNAs coexisting in multiple conformations are averaged into a single apparent structure; alternative conformations present at substantial frequency are difficult to resolve; rare functional states may be invisible; and information on conformational dynamics is lost. With ongoing improvements in single-molecule resolution and multimodal data integration, the combination of high-throughput structure probing and deep learning will enable us to achieve a comprehensive understanding of how RNA folding determines biological function. Such knowledge will have profound implications, ranging from RNA-targeted therapeutics and synthetic RNA engineering to molecular breeding in both animals and plants.

Author contributions

The authors confirm their contributions to this study as follows: study conception and design: Tao J, Yang M; draft manuscript preparation and figures organization: Tao J, Li S; manuscript revision: Tao J, Yang M. All authors reviewed the results and approved the final version of the manuscript.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Acknowledgments

This research was financially supported by the National Natural Science Foundation of China (Grant No. 32470580), the Natural Science Foundation of Anhui Province (Grant No. 2508085MC042), and the Youth Innovation Fund of the University of Science and Technology of China (WK9100250094).

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 4 September 2026; Revised 2 October 2026; Accepted 9 October 2026; Published online 27 October 2026

References

[1] Gilbert W. 1986. Origin of life: The RNA world. *Nature* 319:618
[2] Sharp PA. 2009. The centrality of RNA. *Cell* 136:577–580
[3] Gottesman S, Storz G. 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harbor Perspectives in Biology* 3:a003798
[4] Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry* 81:145–166
[5] Strobel EJ, Watters KE, Loughrey D, Lucks JB. 2016. RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs. *Current Opinion in Biotechnology* 39:182–191
[6] Doherty EA, Doudna JA. 2001. Ribozyme structures and mechanisms. *Annual Review of Biophysics and Biomolecular Structure* 30:457–475
[7] Steitz TA, Steitz JA. 1993. A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences of the United States of America* 90:6498–6502
[8] Walter NG, Engelke DR. 2002. Ribozymes: catalytic RNAs that cut things, make things, and do odd and useful jobs. *Biologist* 49:199–203
[9] Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, et al. 2023. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nature Reviews Molecular Cell Biology* 24:430–447
[10] Weeks KM. 2021. SHAPE directed discovery of new functions in large RNAs. *Accounts of Chemical Research* 54:2502–2517
[11] Zhang H, Ding Y. 2021. Novel insights into the pervasive role of RNA structure in post-transcriptional regulation of gene expression in plants. *Biochemical Society Transactions* 49:1829–1839
[12] Breaker RR. 2012. Riboswitches and the RNA world. *Cold Spring Harbor Perspectives in Biology* 4:a003566
[13] Venkata Subbaiah KC, Hedaya O, Wu J, Jiang F, Yao P. 2019. Mammalian RNA switches: Molecular rheostats in gene regulation, disease, and medicine. *Computational and Structural Biotechnology Journal* 17:1326–1338
[14] Ray PS, Jia J, Yao P, Majumder M, Hatzoglou M, et al. 2009. A stress-responsive RNA switch regulates VEGFA expression. *Nature* 457:915–919
[15] Yao P, Potdar AA, Ray PS, Eswarappa SM, Flagg AC, et al. 2013. The HILDA complex coordinates a conditional switch in the 3'-untranslated region of the VEGFA mRNA. *PLoS Biology* 11:e1001635
[16] McHugh CA, Chen CK, Chow A, Surka CF, Tran C, et al. 2015. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521:232–236
[17] Saha P, Patel S, Tsao LH, Tavares RC, Choi H, et al. 2026. Pnky lncRNA secondary structure maps identify functional regions that control neurogenesis in neural stem cells. *Cell Reports* 45:117451
[18] Kiontke KC, Herrera RA, Vuong E, Luo J, Schwarz EM, et al. 2019. The long non-coding RNA lep-5 promotes the juvenile-to-adult transition by destabilizing LIN-28. *Developmental Cell* 49:542–555. e9
[19] Lawson H, Vuong E, Miller RM, Kiontke K, Fitch DH, et al. 2019. The Makorin *lep-2* and the lncRNA *lep-5* regulate *lin-28* to schedule sexual maturation of the *C. elegans* nervous system. *eLife* 8:e43660
[20] Kwok CK, Ding Y, Shahid S, Assmann SM, Bevilacqua PC. 2015. A stable RNA G-quadruplex within the 5′-UTR of Arabidopsis thaliana ATR mRNA inhibits translation. *Biochemical Journal* 467:91–102
[21] Yang X, Cheema J, Zhang Y, Deng H, Duncan S, et al. 2020. RNA G-quadruplex structures exist and function in vivo in plants. *Genome Biology* 21:226
[22] Park C, Cho HS, Lim Y, Cho CH, Nam H, et al. 2025. Evolution of the *JULGI-SMXL4/5* module for phloem development in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* 122:e2416674122
[23] Zhang Y, Yang M, Duncan S, Yang X, Abdelhamid MAS, et al. 2019. G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Research* 47:11746–11754
[24] Yang X, Yu H, Duncan S, Zhang Y, Cheema J, et al. 2022. RNA G-quadruplex structure contributes to cold adaptation in plants. *Nature Communications* 13:6224
[25] Yang M, Zhu P, Cheema J, Bloomer R, Mikulski P, et al. 2022. *In vivo* single-molecule analysis reveals COOLAIR RNA structural diversity. *Nature* 609:394–399
[26] Kiss T. 2002. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* 109:145–148
[27] Rabany O, Nachmani D. 2023. Small nucleolar (sno)RNA: therapy lays in translation. *Noncoding RNA* 9:35
[28] Cao Y, Wang J, Wu S, Yin X, Shu J, et al. 2022. The small nucleolar RNA SnoR28 regulates plant growth and development by directing rRNA maturation. *The Plant Cell* 34:4173–4190
[29] Bies-Etheve N, Sáez-Vásquez J. 2026. C/D-box snoRNP assembly and trafficking in plants: organelle-derived insights into 2′-O-methylation. *Journal of Experimental Botany* 2:erag162
[30] Thore S, Leibundgut M, Ban N. 2006. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* 312:1208–1211
[31] Bonilla SL, Jang K. 2024. Challenges, advances, and opportunities in RNA structural biology by Cryo-EM. *Current Opinion in Structural Biology* 88:102894

[32] Deng J, Fang X, Huang L, Li S, Xu L, et al. 2023. RNA structure determination: From 2D to 3D. *Fundamental Research* 3:727–737
[33] Ma H, Jia X, Zhang K, Su Z. 2022. Cryo-EM advances in RNA structure determination. *Signal Transduction and Targeted Therapy* 7:58
[34] Aguilar R, Mardones C, Moreno AA, Cepeda-Plaza M. 2025. A guide to RNA structure analysis and RNA-targeting methods. *The FEBS Journal* 292:2751–2766
[35] Spitale RC, Incarnato D. 2023. Probing the dynamic RNA structurome and its functions. *Nature Reviews Genetics* 24:178–196
[36] Tamura M, Holbrook SR. 2002. Sequence and structural conservation in RNA ribose zippers. *Journal of Molecular Biology* 320:455–474
[37] Batey RT, Rambo RP, Doudna JA. 1999. Tertiary motifs in RNA structure and folding. *Angewandte Chemie (International Ed in English)* 38:2326–2343
[38] Rhodes D, Lipps HJ. 2015. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Research* 43:8627–8637
[39] Huang Z, Lin W, Wang L, Su Z. 2026. Molecular insights into dynamic RNA quaternary assemblies. *RNA Biology* 23:1–13
[40] Corley M, Flynn RA, Lee B, Blue SM, Chang HY, et al. 2020. Footprinting SHAPE-eCLIP reveals transcriptome-wide hydrogen bonds at RNA-protein interfaces. *Molecular Cell* 80:903–914. e8
[41] Gabryelska MM, Badrock AP, Lau JY, O'Keefe RT, Crow YJ, et al. 2022. Global mapping of RNA homodimers in living cells. *Genome Research* 32:956–967
[42] Xiao L, Fang L, Kool ET. 2024. 2′-OH as a universal handle for studying intracellular RNAs. *Cell Chemical Biology* 31:110–124
[43] Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, et al. 2016. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165:1267–1279
[44] Zhang M, Li K, Bai J, Velema WA, Yu C, et al. 2021. Optimized photochemistry enables efficient analysis of dynamic RNA structuromes and interactomes in genetic and infectious diseases. *Nature Communications* 12:2344
[45] Aw JGA, Shen Y, Wilm A, Sun M, Lim XN, et al. 2016. *In vivo* mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Molecular Cell* 62:603–617
[46] Sharma E, Sterne-Weiler T, O'Hanlon D, Blencowe BJ. 2016. Global mapping of human RNA–RNA interactions. *Molecular Cell* 62:618–626
[47] Ziv O, Gabryelska MM, Lun ATL, Gebert LFR, Sheu-Gruttadauria J, et al. 2018. COMRADES determines *in vivo* RNA structures and interactions. *Nature Methods* 15:785–788
[48] Van Damme R, Li K, Zhang M, Bai J, Lee WH, et al. 2022. Chemical reversible crosslinking enables measurement of RNA 3D distances and alternative conformations in cells. *Nature Communications* 13:911
[49] Cai Z, Cao C, Ji L, Ye R, Wang D, et al. 2020. RIC-seq for global *in situ* profiling of RNA–RNA spatial interactions. *Nature* 582:432–437
[50] Wu T, Cheng AY, Zhang Y, Xu J, Wu J, et al. 2024. KARR-seq reveals cellular higher-order RNA structures and RNA–RNA interactions. *Nature Biotechnology* 42:1909–1920
[51] Nguyen TC, Cao X, Yu P, Xiao S, Lu J, et al. 2016. Mapping RNA–RNA interactome and RNA structure *in vivo* by MARIO. *Nature Communications* 7:12023
[52] Oleynikov M, Jaffrey SR. 2024. RNA tertiary structure and conformational dynamics revealed by BASH MaP. *eLife* 13:RP98540
[53] Saleem I, Miller T, Kearns L, Hoang A, Meehan J, et al. 2026. Multi-site DMS probing reveals higher-order structure of RNA-protein complexes in living cells. *Molecular Cell* 86:1815–1830. e9
[54] Magnus M, Matelska D, Łach G, Chojnowski G, Boniecki MJ, et al. 2014. Computational modeling of RNA 3D structures, with the aid of experimental restraints. *RNA Biology* 11:522–536
[55] Bernard C, Postic G, Ghannay S, Tahi F. 2024. State-of-the-RNArt: benchmarking current methods for RNA 3D structure prediction. *NAR Genomics and Bioinformatics* 6:lqae048
[56] Wang W, Su B, Peng Z, Yang J. 2026. Integrated experimental and AI innovations for RNA structure determination. *Nature Biotechnology* 44:205–214
[57] Zhang S, Li J, Chen SJ. 2024. Machine learning in RNA structure prediction: Advances and challenges. *Biophysical Journal* 123:2647–2657
[58] Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, et al. 2009. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* 15:189–199
[59] Sharma S, Ding F, Dokholyan NV. 2008. IFoldRNA: Three-dimensional RNA structure prediction and folding. *Bioinformatics* 24:1951–1952
[60] Krokhotin A, Houlihan K, Dokholyan NV. 2015. iFoldRNA v2: folding RNA with constraints. *Bioinformatics* 31:2891–2893
[61] Kerpedjiev P, Höner Zu Siederdissen C, Hofacker IL. 2015. Predicting RNA 3D structure using a coarse-grain helix-centered model. *RNA* 21:1110–1121
[62] Frellsen J, Moltke I, Thiim M, Mardia KV, Ferkinghoff-Borg J, et al. 2009. A probabilistic model of RNA conformational space. *PLoS Computational Biology* 5:e1000406
[63] Šulc P, Romano F, Ouldridge TE, Doye JPK, Louis AA. 2014. A nucleotide-level coarse-grained model of RNA. *Journal of Chemical Physics* 140:235102
[64] Boniecki MJ, Lach G, Dawson WK, Tomala K, Lukasz P, et al. 2016. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Research* 44:e63
[65] Moafinejad SN, de Aquino BRH, Boniecki MJ, Pandaranadar Jeyeram IPN, Nikolaev G, et al. 2024. SimRNAweb v2.0: a web server for RNA folding simulations and 3D structure modeling, with optional restraints and enhanced analysis of folding trajectories. *Nucleic Acids Research* 52:W368–W373
[66] Zhang D, Chen SJ. 2018. IsRNA: an iterative simulated reference state approach to modeling correlated interactions in RNA folding. *Journal of Chemical Theory and Computation* 14:2230–2239
[67] Zhang D, Li J, Chen SJ. 2021. IsRNA1: De *novo* prediction and blind screening of RNA 3D structures. *Journal of Chemical Theory and Computation* 17:1842–1857
[68] Zhang D, Chen SJ, Zhou R. 2021. Modeling noncanonical RNA base pairs by coarse-grained IsRNA2 model. *The Journal of Physical Chemistry B* 125:11907–11915
[69] Jiang H, Xu Y, Tong Y, Zhang D, Zhou R. 2024. IsRNAcirc: 3D structure prediction of circular RNAs based on coarse-grained molecular dynamics simulation. *PLoS Computational Biology* 20:e1012293
[70] Pasquali S, Derreumaux P. 2010. HiRE-RNA: a high resolution coarse-grained energy model for RNA. *The Journal of Physical Chemistry B* 114:11957–11966
[71] Li J, Chen SJ. 2023. RNAJP: enhanced RNA 3D structure predictions with non-canonical interactions and global topology sampling. *Nucleic Acids Research* 51:3341–3356
[72] Wang X, Zhou Z, Yu S, Tan YL, Shi YZ, et al. 2026. TiRNA: a coarse-grained method with temperature and ion effects for RNA structure folding and prediction. *Nucleic Acids Research* 54:gkaf1499
[73] Rother M, Rother K, Puton T, Bujnicki JM. 2011. ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Research* 39:4007–4022
[74] Flores SC, Sherman M, Bruns CM, Eastman P, Altman RB. 2011. Fast flexible modeling of RNA structure using internal coordinates. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 8:1247–1257
[75] Li J, Zhang S, Zhang D, Chen SJ. 2022. Vfold-Pipeline: a web server for RNA 3D structure prediction from sequences. *Bioinformatics* 38:4042–4043
[76] Zhou L, Wang X, Yu S, Tan YL, Tan ZJ. 2022. FebRNA: an automated fragment-ensemble-based model for building RNA 3D structures. *Biophysical Journal* 121:3381–3392
[77] Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, et al. 2012. Automated 3D structure composition for large RNAs. *Nucleic Acids Research* 40:e112
[78] Zhao Y, Huang Y, Gong Z, Wang Y, Man J, et al. 2012. Automated and fast building of three-dimensional RNA structures. *Scientific Reports* 2:734

[79] Das R, Baker D. 2007. Automated de novo prediction of native-like RNA tertiary structures. *Proceedings of the National Academy of Sciences of the United States of America* 104:14664−14669
[80] Watkins AM, Rangan R, Das R. 2020. FARFAR2: improved *de novo* Rosetta prediction of complex global RNA folds. *Structure* 28:963−976. e6
[81] Parisien M, Major F. 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452:51−55
[82] Du Z, Peng Z, Yang J. 2024. RNA threading with secondary structure and sequence profile. *Bioinformatics* 40:btae080
[83] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583−589
[84] Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, et al. 2020. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences of the United States of America* 117:1496−1503
[85] Wang W, Feng C, Han R, Wang Z, Ye L, et al. 2023. trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nature Communications* 14:7266
[86] Pearce R, Omenn GS, Zhang Y. 2022. *De novo* RNA tertiary structure prediction at atomic resolution using geometric potentials from deep learning. *bioRxiv* Preprint: 491755
[87] Sha CM, Wang J, Dokholyan NV. 2024. Predicting 3D RNA structure from the nucleotide sequence using Euclidean neural networks. *Biophysical Journal* 123:2671−2681
[88] Shen T, Hu Z, Sun S, Liu D, Wong F, et al. 2024. Accurate RNA 3D structure prediction using a language model-based deep learning approach. *Nature Methods* 21:2287−2298
[89] Kagaya Y, Zhang Z, Ibtehaz N, Wang X, Nakamura T, et al. 2025. NuFold: end-to-end approach for RNA tertiary structure prediction with flexible nucleobase center representation. *Nature Communications* 16:881
[90] Li Y, Feng C, Zhang X, Tsukiyama S, Feng D, et al. 2026. DRfold2 is a deep learning-based tool that enables efficient and accurate RNA structure prediction. *PLoS Biology* 24:e3003659
[91] Tarafder S, Bhattacharya D. 2026. RNAbpFlow: base pair-augmented SE(3) flow matching for conditional RNA 3D structure generation. *Nature Methods* 23:1349−1358
[92] Li Z, Zhu J, Hong X, Mu J, Zheng Z, et al. 2025. DynaRNA: accurate dynamic RNA conformation ensemble generation with diffusion model. *Communications Biology* 8:1472
[93] Baek M, McHugh R, Anishchenko I, Jiang H, Baker D, et al. 2024. Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nature Methods* 21:117−121
[94] Abramson J, Adler J, Dunger J, Evans R, Green T, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630:493−500
[95] Roche R, Tarafder S, Bhattacharya D. 2025. Single-sequence protein-RNA complex structure prediction by geometric attention-enabled pairing of biological language models. *Cell Systems* 16:101400
[96] Das R, Kretsch RC, Simpkin AJ, Mulvaney T, Pham P, et al. 2023. Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins: Structure, Function, and Bioinformatics* 91:1747−1770
[97] Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, et al. 2012. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* 18:610−625
[98] Zhang Y, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins* 57:702−710
[99] Mariani V, Biasini M, Barbato A, Schwede T. 2013. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* 29:2722−2728
[100] Childs-Disney JL, Yang X, Gibaut QMR, Tong Y, Batey RT, et al. 2022. Targeting RNA structures with small molecules. *Nature Reviews Drug Discovery* 21:736−762
[101] Childs-Disney JL, Hoskins J, Rzuczek SG, Thornton CA, Disney MD. 2012. Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. *ACS Chemical Biology* 7:856−862
[102] Wen M, Liang X, Shi K, Fei L, Wang Y, et al. 2026. Global atlas of enhancer-promoter interactome in cotton genome revealed by profiling RNA-RNA spatial interactions. *Genome Biology* 27:12

Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit https://creativecommons.org/licenses/by/4.0/.