

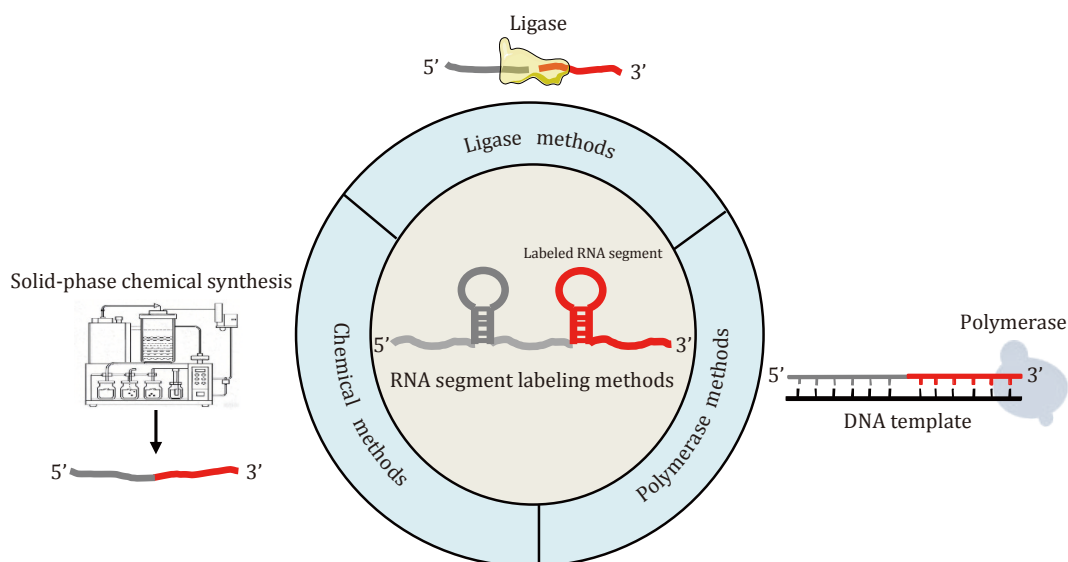
Methods and advances in RNA segment-selective labelling

Zixuan Hu^{1#}, Shiping Yang^{1#}, Xianyang Fang¹✉

¹ Key Laboratory of RNA Science and Engineering, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

Received: 16 July 2025 / Accepted: 23 September 2025

Graphic abstract



Abstract RNA is fundamental to all life, governing gene expression, regulating biological processes, and influencing growth, development, and cellular metabolism. However, elucidating the precise mechanisms of these functions requires methods to probe the complex structure, dynamics, and interactions of RNA molecules. In recent years, RNA segment-selective labeling has emerged as an essential technique for studying RNA structure, function, and dynamics. We classify various RNA segmental labeling methods based on their underlying principles into three categories: chemical methods (including solid-phase synthesis labeling and post-synthesis RNA chemical ligation), ligase-based methods (such as T4 RNA/DNA ligases, deoxyribozymes, and ribozymes), and polymerase-based methods (by DNA polymerase and T7 RNA polymerase). Each method has its unique characteristics and needs to be selected according to specific experimental requirements. Despite significant progress, current methods still face challenges. This review focuses on RNA segment-selective labeling techniques, exploring their underlying principles, advantages, limitations, and applications. The aim is to provide a foundation for developing novel RNA segmental labeling techniques with broader application potential in the future.

Zixuan Hu and Shiping Yang contributed equally to this work.

✉ Correspondence: fangxy@ibp.ac.cn (X. Fang)

Keywords RNA segmental labelling, RNA structures, Chemical ligation, Enzymatic ligation, Biophysical studies

INTRODUCTION

RNA stands as a cornerstone of life, orchestrating vital processes from genetic information transfer to dynamic regulation of cellular functions. Beyond its classical role as a messenger between DNA and proteins, RNA exhibits remarkable functional versatility, including catalyzing biochemical reactions (*e.g.*, ribozymes) (Serganov and Patel 2007), modulating gene expression (*e.g.*, miRNAs) (Bartel 2018), and serving as a structural scaffold (*e.g.*, ribosomal RNA) (Staley and Woolford 2009). The functional diversity of RNAs arises from their remarkable ability to form complex hierarchical structures on their own or in complex with binding partners and undergo structural changes in response to variations of cellular and environmental conditions (Ganser *et al.* 2019; Xu *et al.* 2022). Consequently, deciphering RNA structure-function relationships has become a pivotal frontier in life sciences, essential for unraveling the mysteries of life and developing RNA-based therapies.

In-depth research into the higher-order structure and conformational dynamics of RNA is undoubtedly the key to unlocking the mysteries of RNA's functional mechanisms. However, how to obtain this key remains a formidable challenge. Traditional structural biology techniques, such as X-ray crystallography (XRC) and cryo-electron microscopy (Cryo-EM), struggle with RNA's inherent flexibility, transient conformational states, and susceptibility to degradation (Westhof and Leontis 2021). Nuclear magnetic resonance (NMR) spectroscopy and small-angle X-ray and neutron scattering (SAXS/SANS) are powerful tools for characterizing the structure and dynamic processes of RNA or RNA-ligand complexes (Hu *et al.* 2022). However, NMR is limited to RNA sequences that are 40–60 nucleotides in length. As the molecular weight increases, severe signal overlap occurs in the NMR spectra of RNA, which restricts its use in structural analysis of longer RNA molecules and RNA-protein complexes (Barnwal *et al.* 2017). SAXS/SANS, on the other hand, suffers from relatively low resolution (Fang *et al.* 2015). These obstacles hinder our ability to map RNA structural landscapes and their functional implications.

RNA-specific labeling techniques, including segment-selective labeling (also referred to as segment-selective labeling) and site-specific labeling, provide promising solutions to these challenges. RNA segment-selective labeling aims to precisely label specific regions within

RNA molecules. These target regions can consist of one or more segments, which are labeled using various labeling agents such as fluorescent dyes, radioactive isotopes, chemical modifications, or other detectable molecular tags (Fig. 1A). Site-specific labeling targets individual nucleotide positions within an RNA molecule, allowing for precise localization of individual nucleotides and providing information about local interactions (Huang and Fang 2023), but may lack the context of larger functional domains. In contrast, RNA segmental labeling focuses on larger, defined regions within the RNA molecule, which expands our ability to study RNA structure and dynamics at the domain level, thereby providing a more comprehensive understanding of RNA overall structural dynamics and functional properties. Consequently, segment-selective labeling is often employed to study the relative orientation, dynamics, and interactions of entire structural domains, while site-specific labeling is ideal for interrogating local structural features or nucleotide-level interactions. The strategic application of RNA segment-selective labeling is revolutionizing RNA structural biology. By isotopically labeling specific RNA segments, NMR spectra can be simplified, helping to avoid confusion caused by signal overlap, thus breaking through the molecular weight limit of RNA that can be studied by NMR (Barnwal *et al.* 2017). Selective deuteration of RNA segments, combined with contrast variation, enables the determination of the relative spatial positioning of different structural domains or components within multi-domain RNA and its complexes in SANS studies (Hennig *et al.* 2013). For XRC, incorporating 2'-methylseleno modifications into specific regions of RNA molecules can, to some extent, aid in crystallization (Höbartner *et al.* 2005). RNA segmental labeling is a powerful tool that expands the application range of NMR, SANS, and XRC in RNA structural biology. By applying RNA segment-selective labeling techniques, researchers can more easily obtain high-resolution RNA structures, study conformational dynamics, and explore RNA-protein interactions, all of which are difficult to achieve with traditional methods.

This review will primarily focus on the methodologies of RNA segment-selective labeling. We will categorize these methods based on their underlying labeling principles (Fig. 1B), providing a comprehensive overview of existing techniques. For each method, we will discuss its fundamental principles, inherent advantages and limitations, and relevant

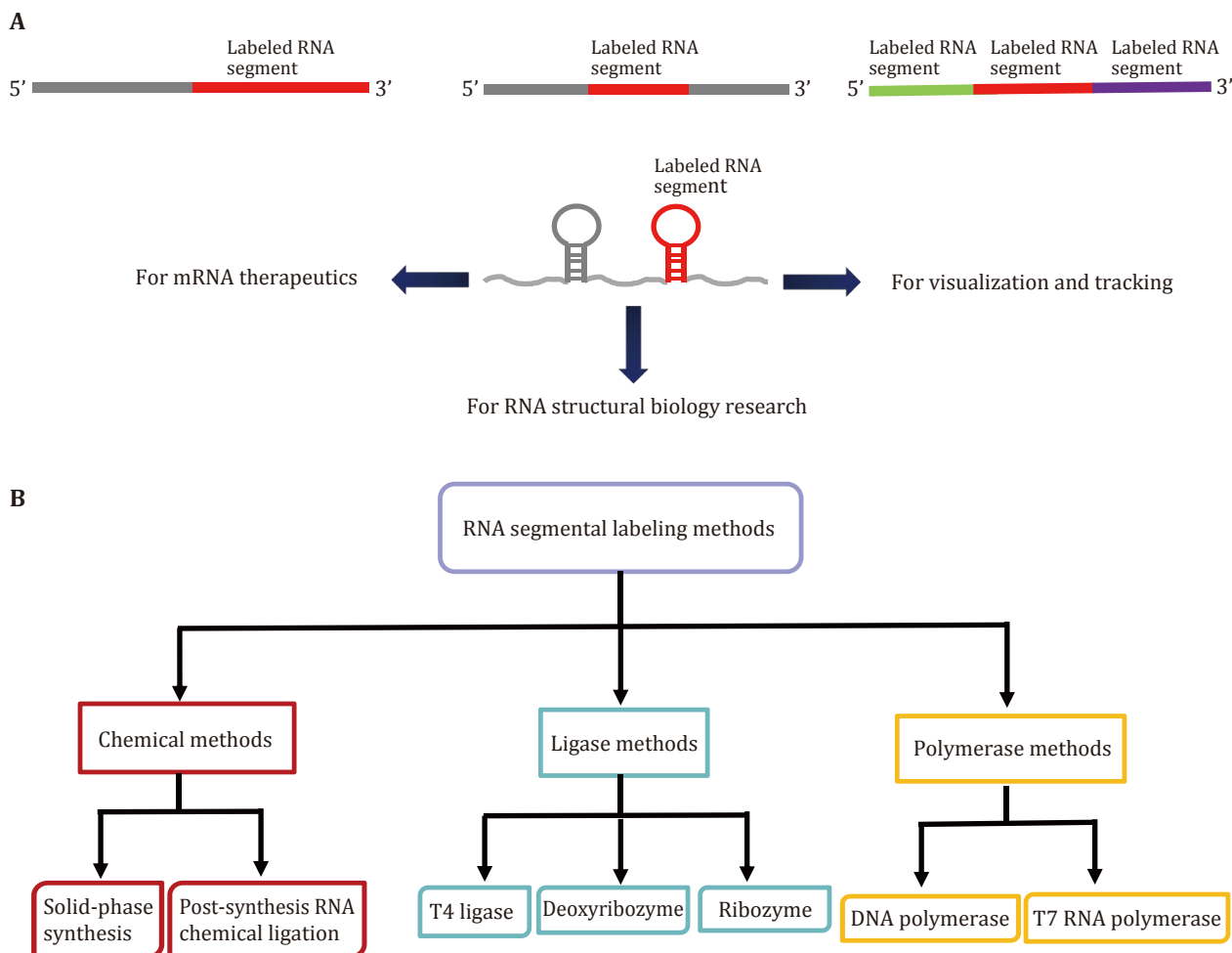


Fig. 1 A Schematic illustration of RNA labelled segments and their potential applications. B Classification of RNA segmental labeling methods

applications, particularly within the context of RNA structural biology. Acknowledging that challenges and limitations still exist despite the progress in RNA segmental labeling, this work aims to serve as a valuable resource and inspiration for future innovations in RNA segment-selective labeling methodologies.

CHEMICAL METHODS

Solid phase synthesis

The most common chemical method for RNA segmental labeling is solid phase chemical synthesis by phosphoramidite approaches (Gait 1984), in which modified nucleotides are used as substrates to introduce the label groups to RNA (Chow *et al.* 2008). In this strategy, an activated nucleoside phosphoramidite reaction with

a solid-support-bound nucleoside is used to synthesize the desired RNA sequence in the 3' to 5' direction. This technique facilitates the synthesis of RNA with either whole labelling or specific labelling, as demonstrated by the introduction of isotope labels (Becette *et al.* 2019), fluorophores (Kirk *et al.* 2001), spin labels (Domnick *et al.* 2020; Höbartner *et al.* 2012; Weinrich *et al.* 2017), and other functional groups (Fig. 2).

The phosphoramidite method involves removing the 4,4'-dimethoxytrityl (DMTr) protecting group from the 5'-hydroxyl (OH) of the 3'-nucleoside using trichloroacetic or dichloroacetic acid. Subsequently, the nucleophilic 5'-OH attacks the activated phosphoramidite, forming a phosphite triester bond that couples the two nucleosides. Unreacted 3'-nucleosides are then capped by acetylation of the 5'-OH groups to ensure the accuracy of the RNA sequence. Finally, the phosphite triester is oxidized to a phosphotriester with either

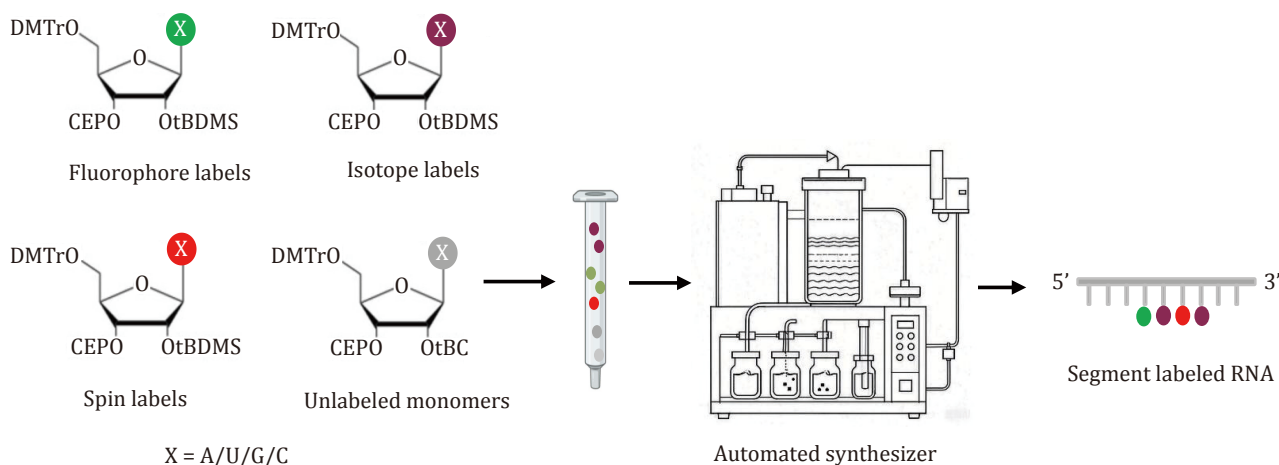


Fig. 2 The labeled RNA segments were synthesized by a solid-phase chemical synthesizer. Fluorophore-labeled monomers (green), isotope-labeled monomers (purple), spin-labeled monomers (red), and unlabeled monomers (gray). X denotes different nucleotides (A, U, G, or C). These monomers undergo a controlled assembly process in the synthesizer, where they are incorporated into the RNA sequence, resulting in an RNA molecule with specifically labeled segments

iodine (I_2) or tert-butyl hydroperoxide. These steps repeat until the entire RNA sequence is synthesized (Becette *et al.* 2019; Glazier *et al.* 2020). The addition of protective groups to the 2'-hydroxyl group, such as the tert-butyldimethylsilyl (tBDMS) group, [(triisopropylsilyl) oxy]methyl (TOM) group, and 2'-cyanoethoxymethyl (CEM) group, is employed as a means of precluding the occurrence of side reactions in the synthesis of RNA (Somoza 2008). At the same time, the efficiency of long RNA coupling is reduced due to the large steric hindrance generated by these protective groups (El-Sagheer and Brown 2010). Phosphodiester backbone cleavage and 3' to 2' phosphate migration also take place during the removal of 2'-protecting groups (El-Sagheer and Brown 2010), which makes large quantities of produce of target-modified RNA more difficult. Therefore, RNA synthesis over 50 nucleotides faces problems such as poor yield and impure product.

Isotopically labeled phosphoramidite monomers are widely used in RNA structural biology studies. Micura *et al.* first synthesized 2'-Se-methyl-labeled nucleoside phosphoramidite building blocks and improved the RNA solid-phase synthesis cycle (Höbartner *et al.* 2005). This advancement allowed the chemical synthesis of multiple Se-labeled RNAs, making them suitable for X-ray crystallography studies (Höbartner *et al.* 2005). Solid-phase synthesis enables segment-selective labelling by precisely controlling the incorporation of modified monomers during the synthesis cycle. For example, when targeting a specific functional domain of RNA (*e.g.*, the P3 stem of the PreQ class-II riboswitch), only labeled monomers (*e.g.*, $^{15}N(3)$ -uridine for isotope labelling) are used during the synthesis of this domain's

nucleotide sequence, while unlabeled monomers are used for other regions (Neuner *et al.* 2015). This strategy avoids non-specific labeling and ensures that the label is confined to the region of interest – consistent with the technique's ability to generate RNA with “either whole labelling or specific labelling”. Strebiter *et al.* demonstrated the potential of investigating the conformational dynamics of the epsilon (ϵ) RNA element of the duck Hepatitis B virus (dHBV ϵ) with atom-specific ^{13}C -labeling (Strebiter *et al.* 2018). Similarly, the introduction of fluorescent probes like 2-Aminopurine (ZAP) incorporated into the hammerhead ribozyme cleavage site can be used as a real-time probe of enzymatic cleavage and inhibition (Kirk *et al.* 2001). Spin labels are stably introduced into RNA by solid-phase synthesis for electron paramagnetic resonance (EPR) study of RNA structure (Domnick *et al.* 2020; Höbartner *et al.* 2012; Weinrich *et al.* 2017). However, the direct incorporation of modified nucleotides encounters several challenges. Reaction conditions restrict the types of nucleoside modifications that can be introduced, and the synthesis length is limited by reaction efficiency.

Post-synthetic RNA chemical ligation

During solid phase synthesis, RNA molecules are typically limited to a length of 70–80 nucleotides (Obi and Chen 2021). Beyond this range, synthesis becomes significantly more challenging and costly. Post-synthetic chemical ligation offers a practical solution by first obtaining two RNA fragments through synthesis or transcription and then joining them using chemical

reactions to produce segmentally labeled RNA.

Post-synthetic chemical ligation creates non-natural linkages to join two RNA strands, with strategies including periodate chemistry, thiol chemistry, click chemistry, and BrCN/EDC-based chemical ligation methods (Depmeier *et al.* 2021). This approach shares conceptual parallels with native chemical ligation (NCL) in protein synthesis, a well-established technique for joining unprotected peptide segments via thioester-mediated transacylation (Dawson *et al.* 1994). Both RNA chemical ligation and protein NCL enable the assembly of large biomolecules from smaller fragments, facilitating access to modified or otherwise inaccessible constructs, but their underlying mechanisms and challenges differ substantially. In protein NCL, the reaction proceeds via nucleophilic attack of a cysteine thiol on a thioester group in the target peptide, followed by a spontaneous S→N acyl shift to form a native peptide bond – all under near-physiological conditions without requiring protecting groups (Conibear *et al.* 2018). In contrast, RNA chemical ligation relies on distinct chemistries: periodate oxidation cleaves ribose 2',3'-diols to generate dialdehydes, which undergo reductive amination to form morpholino linkages (Kurata *et al.* 2003); and click chemistry (*e.g.*, CuAAC) uses azide-alkyne cycloaddition to generate stable triazole linkages (El-Sagheer and Brown 2010). Unlike NCL, RNA ligation often requires pre-installed functional groups (*e.g.*, azides, alkynes) via solid-phase synthesis, analogous to thioester installation in peptides for NCL.

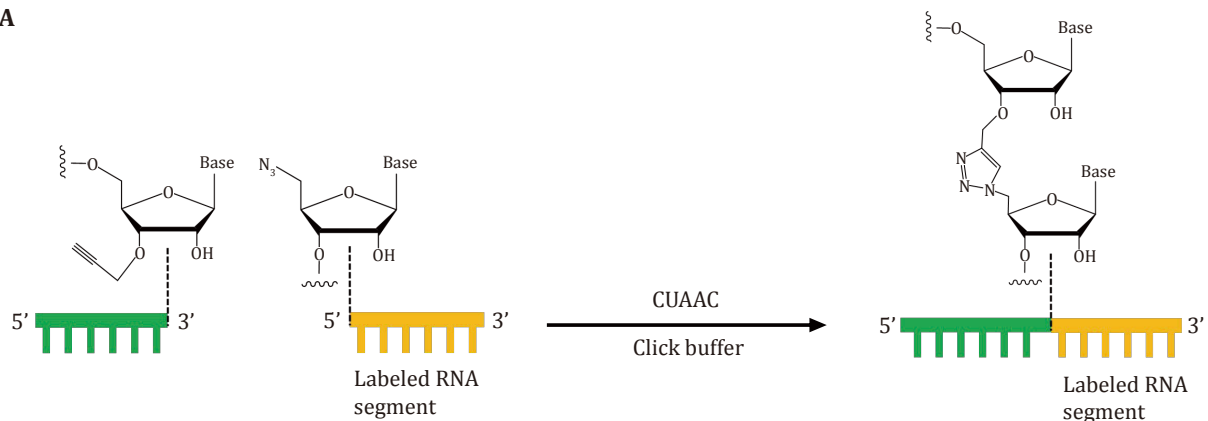
The periodate reaction specifically targets the ribose 2',3'-diol, making it RNA-specific. Wincott and colleagues utilized this method to produce morpholino-linked ribozymes (Bellon *et al.* 1996). They found that ribozymes with morpholino linkages retained normal functionality and offered higher yields and easier purification compared to RNA synthesis (Bellon *et al.* 1996). Thiol-based chemical ligation features the formation of a disulfide bond to covalently link two RNA fragments. Collins *et al.* developed a method to join two RNA strands through disulfide bonds under Mg²⁺-free conditions and applied it to ligate RG ribozymes (Jaikaran *et al.* 2008). Comparing the wild-type RG ribozyme with the ligated RGssG ribozyme, they observed that the disulfide bond had no adverse effect on cis-cleavage kinetics (Jaikaran *et al.* 2008).

Click chemistry proceeds under mild reaction conditions and in suitable solvents, offering high yields with negligible by-product formation. It provides new opportunities for the stable and efficient ligation of RNA fragments (Paredes and Das 2011). Among them, the copper (I)-catalyzed azide-alkyne cycloaddition (CuAAC) reaction, which was discovered independently

by Sharpless and Meldal in 2002, could selectively introduce further modifications like fluorophores, affinity tags (*i.e.* biotin) and so on. As shown in Fig. 3A, two separate RNA fragments, with their 3' and 5' ends modified by alkyne and azide (N₃) group, respectively, are joined through a stable triazole ring formed via a CuAAC click reaction. This method is commonly used to ligate shorter RNA fragments to obtain longer, modified, or segmentally labeled RNA molecules. The non-native triazole bonds introduced by click chemistry do not affect RNA function and can even mimic the natural phosphodiester bond, allowing them to be read through by DNA and RNA polymerases (El-Sagheer and Brown 2015; Paredes and Das 2011). Brown *et al.* used DNA containing triazole linkages as a template for PCR amplification (El-Sagheer and Brown 2012). Their results demonstrated that triazole linkages do not interfere with *in vitro* PCR amplification of DNA. When the linkage site is located within the coding region, RNA products transcribed from the template can also be used for protein expression (El-Sagheer and Brown 2012). For RNA click ligation, El-Sagheer and Brown mediated the binding of 3' alkynes and 5' azides in hammerhead ribozymes by using splints (El-Sagheer and Brown 2010). The click hammerhead ribozyme cleaved its substrate with similar efficiency and specificity to the native ribozyme, though the unnatural triazole linkage is located at the active site. This result proves that the triazole bond, as described in Fig. 3A, is biocompatible. Besides, the ligation of hammerhead ribozymes, approximately 100 nucleotides in length, demonstrated that the production of long RNA through CuAAC chemical ligation is feasible. Frommer *et al.* demonstrated the binding of the flavine mononucleotide (FMN) to the flavine mononucleotide by preparing the ypaA aptamer using CuAAC chemical ligation (Frommer *et al.* 2014). For the 129-nt ypaA aptamer, the click connection holds regardless of whether the azide is at the 5'-terminal of the oligonucleotide or the alkyne is at the 3'-terminal of the oligonucleotide. At the same time, despite having two triazole skeleton bonds, it retains full function in binding its cognate ligand.

In terms of the specific operation of introducing azide and alkyne modifications into RNA, solid phase chemical synthesis is the main pathway. But the azide group is unstable under phosphoramidite conditions. Sanzone *et al.* describe the introduction of azide groups through substitution reactions with NaN₃ after solid phase chemical synthesis (Sanzone *et al.* 2012). H-phosphonate or phosphotriester strategies have been successfully employed. Specially designed phosphoramidates have also been shown to be compatible with

A



B

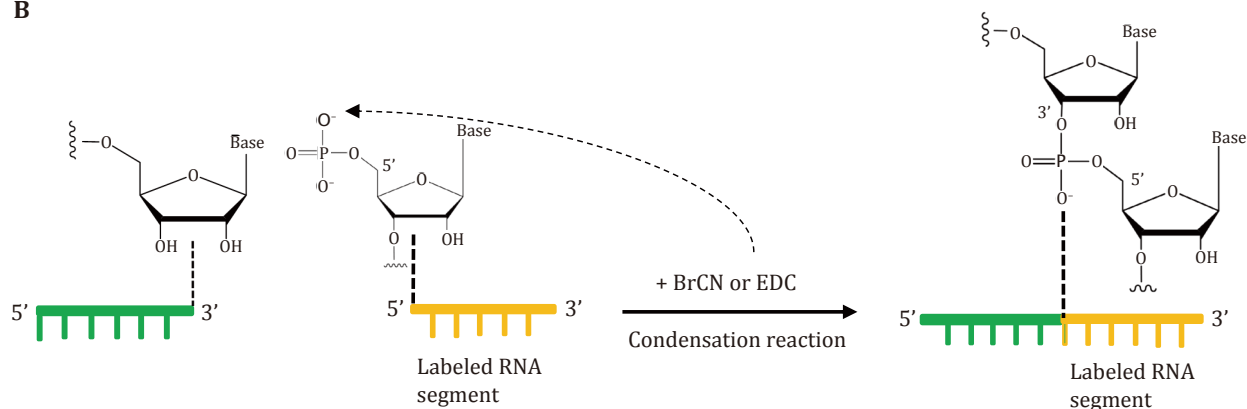


Fig. 3 **A** Click chemistry ligation for RNA segmental labeling. Two separately synthesized RNA fragments, one containing an alkyne-modified nucleotide and the other containing an azide-modified nucleotide, undergo CuAAC reaction to form a triazole-linked segment-labeled RNA. **B** Chemical ligation of the distal ends of linear RNA precursors can be achieved using cyanogen bromide (BrCN) or 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC). Subsequent condensation reactions result in ligation via 3'-5' phosphodiester bonds

azides under modified solid-phase synthesis conditions in this way (Müggenburg and Müller 2022). Obviously, as long as the azido modification is not part of the same molecule as the phosphoramidite group, the azide remains stable during oligonucleotide synthesis. Therefore, 3'-terminal azide-modified oligonucleotides can usually be obtained by solid-phase synthesis of standard phosphoramidites. In contrast, the alkyne group can be easily incorporated into the 5' or 3' end of RNA through solid-phase synthesis (Paredes and Das 2011). Nevertheless, the introduction of azide and alkynyl groups still faces the length limitation of solid-phase chemical synthesis itself. Due to the cytotoxicity of copper ions, CuAAC is limited to *in vitro* RNA modification and ligation (Agard *et al.* 2006). The development of the strain-promoted azide-alkyne cycloaddition (SPAAC) reaction, also known as Cu-free click chemistry, has provided a solution to this issue. SPAAC was pioneered by Bertozzi and used for the

covalent modification of biomolecules in living systems. Examination of the click reactions of alkynes with oligonucleotides derivatized at the 5' terminus, 3' terminus, and at an internal position proves the feasibility of RNA ligation through SPAAC. For instance, Nainar *et al.* reported the use of the SPAAC method for temporary labeling of nascent RNA in living cells (Nainar *et al.* 2017). Therefore, SPAAC presents a promising strategy for adapting RNA ligation for use within living systems.

RNA segment-selective labeling is not limited to small molecular tags; it also enables the incorporation of nanoscale labels such as gold nanoparticles (AuNPs) and quantum dots (QDs), which offer unique optical, electronic, and structural properties. AuNPs, with their high extinction coefficients and tunable surface plasmon resonance, enhance sensitivity in RNA imaging and sensing (Kumalasari *et al.* 2024). QDs, characterized by photostability and size-tunable emission,

enable long-term tracking of RNA dynamics in live cells (Jaiswal *et al.* 2003). Incorporating these nanoscale labels via segmental labeling requires site-specific conjugation to RNA segments, often using click chemistry or thiol-mediated ligation (Sanzone *et al.* 2012). For example, AuNPs functionalized with azide groups can be ligated to alkyne-modified RNA segments via CuAAC, ensuring precise attachment to target regions. QDs conjugated to oligonucleotides complementary to a labeled RNA segment can hybridize selectively, leveraging segmental labeling for spatial control. These labels expand applications into single-molecule imaging, targeted drug delivery, and ultrasensitive biosensing, underscoring the versatility of RNA segmental labeling.

The chemical ligation of RNA can also be achieved using cyanogen bromide (BrCN) or a 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) as activation agents. These reagents promote the activation of phosphate groups on RNA to react with nearby hydroxyl groups, facilitating the formation of phosphodiester bonds through condensation reactions (Paredes *et al.* 2011). However, despite the theoretical efficiency of this method, it is significantly limited in practice due to its low ligation efficiency when applied to RNA. To address this inefficiency, splints are often employed to stabilize the alignment of RNA fragments and enhance the likelihood of successful ligation (Dolinnaya and Shabarova 1996). Another drawback of this approach is the unintended formation of 2'-5' phosphodiester linkages, which deviate from the natural 3'-5' linkages in RNA and may compromise the functionality of the ligated product (Liu *et al.* 2016). Fig. 3B demonstrates the stepwise process, highlighting the reaction mechanism and the resulting ligation at the distal ends of RNA precursors.

LIGASE-BASED METHODS

Compared to the previously mentioned chemical methods, ligase-based methods offer several advantages, including simplicity of operation, high ligation efficiency, and the formation of natural phosphodiester bonds that can be recognized by all nucleic acid processing enzymes. Among the ligases capable of ligating RNA, the three most widely used enzymes are derived from the genome of bacteriophage T4, namely T4 DNA ligase (T4 Dnl), T4 RNA ligase 1 (T4 Rnl1), and T4 RNA ligase 2 (T4 Rnl2) (Pascal 2008). In addition, some DNA-based deoxyribozymes and RNA-based ribozymes also possess the function of catalyzing RNA ligation.

T4 ligase

T4 DNA ligase

T4 Dnl utilizes adenosine triphosphate (ATP) as a cofactor, can ligate the ends of sequences at nicks in double-stranded DNA (dsDNA), double-stranded RNA, or DNA-RNA hybrid strands. The catalytic mechanism of T4 DNA ligase consists of three steps. First, the ϵ -amino group of a highly conserved Lys residue in the ligase performs a nucleophilic attack on the phosphate ester bond of the cofactor ATP or nicotinamide adenine dinucleotide (NAD⁺), forming a covalent ligase-AMP intermediate and releasing pyrophosphate (PPi) or nicotinamide mononucleotide (NMN). Subsequently, the AMP moiety is transferred from the ligase-AMP intermediate to the 5' phosphate group of the nicked DNA, resulting in the formation of an adenylated AMP-DNA intermediate. Finally, the 3' hydroxyl group of the opposite strand attacks the adenylated phosphate, leading to the formation of a phosphodiester bond and the release of AMP, thereby completing the ligation reaction, and joining the two nucleic acid strands (Williamson and Leiros 2019). T4 Dnl requires the assistance of a bridge strand to connect the unlabeled RNA fragment and the labeled RNA fragment (Fig. 4A). The labeled RNA fragment can be obtained through *in vitro* transcription or chemical synthesis. The RNA synthesized by *in vitro* transcription typically has a 5' triphosphate group. Therefore, before ligation, RNA pyrophosphohydrolase (RppH) is used to remove the pyrophosphate from the 5' end of the triphosphorylated RNA, leaving a 5' monophosphate RNA. It is important to note that the splint and RNA sequence must ensure absolute base complementarity to avoid any gaps at the ligation site (Lohman *et al.* 2011). Upon completion of the ligation, deoxyribonuclease (DNase) can be used to remove the DNA splint.

Allain's group utilized T4 DNA ligase to connect RNA fragments modified with 4-thiouridine and unmodified RNA fragments, thereby producing large RNA molecules containing two spin labels (Duss *et al.* 2014). Subsequently, they measured the long-range distances within these large RNA molecules using EPR spectroscopy. The researchers explored two strategies: first, ligating the RNA fragments followed by spin labeling; and second, spin labeling the fragments prior to ligation. By employing T4 DNA ligase, Allain's group successfully ligated multiple RNA fragments to generate the desired large spin-labeled RNA. In this study, the yield of the ligation reaction using T4 DNA ligase typically ranged between 20%–40% (Duss *et al.* 2014).

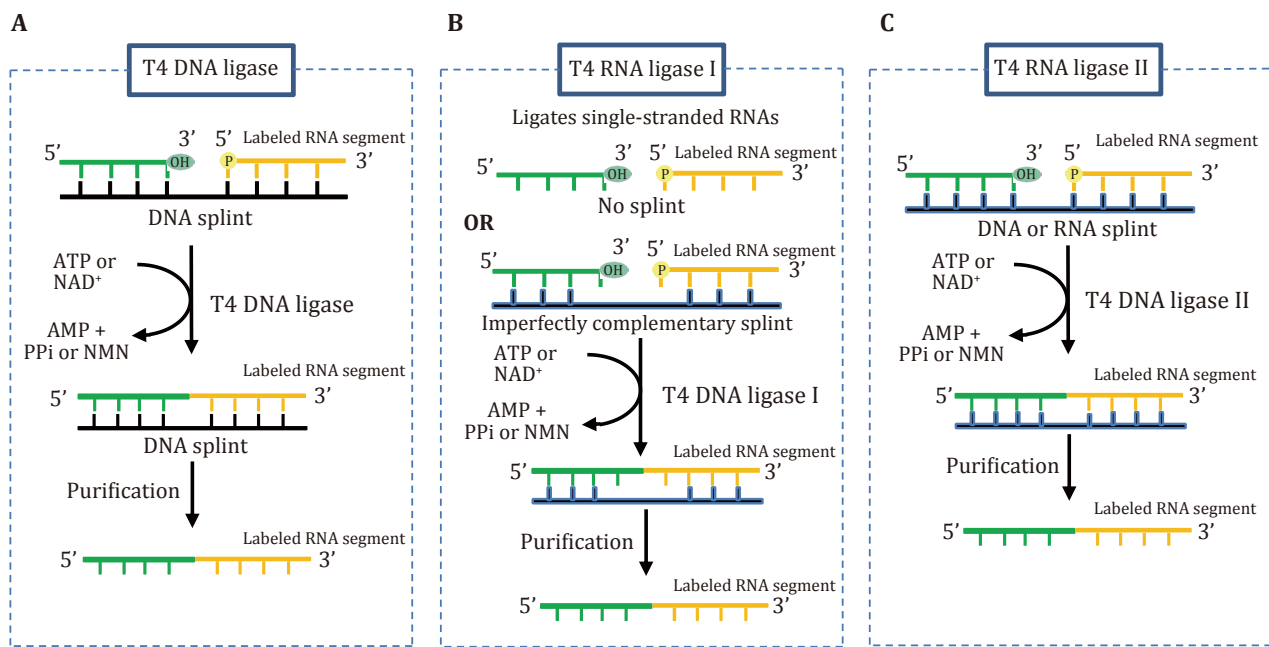


Fig. 4 T4 ligase method for ligating segment-labeled RNA. **A** T4 DNA Ligase mediated ligation of two RNA strands. **B** T4 RNA Ligase I mediated ligation of two RNA strands. **C** T4 RNA Ligase II mediated ligation of two RNA strands

Chen *et al.* employed T4 Dnl with a splint to synthesize isotope-labeled C-C chemokine receptor type 5 (CCR5) RNA fragments (Chen *et al.* 2018). The CCR5 fragment is 96 nucleotides in length and contains a pseudoknot domain. The researchers conducted NMR experiments on the labeled CCR5 to measure the exchange rates and populations of different conformational states. By synthesizing specifically labeled CCR5 RNA fragments using T4 Dnl, this study overcame the challenge of spectral overlap often encountered in NMR analyses of large RNA molecules. The experiments demonstrated that the interaction between A90-C1' of the CCR5 pseudoknot and microRNA-1224 stimulates-1 ribosomal frameshifting, which is crucial for RNA-RNA molecular recognition.

T4 RNA ligase

T4 RNA ligase, an RNA end-joining enzyme involved in RNA repair, splicing, and editing processes, belongs to the same covalent nucleotidyl transferase superfamily as T4 DNA ligase (Ho *et al.* 2004). The catalytic mechanism of T4 RNA ligase is highly similar to that of T4 DNA ligase, requiring ATP as a cofactor. It facilitates the formation of a phosphodiester bond between the 5'-phosphate end of the donor RNA and the 3'-hydroxyl end of the acceptor RNA through a classic three-step reaction analogous to DNA ligase, thereby joining the two RNA segments together (Nandakumar *et al.* 2006).

T4 Rnl1 primarily catalyzes the ligation of single-stranded nucleic acids, including single-stranded RNA, single-stranded RNA-DNA hybrids, and single-stranded DNA. T4 Rnl1 exhibits the highest ligation efficiency for RNA-RNA ligation, followed by RNA-DNA ligation, with the lowest efficiency observed for DNA-DNA ligation. Consequently, T4 Rnl1 is seldom used for ligating single-stranded DNA (Middleton *et al.* 1985). Due to its preference for single-stranded RNA substrates, the enzyme does not require the addition of a splint during the ligation process (Fig. 4B). If the single-stranded ends of the donor and acceptor RNA molecules cannot be brought into close proximity, the catalytic efficiency of the enzyme is significantly reduced. In such cases, a splint that is complementary to both the donor and acceptor RNA sequences can be designed (Fig. 4B). It is important to note that the splint sequence should not be fully complementary, and the RNA splint should retain 2–3 unpaired nucleotides at each end, while the DNA splint should retain a longer single-stranded nucleotide sequence. When it comes to the ligation reaction, T4 Rnl1 exhibits preferences for the nucleic acid ends of its substrates. The ligation efficiency for the 3' hydroxyl end bases is in the order of A > C > G > U, while the efficiency for the 5' phosphate groups is in the order of pA > pG ≈ pC > pU (McLaughlin *et al.* 1982). Wang *et al.* constructed branched oligonucleotides containing dual cap structures and incorporated them into the 5' end of messenger RNA (mRNA) using T4

Rnl1 to form multi-capped mRNA. Additionally, they systematically screened modifications to the 5' untranslated region (UTR) and 3' UTR, using a T4 Rnl1-based segment ligation method. Ultimately, combining the dual-cap topology engineering and segmental chemical modifications of the mRNA 5' UTR and 3' UTR, they achieved nearly a tenfold increase in peak protein expression levels of the luciferase reporter mRNA (Chen *et al.* 2025). Stark *et al.* designed a splint structure that maintains the nucleotide splice junctions in a single-stranded state, enabling near-complete completion of multiple ligation reactions, typically within 30 min. Using T4 Rnl1, the researchers successfully generated a 128-nt RNA from three re-synthesized RNA fragments, demonstrating the potential of this method for constructing larger RNA molecules (Stark *et al.* 2006).

T4 Rnl2 is an ATP-dependent nucleic acid ligase discovered and characterized by Ho *et al.* in 2002 (Ho and Shuman 2002). It shares high homology with T4 Rnl1 but exhibits distinct substrate specificity. In addition to its activity in catalyzing the ligation of single-stranded RNA and single-stranded DNA, T4 Rnl2 demonstrates higher catalytic activity and ligation efficiency when dealing with nicked double-stranded RNA substrates. Therefore, when using T4 Rnl2 for RNA ligation, a splint is required as an auxiliary, and this splint must be perfectly complementary to the RNA chain ends (Fig. 4C). Lebars *et al.* designed a natural RNA fragment and an RNA fragment modified with 6-thioguanosine (Lebars *et al.* 2014). They then used T4 Rnl2, assisted by a complementary 43-nt DNA splint, to ligate the two RNA fragments, forming a full-length RNA ligation product containing specific thio-modified nucleotides. The ligation product was subsequently purified, with an estimated ligation yield of 50%. This enzymatic method can be used for site-specific spin labeling of long RNA, with yields sufficient for NMR studies (Lebars *et al.* 2014). Combined with isotopic labeling, this method holds significant potential for investigating the structure and dynamics of RNA complexes. Manz *et al.* constructed five RNA fragments labeled with fluorescence resonance energy transfer (FRET) dyes, which were annealed with DNA splints and then ligated using T4 Rnl2 to obtain a full-length 169-nt S-adenosyl-L-methionine (SAM)-I riboswitch (Manz *et al.* 2017). Using single-molecule FRET microscopy, the research team identified four conformational states of the SAM-I riboswitch under different conditions. The conformational energy landscape of the SAM-I riboswitch was influenced by Mg^{2+} and SAM ligand concentrations, and the dynamic changes induced by ligand binding were crucial for its gene regulatory function (Manz *et al.* 2017).

Deoxyribozyme

Deoxyribozymes, also known as DNAzymes, are synthetic DNA molecules selected *in vitro* through directed evolution for their catalytic functions. First discovered by Breaker and Joyce in 1994 (Breaker and Joyce 1994), these DNA molecules were found to catalyze the cleavage of RNA phosphodiester bonds. Since then, hundreds of deoxyribozymes have been identified, expanding the range of reactions catalyzed by DNAzymes. Among their various applications, RNA ligation is a significant function due to its potential in studying RNA structure and function. In 2003, Silverman *et al.* first reported a deoxyribozyme with RNA ligase activity. However, this enzyme could only form non-native 2'-5' phosphodiester bonds (Flynn-Charlebois *et al.* 2003). In 2005, Silverman further identified DNAzymes capable of forming native 3'-5' linkages, named 9DB1 and 7DE5 (Purtha *et al.* 2005). Subsequently, through directed evolution based on 9DB1, a shorter and more catalytically efficient deoxyribozyme, 9DB1*, was developed (Wawrzyniak-Turek and Höbartner 2014). The 9DB1* consists of a central catalytic domain of 31-nt flanked by variable binding arms (Fig. 5). These binding arms act like a splint, maintaining the proper orientation and proximity of the RNA fragments to ensure correct substrate alignment for the catalytic reaction. Upon binding to the RNA substrate, 9DB1* folds into a tertiary structure, bringing the 3'-OH group of one RNA fragment (the acceptor) close to the 5'-triphosphate group of another RNA fragment (the donor) (Höbartner 2019). This proximity facilitates a nucleophilic attack of the 3'-OH on the 5'-triphosphate, a reaction catalyzed by metal ions, typically divalent cations such as Mg^{2+} or Mn^{2+} (Aranda *et al.* 2019). This attack results in the formation of a new 3'-5' phosphodiester bond, thus producing a covalently linked RNA product (Fig. 5). 9DB1* has sequence restrictions at the ligation site: the 3' end of the acceptor can be any nucleotides except cytidine (C), while the 5' end of the donor RNA must be either 5'-GA or 5'-AA (Wawrzyniak-Turek and Höbartner 2014).

Höbartner *et al.* successfully utilized 9DB1* to ligate RNA molecules ranging from 26 to 118-nt in length, including two spin-labeled SAM riboswitch variants: SAM-III (53-nt) and SAM-I (118-nt) (Wawrzyniak-Turek and Höbartner 2014). The 5'-triphosphate donor RNA was prepared by *in vitro* transcription (IVT), while the labeled acceptor RNA was synthesized using solid-phase methods. Under the catalysis of 9DB1*, these RNAs were ligated to form full-length SAM-III and SAM-I, achieving yields of up to 70%. Compared to the traditional T4 Dnl ligase, 9DB1* catalyzed ligation does

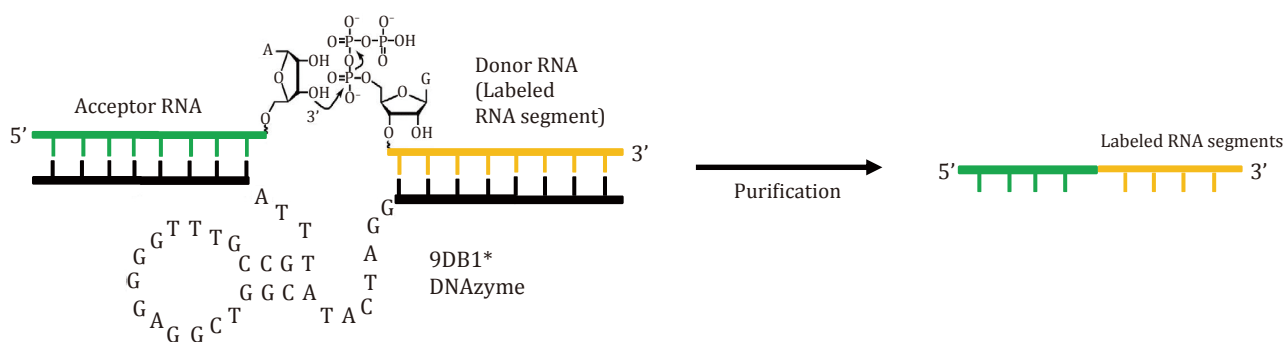


Fig. 5 9DB1* can catalyze nucleophilic attack of the 3'-hydroxyl group on the 5'-triphosphate by divalent metal ions (Mg^{2+} or Mn^{2+}), forming a natural phosphodiester bond between two RNA strands. The 9DB1* DNA enzyme consists of a central catalytic domain with two arms on either side. These arms hybridize with the RNA substrate through base pairing, leaving only two unpaired A and G at the junction site

not require a 5'-monophosphate group, and it yields higher efficiency with fewer byproducts. Thus, 9DB1* provides a new and more reliable tool for EPR analysis of long modified RNAs and has been successfully applied to the study of the dynamics and secondary structure of SAM riboswitch domains (Wawrzyniak-Turek and Höbartner 2014).

Ribozyme

In addition to deoxyribozymes, there are also ribozymes in nature that can catalyze RNA ligation reactions (Lilley 2005). Through directed evolution and rational design, scientists have engineered these natural ribozymes to obtain variants with higher catalytic efficiency and greater substrate specificity (Bartel and Szostak 1993). Müller's group pioneered the design and construction of a twin ribozyme for RNA labeling (Welz *et al.* 2003). This twin ribozyme is created by tandemly fusing two hairpin ribozymes, which can excise specific regions of the target RNA and replace them with synthetic RNA sequences. The hairpin ribozyme, serving as the core component of the twin ribozyme, is a naturally occurring small catalytic RNA capable of reversible cleavage and re-ligation of RNA fragments. While the hairpin ribozyme exhibits high flexibility in substrate sequence recognition, a guanine nucleotide must be present at the 3' position of the cleavage/ligation site. The cleavage reaction occurs through the 2'-OH group of the nucleotide attacking the adjacent phosphodiester bond, resulting in a 2',3'-cyclic phosphodiester bond at one fragment and a free 5'-OH group at the other. Re-ligation is facilitated by the free 5'-OH group attacking the 2',3'-cyclic phosphodiester, forming a 3',5'-phosphodiester bond and a free 2'-OH (Ivanov *et al.* 2005). Müller's group fused two hairpin

ribozymes together, each flanked by two recognition arms complementary to the target RNA sequence. One arm also serves as a linker, covalently connecting the two ribozymes. Upon annealing to the target RNA, each hairpin ribozyme cleaves at the corresponding cleavage site. The linking arm, designed with non-paired tetraloops, destabilizes the cleaved products. Subsequently, the labeled RNA fragment to be inserted replaces the excised fragment (Fig. 6).

The twin ribozyme has been used to replace a 16-nt segment in an *in vitro* transcribed 145-nt RNA with a 20-nt labeled RNA fragment. The labeled RNA fragment was prepared using solid-phase synthesis and incorporated an amino-modified deoxythymidine unit. The introduced amino modification was functionalized with various fluorescent dyes (such as fluorescein, TAMRA, Cy5) and affinity probes (such as biotin). Under reaction conditions at 37°C, the yields were approximately 11% and 18%. Increasing the temperature to 55°C led to a higher yield; however, the optimal reaction conditions were achieved at 47°C, where up to 53% yield of the dye-labeled 145-mer RNA was obtained (Vauléon *et al.* 2005). The use of ribozyme ligation overcomes the length limitations of RNA solid-phase synthesis and can be applied to construct RNA probes, RNA aptamers, and other constructs.

In summary, ligase-based methods provide a versatile toolkit for RNA segmental labeling, each with distinct advantages and ideal application scenarios. T4 DNA ligase is highly effective for ligating nicks in double-stranded RNA/DNA hybrids and is widely used for its ability to generate natural phosphodiester bonds with the aid of a splint, making it a mainstream choice for preparing large RNAs for biophysical studies like NMR and EPR. T4 RNA ligase 1 excels at joining single-stranded RNA ends without a splint, offering simplicity

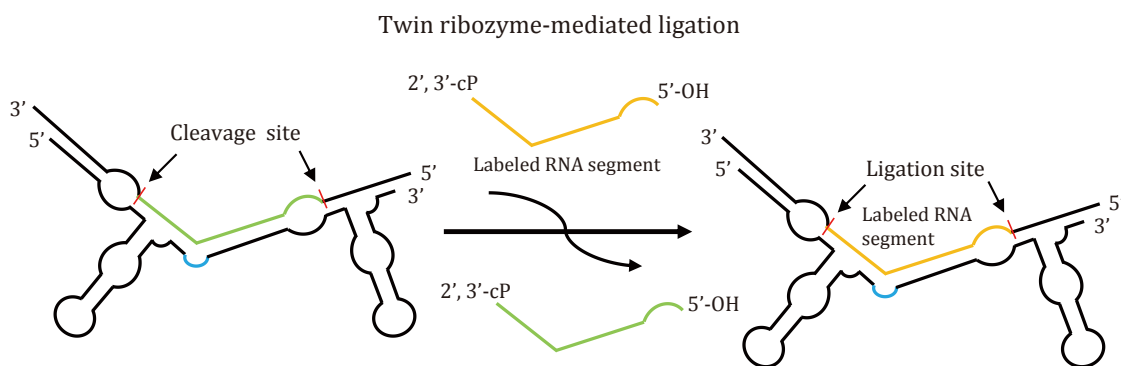


Fig. 6 A schematic diagram of RNA segmental labeling using a twin ribozyme. The labeled RNA fragments are used as substrates, and through two consecutive cleavage and ligation events, the corresponding segments in the target RNA are replaced

for constructing chimeric RNAs like multi-capped mRNAs, though its efficiency is influenced by terminal nucleotides. T4 RNA ligase 2 is particularly valuable for its high efficiency with splinted, nicked double-stranded RNA substrates, facilitating the incorporation of specific modifications for structural biology. Deoxy-ribozymes (*e.g.*, 9DB1*) offer a purely chemical and splint-free alternative with high ligation efficiency and no requirement for a 5'-monophosphate, but their application is constrained by specific sequence requirements at the junction site. Conversely, ribozymes (*e.g.*, the twin ribozyme) enable the precise replacement of internal RNA segments through cleavage-religation mechanisms, overcoming length limitations of solid-phase synthesis, though often with moderate yields. The choice among these methods ultimately depends on the experimental requirements, including the RNA structure (single-stranded vs. double-stranded at the junction), the need for natural linkages, sequence constraints, desired yield, and the specific application, from structural analysis to the construction of functional RNA constructs.

POLYMERASE-BASED METHOD

Segmental labeling of RNA using ligase methods is a type of post-transcriptional labeling, which often requires additional steps that may result in RNA loss or contamination. This method is not suitable when large amounts of labeled RNA or precise control over the type, and location of the label is required. Transcription mediated by engineered DNA polymerase and T7 RNA polymerase allows direct labeling during synthesis. This avoids additional steps, reduces the possibility of RNA loss or degradation, and can accommodate various RNA sequences and label types.

DNA polymerase

DNA polymerases are generally considered to be used for synthesizing DNA rather than RNA. They synthesize DNA from a DNA template by incorporating deoxyribonucleotides (dNTPs) and lack the ability to process ribonucleotides (rNTPs), making them unable to directly produce RNA. However, through engineering of thermophilic DNA polymerases, both the SFM4-3 polymerase from the A-family of *Thermococcus aquaticus* (Taq) and the TGK polymerase from the B-family of *Thermococcus gorgonarius* have acquired the ability to synthesize RNA with modified bases (Brunderová *et al.* 2024). Both the enzymes SFM4-3 and TGK can extend RNA from a DNA template. This process does not require a specific RNA initiation sequence, offering a new option for the synthesis of RNA sequences that do not start with G (Christensen *et al.* 2022; Wen *et al.* 2024). The initial RNA Segment 1 can be transcribed *in vitro* by T7 RNA polymerase or synthesized chemically. To obtain long RNA sequences that do not start with G, the following steps can be followed. First, perform a template annealing step at 95°C, allowing the DNA primer to bind to the DNA template. Then, use TGK or SFM4-3 polymerase to extend RNA from the 3' end of the primer. Finally, remove the template with DNase and purify the transcribed RNA. In the segment-selective modification step, RNA Segment 1 first anneals with a new DNA template, allowing the RNA fragment to bind to the DNA template. Next, extend the RNA using TGK or SFM4-3 in the presence of a specific rNTPs pool. Finally, remove the template again and purify the RNA containing the labeled segment. If multiple segments labeled RNAs are required, anneal with a new DNA template and repeat the process (Fig. 7).

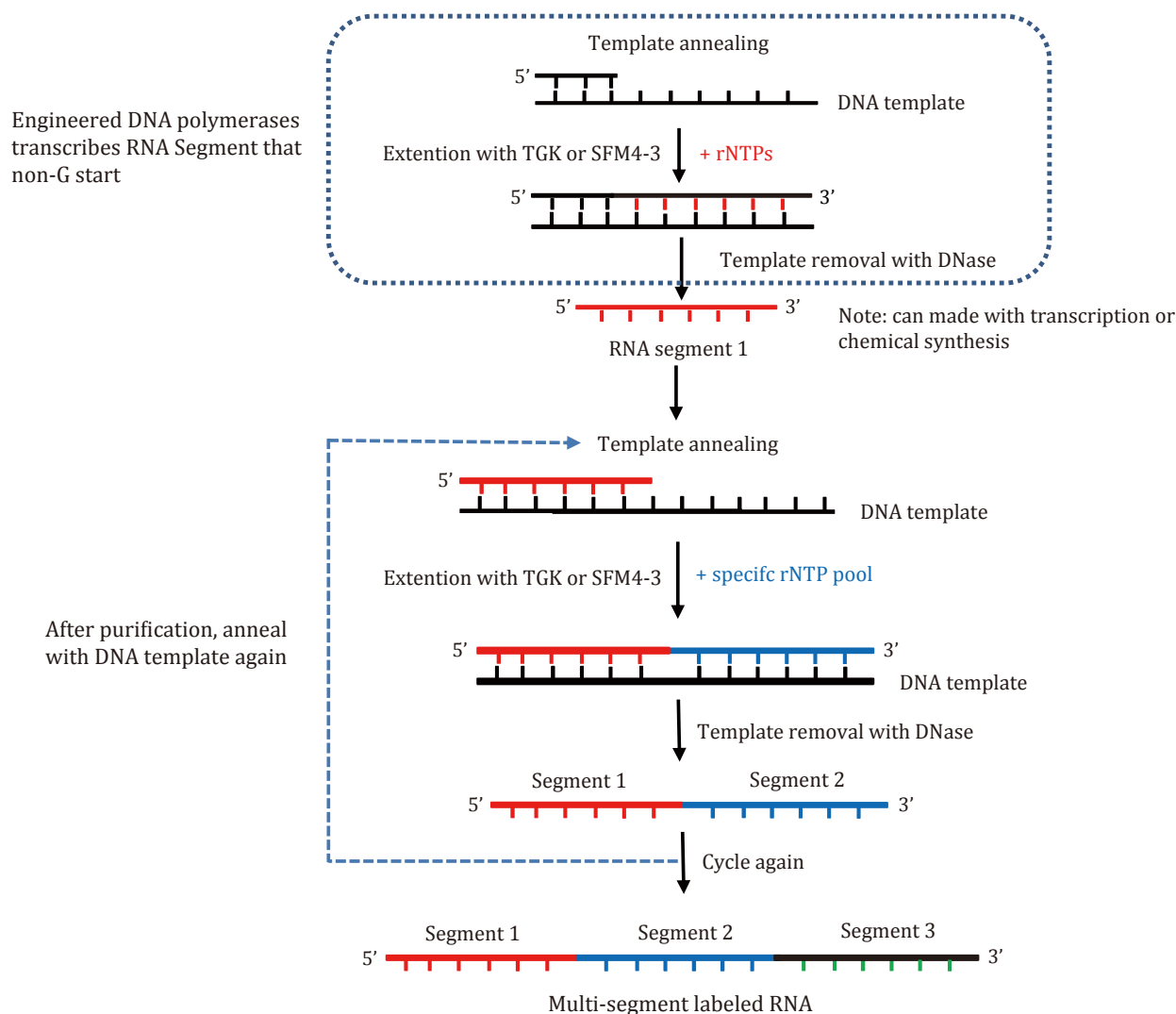


Fig. 7 DNA polymerase-based RNA segmental labeling. RNA segment 1 (red) anneals with the DNA template (black) and is extended using a specific rNTPs pool to generate segment 2 (blue). After removing the template with DNase, RNA of segment 1 + 2 can undergoes another round of extension to generate segment 3 (green). If RNA segment 1 requires a non-G start, an engineered DNA polymerase can be used to transcribe the RNA segment from a DNA template

Haslecker and colleagues utilized a TKG-based labeling scheme (SegModTeX) to construct multi-segment RNA, including two different multi-segment labeling experiments on 7SK snRNA (324-nt) (Haslecker *et al.* 2023). One experiment involved labeling nucleotides from positions 181 to 253, while fully deuterating the segments at both ends (7SK snRNA_{180/253}). The other experiment made five adenosines between positions 149 and 178 visible, while fully deuterating the other bases (7SK snRNA_{148/178}). After labeling selective segments of 7SK snRNA using this method, the NMR spectra were significantly simplified, enhancing the

efficiency of NMR in structural analysis of multi-domain RNAs. This provided a powerful tool for in-depth studies of RNA structure and dynamics. SegModTeX can also introduce various biologically relevant modified NTPs, such as N⁶-methyladenosine (m⁶A), pseudouridine (Ψ), and 2-¹⁹F-ATP, further broadening its application scope (Haslecker *et al.* 2023). Hocek *et al.* designed a series of base-modified rNTPs (rN^XTPs) and tested the incorporation efficiency of rN^XTPs by SFM4-3 and TKG (Brunderová *et al.* 2024). They found that TKG outperformed SFM4-3 in terms of incorporation efficiency, activity, and fidelity. The group then used an

adenine riboswitch as a model, introducing two fluorescent labels into the loop-1 and loop-2 regions using T7 RNA polymerase (T7 RNAP), and synthesized a FAM-Cy5-Cy3 triple-labeled riboswitch. They studied RNA conformational changes using FRET, and the results showed expected conformational changes in the riboswitch arms. Finally, they used T7 RNAP to synthesize mRNAs with various modifications, including unmodified, fully modified, gene-modified, and single-site modified mRNAs. *In vitro* and *in vivo* translation experiments demonstrated that mRNAs with a single-site 5-methylcytosine modification enhanced translation efficiency, while fully modified mRNAs hindered translation (Brunderová *et al.* 2024). These results highlight the potential applications of this method in RNA biochemical research, biotechnology, and the production of modified RNA.

T7 RNA polymerase

In vitro transcription mediated by T7 RNA polymerase (T7 RNAP) can generate multiple RNA copies and ultralong sequences from a single dsDNA template molecule (Beckert and Masquida 2011; Pokrovskaya and Gurevich 1994). During IVT, specific chemically labeled rNTPs can be added to introduce chemical modifications into the RNA product. However, the incorporation sites of these labels are random, making it challenging to achieve labeling in specific regions. To address this issue, a technique called position-selective labeling of RNA (PLOR) was developed (Liu *et al.* 2015, 2018). This method enables the synthesis of site- and region-specific labeled RNA through IVT. The synthesis of RNA using PLOR involves three main stages: initiation, extension, and termination. Transcription occurs on a 5'-Biotin-DNA template immobilized on streptavidin agarose beads (Fig. 8). The 5'-Biotin-DNA template contains a T7 promoter sequence. A spacer of 15–30 bp is placed between the biotin and the T7 promoter sequence to eliminate the steric hindrance caused by the beads, improving transcription efficiency. In the initiation or extension phase, adding three or fewer rNTPs in a single cycle causes the transcription process to pause due to the lack of the necessary rNTP. During the initiation phase, it is essential to ensure that RNA elongation reaches at least 9 bases, with the optimal range being 11 to 13 bases, before the first polymerase pause occurs. This step is crucial for forming a stable elongation complex (RNA polymerase-DNA-RNA ternary complex). Only in this way can it be ensured that the complex remains intact when the remaining rNTPs are washed away in subsequent steps. Solid phase extraction is then repeatedly performed to wash the

beads, thoroughly removing rNTPs from the elongation complex. During the extension phase, new rNTPs need to be added, which can be either unlabeled or labeled. Since one specific rNTP is still missing during this process, the polymerase will pause at another position in the elongation process. Subsequently, the washing step is repeated. This cycle of extension and stopping can be repeated multiple times until all labeled rNTPs are successfully incorporated into the designated regions. After transcription is complete, the transcript dissociates from the complex and can be collected along with the liquid phase. Following thorough cleaning, the DNA template and T7 RNAP can be recovered for reuse in the next PLOR cycle. This approach allows for higher RNA yield, as only one transcript is produced per template and cycle.

Wang *et al.* successfully synthesized a 71-nt aptamer domain and a 104-nt turnip crinkle virus RNA element using PLOR (Liu *et al.* 2015, 2018). In these synthesized products, different segments or nucleotides were specifically labeled with isotope-containing nucleotides or nucleotides with fluorescent or chemical modifications. The segmental labeling of the 71-nt RNA significantly improved the issue of NMR spectral overlap (Liu *et al.* 2015). Liu *et al.* optimized the PLOR technique and found that the first six nucleotides are crucial for the synthesis of RNA transcripts based on PLOR and determined that 5'-GGG is the ideal starting sequence. PLOR can precisely incorporate rNTPs with various modifications or labels, accepted by T7 RNAP and its engineered variants, into specific segments or sites. Recently, there have been reports of selectively introducing 5'-iodouridine into riboA71 using PLOR to assist crystallization (Stagno *et al.* 2019).

SUMMARY AND PERSPECTIVES

Segmental labeling of RNA is crucial for studying RNA structure and function. Through the efforts and innovations of scientists, significant advancements have been made in the development of RNA segmental labeling methods. However, each of the current segmental labeling techniques has its own strengths and weaknesses, so selecting a method for synthesizing RNA with segmental labels requires careful consideration based on specific synthesis needs and application scenarios. To facilitate the selection of an appropriate method for a given experimental goal, the key characteristics of the techniques discussed herein – including applicable RNA length, yield, advantages, and disadvantages – are summarized in Table 1.

The main advantage of chemical synthesis is the high

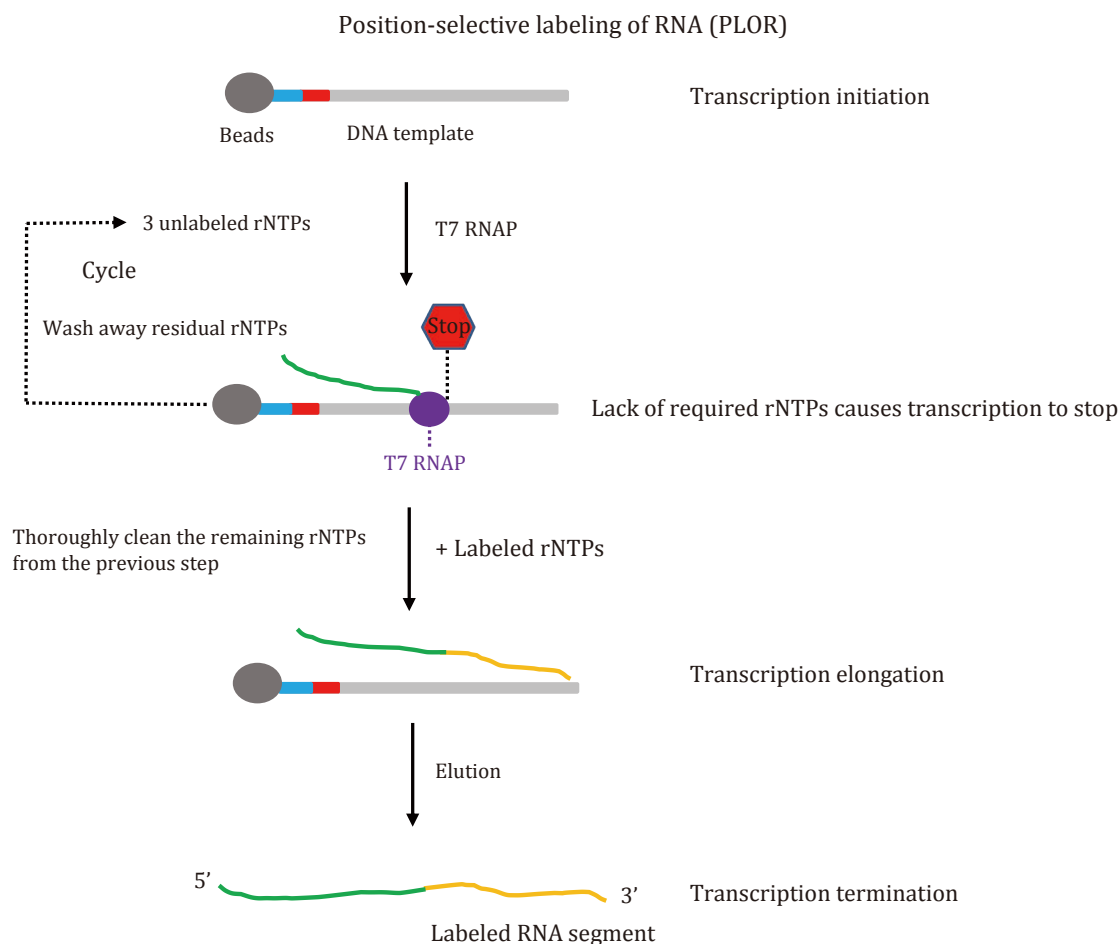


Fig. 8 PLOR-based RNA segmental labeling. PLOR consists of three stages: initiation, extension, and termination. Transcription occurs on a DNA template immobilized on beads. In a single cycle of the initiation or extension phase, three or fewer rNTPs are added, causing transcription to pause due to the absence of the required rNTP. After removing residual rNTPs and introducing labeled rNTPs, extension resumes, generating segment-labeled RNA

purity of the resulting RNA, and there is a wide range of commercial services that specialize in custom synthesis. However, the length of the produced RNA molecules is typically limited to 70–80 nucleotides; beyond this range, the yield significantly decreases, and the cost increases. Post-synthetic chemical ligation enables the preparation of long RNA molecules with segment-selective modifications from smaller RNA fragments. However, it introduces non-natural linkages and involves the use of highly toxic reagents, raising chemical safety concerns. Although the triazole bond produced by CuAAC can mimic the phosphodiester bond and does not interfere with RNA functionality, click chemistry linkers often need to be introduced through chemical synthesis, and their ligation efficiency is relatively low. Compared to chemical synthesis, ligase methods can generate larger linear RNA molecules. Unlike some chemical ligation methods, ligase methods

do not cause damage to RNA molecules and do not introduce toxic substances. The T4 ligase method is currently the most mainstream method for RNA ligation. T4 ligase can form phosphodiester bonds between the 3' and 5' ends of RNA and is applicable for linking RNA fragments of various lengths and sequences. However, this method has some drawbacks. The drawbacks of this method are that the 5' end requires a monophosphate group, a splint is needed, and ligation efficiency may be affected by the length, sequence, and structure of the RNA, resulting in low efficiency. Additionally, the low enzyme turnover rate and high cost are challenges. The use of deoxyribozymes and ribozymes for ligation is less widespread and is associated with challenges such as difficulties in ligating large RNA molecules and low ligation efficiency. DNA polymerase-driven labeling methods are notable for their ability to use multi-step extension to integrate

Table 1 Comparison of methods for RNA segmental labeling

Method category	Method name	Applicable RNA length and yield	Advantages	Disadvantages
Chemical methods	Solid-phase synthesis	<ul style="list-style-type: none"> • ≤70–80 nt • Decreases with length (>50 nt: low) 	High purity; Ultimate flexibility in modification; Commercially available	Low yield for long RNAs; Requires protected phosphoramidites
	Post-synthetic chemical ligation	<ul style="list-style-type: none"> • >80 nt (assembled from shorter fragments) • Variable (CuAAC: moderate; BrCN/EDC: low) 	Extends length beyond solid-phase limits; Access to long modified RNAs	Non-native linkages; Toxic reagents (<i>e.g.</i> , Cu)
Ligase-based methods	T4 DNA ligase	<ul style="list-style-type: none"> • >80 nt • 20%–40% 	Forms natural phosphodiester bonds; Suitable for large RNAs	Requires 5'-monophosphate & splint; Sensitive to RNA structure; Low turnover
	T4 RNA ligase 1	<ul style="list-style-type: none"> • >80 nt • High (with optimal terminal bases) 	No splint needed for ssRNA; High efficiency for RNA-RNA ligation	Efficiency depends on terminal bases (3' end: A>C>G>U); Low for DNA-DNA ligation
	T4 RNA ligase 2	<ul style="list-style-type: none"> • >80 nt • High (with optimal terminal bases) 	High activity for dsRNA nicks	Requires perfectly complementary splint; No ssRNA activity
	Deoxyribozyme (9DB1*)	<ul style="list-style-type: none"> • 26–118 nt • Up to 70% 	No 5' monophosphate required; Fewer byproducts than T4 ligases	Sequence restrictions (donor 5' end: GA/AA; acceptor 3' end: not C)
	Ribozye (Twin)	<ul style="list-style-type: none"> • >80 nt • Up to 53% (at 47°C) 	Can replace internal segments; No external ligase needed	Complex design and preparation require G at ligation site; Limited to specific RNA sequences
Polymerase-based methods	Engineered DNA Polymerase (TGK/SFM4-3)	<ul style="list-style-type: none"> • >80 nt • Efficient (consistent with template) 	No sequence restrictions; Introduces modified NTPs (m ⁶ A, Ψ); Multi-segment labeling	Length limited by DNA template; One transcript per template
	T7 RNAP (PLOR)	<ul style="list-style-type: none"> • >80 nt • Moderate (improves with cycles) 	Can produce long, multi-segment labeled RNA; Template and enzyme are reusable	Complex protocol; high cost (labeled rNTPs, washing); Requires technical expertise

large numbers of modified rNTPs into any desired position without sequence restrictions. The downside is that the length is limited by the template, and each template can only transcribe one RNA, so multiple RNA copies cannot be generated. PLOR can synthesize long modified RNAs that cannot be produced by solid-phase chemical synthesis or other methods and can generate multiple transcripts through repeated cycles. However, this method has the drawback of being relatively complex and requiring high technical expertise. The PLOR technique requires the use of saturated amounts of labeled rNTPs pools, and each cycle needs thorough washing, leading to high material and equipment costs.

To better understand the biological activity of RNA and its mechanisms of action in living systems, RNA segmental labeling methods have become important

tools in RNA research. Currently, the use of RNA segmental labeling methods for biophysical studies of RNA has become feasible, such as large RNA structural biology studies using FRET, NMR, or EPR spectroscopy. RNA segment-selective labeling also provides more accurate data for RNA visualization and tracking, as it allows scientists to trace the location, modifications, and interactions of RNA with high specificity and sensitivity (Wang *et al.* 2020).

Additionally, in the development of RNA-based therapies, segmental labeling is crucial for elucidating structure-function relationships and for engineering optimized therapeutic agents. For instance, in mRNA therapeutics, segmental labeling and modification allow for the independent optimization of the 5' UTR, coding sequence, and 3' UTR to enhance translation efficiency

and stability while minimizing immunogenicity – a level of control unattainable with random global modification (Brunderová *et al.* 2024). Furthermore, in viral RNA research, segment-selective labeling has been applied to study complex structural elements. By selectively labeling these domains, researchers can use techniques like NMR or smFRET to probe their dynamics and interactions with potential antiviral drugs, guiding rational therapeutic design (Haslecker *et al.* 2023). Although RNA segmental labeling holds great potential in biomedicine, various methods still have limitations, meaning that practical applications still face a series of challenges. From current research progress, segmental labeling of large natural RNA molecules and intracellular RNA remains challenging. The lack of these labeling strategies is detrimental to the development of RNA-based diagnostics and therapeutic applications. The development of RNA segmental labeling methods holds substantial innovation potential. In the future, cross-disciplinary efforts combining chemistry and biotechnology are expected to achieve higher efficiency, lower costs, and better biocompatibility in RNA segmental labeling. With the continuous advancement of scientific research and technology, the future applications of RNA segmental labeling methods will be very broad, contributing greatly to human health and the development of life sciences.

Acknowledgements This work was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB0570300), the National Key Research and Development Project of China (2021YFA1301500, 2021YFC2302401), and the National Natural Science Foundation of China (82341081).

Compliance with Ethical Standards

Conflict of interest Zixuan Hu, Shiping Yang, and Xianyang Fang declare that they have no conflict of interests.

Human and animal rights and informed consent This article does not contain any studies with human or animal subjects performed by the any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agard NJ, Baskin JM, Prescher JA, Lo A, Bertozzi CR (2006) A comparative study of bioorthogonal reactions with azides. *ACS Chem Biol* 1(10): 644–648
- Aranda J, Terrazas M, Gómez H, Villegas N, Orozco M (2019) An artificial DNzyme RNA ligase shows a reaction mechanism resembling that of cellular polymerases. *Nat Catal* 2(6): 544–552
- Barnwal RP, Yang F, Varani G (2017) Applications of NMR to structure determination of RNAs large and small. *Arch Biochem Biophys* 628: 42–56
- Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* 261(5127): 1411–1418
- Bartel DP (2018) Metazoan MicroRNAs. *Cell* 173(1): 20–51
- Becette O, Olingenski LT, Dayie TK (2019) Solid-phase chemical synthesis of stable isotope-labeled RNA to aid structure and dynamics studies by NMR spectroscopy. *Molecules* 24(19): 3476. <https://doi.org/10.3390/molecules24193476>
- Beckert B, Masquida B (2011) Synthesis of RNA by *in vitro* transcription. *Methods Mol Biol* 703: 29–41
- Bellon L, Workman C, Scherrer J, Usman N, Wincott F (1996) Morpholino-linked ribozymes: a convergent synthetic approach. *J Am Chem Soc* 118(15): 3771–3772
- Breaker RR, Joyce GF (1994) A DNA enzyme that cleaves RNA. *Chem Biol* 1(4): 223–229
- Brunderová M, Havlíček V, Matyašovský J, Pohl R, Poštová Slavětinská L, Krömer M, Hocek M (2024) Expedient production of site specifically nucleobase-labelled or hypermodified RNA with engineered thermophilic DNA polymerases. *Nat Commun* 15(1): 3054. <https://doi.org/10.1038/s41467-024-47444-9>
- Chen B, Longhini AP, Nußbaumer F, Kreutz C, Dinman JD, Dayie TK (2018) CCR5 RNA pseudoknots: residue and site -specific labeling correlate internal motions with microRNA binding. *Chem Eur J* 24(21): 5462–5468
- Chen H, Liu D, Aditham A, Guo J, Huang J, Kostas F, Maher K, Friedrich MJ, Xavier RJ, Zhang F, Wang X (2025) Chemical and topological design of multicapped mRNA and capped circular RNA to augment translation. *Nat Biotechnol* 43(7): 1128–1143
- Chow CS, Mahto SK, Lamichhane TN (2008) Combined approaches to site-specific modification of RNA. *ACS Chem Biol* 3(1): 30–37
- Christensen TA, Lee KY, Gottlieb SZP, Carrier MB, Leconte AM (2022) Mutant polymerases capable of 2' fluoro-modified nucleic acid synthesis and amplification with improved accuracy. *RSC Chem Biol* 3(8): 1044–1051
- Conibear AC, Watson EE, Payne RJ, Becker CFW (2018) Native chemical ligation in protein synthesis and semi-synthesis. *Chem Soc Rev* 47(24): 9046–9068
- Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Synthesis of proteins by native chemical ligation. *Science* 266(5186): 776–779
- Depmeier H, Hoffmann E, Bornewasser L, Kath-Schorr S (2021) Strategies for covalent labeling of long RNAs. *ChemBioChem* 22(19): 2826–2847
- Dolinnaya NG, Shabarova ZA (1996) Chemical ligation as a method for the assembly of double-stranded nucleic acids: modifications and local structure studies. *Russ Chem Bull* 45(8): 1787–1809
- Domnick C, Eggert F, Wuebben C, Bornewasser L, Hagelueken G, Schiemann O, Kath-Schorr S (2020) EPR distance measurements on long non-coding RNAs empowered by genetic

- alphabet expansion transcription. *Angew Chem Int Ed* 59(20): 7891–7896
- Duss O, Yulikov M, Jeschke G, Allain FHT (2014) EPR-aided approach for solution structure determination of large RNAs or protein–RNA complexes. *Nat Commun* 5(1): 3669. <https://doi.org/10.1038/ncomms4669>
- El-Sagheer AH, Brown T (2010) New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *Proc Natl Acad Sci USA* 107(35): 15329–15334
- El-Sagheer AH, Brown T (2012) Click nucleic acid ligation: applications in biology and nanotechnology. *Acc Chem Res* 45(8): 1258–1267
- El-Sagheer AH, Brown T (2015) A triazole linkage that mimics the DNA phosphodiester group in living systems. *Quart Rev Biophys* 48(4): 429–436
- Fang X, Stagno JR, Bhandari YR, Zuo X, Wang YX (2015) Small-angle X-ray scattering: a bridge between RNA secondary structures and three-dimensional topological structures. *Curr Opin Struct Biol* 30: 147–160
- Flynn-Charlebois A, Wang Y, Prior TK, Rashid I, Hoadley KA, Coppins RL, Wolf AC, Silverman SK (2003) Deoxyribozymes with 2'–5' RNA ligase activity. *J Am Chem Soc* 125(9): 2444–2454
- Frommer J, Hieronymus R, Selvi Arunachalam T, Heeren S, Jenckel M, Strahl A, Appel B, Müller S (2014) Preparation of modified long-mer RNAs and analysis of FMN binding to the *ypaA* aptamer from *B. subtilis*. *RNA Biol* 11(5): 609–623
- Gait MJ (1984) An introduction to modern methods of DNA synthesis. In: Gait MJ (ed.) *Oligonucleotide synthesis: a practical approach*. Oxford: IRL Press, pp 1–22
- Ganser LR, Kelly ML, Herschlag D, Al-Hashimi HM (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat Rev Mol Cell Biol* 20(8): 474–489
- Glazier DA, Liao J, Roberts BL, Li X, Yang K, Stevens CM, Tang W (2020) Chemical synthesis and biological application of modified oligonucleotides. *Bioconjugate Chem* 31(5): 1213–1233
- Haslecker R, Pham VV, Glänzer D, Kreutz C, Dayie TK, D'Souza VM (2023) Extending the toolbox for RNA biology with SegModTeX: a polymerase-driven method for site-specific and segmental labeling of RNA. *Nat Commun* 14(1): 8422. <https://doi.org/10.1038/s41467-023-44254-3>
- Hennig J, Wang I, Sonntag M, Gabel F, Sattler M (2013) Combining NMR and small angle X-ray and neutron scattering in the structural analysis of a ternary protein–RNA complex. *J Biomol NMR* 56(1): 17–30
- Ho CK, Shuman S (2002) Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc Natl Acad Sci USA* 99(20): 12709–12714
- Ho CK, Wang LK, Lima CD, Shuman S (2004) Structure and mechanism of RNA ligase. *Structure* 12(2): 327–339
- Höbartner C, Rieder R, Kreutz C, Puffer B, Lang K, Polonskaia A, Serganov A, Micura R (2005) Syntheses of RNAs with up to 100 nucleotides containing site-specific 2'-methylseleno labels for use in X-ray crystallography. *J Am Chem Soc* 127(34): 12035–12045
- Höbartner C, Sicoli G, Wachowius F, Gophane DB, Sigurdsson ST (2012) Synthesis and characterization of RNA containing a rigid and nonperturbing cytidine-derived spin label. *J Org Chem* 77(17): 7749–7754
- Höbartner C (2019) How DNA catalyses RNA ligation. *Nat Catal* 2(6): 483–484
- Hu Y, Wang Y, Singh J, Sun R, Xu L, Niu X, Huang K, Bai G, Liu G, Zuo X, Chen C, Qin PZ, Fang X (2022) Phosphorothioate-based site-specific labeling of large RNAs for structural and dynamic Studies. *ACS Chem Biol* 17(9): 2448–2460
- Huang K, Fang X (2023) A review on recent advances in methods for site-directed spin labeling of long RNAs. *Int J Biol Macromol* 239: 124244. <https://doi.org/10.1016/j.ijbiomac.2023.124244>
- Ivanov SA, Vauléon S, Müller S (2005) Efficient RNA ligation by reverse-joined hairpin ribozymes and engineering of twin ribozymes consisting of conventional and reverse-joined hairpin ribozyme units. *FEBS J* 272(17): 4464–4474
- Jaikaran D, Smith MD, Mehdizadeh R, Olive J, Collins RA (2008) An important role of G638 in the *cis*-cleavage reaction of the *Neurospora* VS ribozyme revealed by a novel nucleotide analog incorporation method. *RNA* 14(5): 938–949
- Jaiswal JK, Mattoussi H, Mauro JM, Simon SM (2003) Long-term multiple color imaging of live cells using quantum dot bioconjugates. *Nat Biotechnol* 21(1): 47–51
- Kirk SR, Luedtke NW, Tor Y (2001) 2-Aminopurine as a real-time probe of enzymatic cleavage and inhibition of hammerhead ribozymes. *Bioorg Med Chem* 9(9): 2295–2301
- Kumalasari MR, Alfanaar R, Andreani AS (2024) Gold nanoparticles (AuNPs): a versatile material for biosensor application. *Talanta Open* 9: 100327. <https://doi.org/10.1016/j.talo.2024.100327>
- Kurata S, Ohtsuki T, Suzuki T, Watanabe K (2003) Quick two-step RNA ligation employing periodate oxidation. *Nucleic Acids Res* 31(22): e145. <https://doi.org/10.1093/nar/gng145>
- Lebars I, Vileno B, Bourbigot S, Turek P, Wolff P, Kieffer B (2014) A fully enzymatic method for site-directed spin labeling of long RNA. *Nucleic Acids Res* 42(15): e117. <https://doi.org/10.1093/nar/gku553>
- Lilley DMJ (2005) Structure, folding and mechanisms of ribozymes. *Curr Opin Struct Biol* 15(3): 313–323
- Liu Y, Holmstrom E, Zhang J, Yu P, Wang J, Dyba MA, Chen D, Ying J, Lockett S, Nesbitt DJ, Ferré-D'amaré AR, Sousa R, Stagno JR, Wang YX (2015) Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature* 522(7556): 368–372
- Liu Y, Sousa R, Wang YX (2016) Specific labeling: an effective tool to explore the RNA world. *BioEssays* 38(2): 192–200
- Liu Y, Holmstrom E, Yu P, Tan K, Zuo X, Nesbitt DJ, Sousa R, Stagno JR, Wang YX (2018) Incorporation of isotopic, fluorescent, and heavy-atom-modified nucleotides into RNAs by position-selective labeling of RNA. *Nat Protoc* 13(5): 987–1005
- Lohman GJS, Tabor S, Nichols NM (2011) DNA ligases. *Curr Protoc Mol Biol* 94: 3.14. 1–3.14. 7. <https://doi.org/10.1002/0471142727.mb0314s94>
- Manz C, Kobitski AY, Samanta A, Keller BG, Jäschke A, Nienhaus GU (2017) Single-molecule FRET reveals the energy landscape of the full-length SAM-I riboswitch. *Nat Chem Biol* 13(11): 1172–1178
- McLaughlin LW, Romaniuk E, Romaniuk PJ, Neilson T (1982) The effect of acceptor oligoribonucleotide sequence on the T₄ RNA ligase reaction. *Eur J Biochem* 125(3): 639–643
- Middleton T, Herlihy WC, Schimmel PR, Munro HN (1985) Synthesis and purification of oligoribonucleotides using T₄ RNA ligase and reverse-phase chromatography. *Anal Biochem* 144(1): 110–117
- Müggenburg F, Müller S (2022) Azide-modified nucleosides as versatile tools for bioorthogonal labeling and functionalization. *Chem Rec* 22(5): e202100322. <https://doi.org/10.1002/tcr.202100322>
- Nainar S, Kubota M, McNitt C, Tran C, Popik VV, Spitale RC (2017) Temporal labeling of nascent RNA using photoclick chemistry in live cells. *J Am Chem Soc* 139(24): 8090–8093
- Nandakumar J, Shuman S, Lima CD (2006) RNA ligase structures reveal the basis for RNA specificity and conformational

- changes that drive ligation forward. *Cell* 127(1): 71–84
- Neuner S, Santner T, Kreutz C, Micura R (2015) The “speedy” synthesis of atom -specific ¹⁵N imino/amido -labeled RNA. *Chem Eur J* 21(33): 11634–11643
- Obi P, Chen YG (2021) The design and synthesis of circular RNAs. *Methods* 196: 85–103
- Paredes E, Das SR (2011) Click chemistry for rapid labeling and ligation of RNA. *ChemBioChem* 12(1): 125–131
- Paredes E, Evans M, Das SR (2011) RNA labeling, conjugation and ligation. *Methods* 54(2): 251–259
- Pascal JM (2008) DNA and RNA ligases: structural variations and shared mechanisms. *Curr Opin Struct Biol* 18(1): 96–105
- Pokrovskaya ID, Gurevich VV (1994) *In vitro* transcription: preparative RNA yields in analytical scale reactions. *Anal Biochem* 220(2): 420–423
- Purtha WE, Coppins RL, Smalley MK, Silverman SK (2005) General deoxyribozyme-catalyzed synthesis of native 3'-5' RNA linkages. *J Am Chem Soc* 127(38): 13124–13125
- Sanzone AP, El-Sagheer AH, Brown T, Tavassoli A (2012) Assessing the biocompatibility of click-linked DNA in *Escherichia coli*. *Nucleic Acids Res* 40(20): 10567–10575
- Serganov A, Patel DJ (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet* 8(10): 776–790
- Somoza Á (2008) Protecting groups for RNA synthesis: an increasing need for selective preparative methods. *Chem Soc Rev* 37(12): 2668–2675
- Stagno JR, Yu P, Dyba MA, Wang YX, Liu Y (2019) Heavy-atom labeling of RNA by PLOR for *de novo* crystallographic phasing. *PLoS One* 14(4): e0215555. <https://doi.org/10.1371/journal.pone.0215555>
- Staley JP, Woolford JL Jr (2009) Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Curr Opin Cell Biol* 21(1): 109–118
- Stark MR, Pleiss JA, Deras M, Scaringe SA, Rader SD (2006) An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA* 12(11): 2014–2019
- Strebitzer E, Nußbaumer F, Kremser J, Tollinger M, Kreutz C (2018) Studying sparsely populated conformational states in RNA combining chemical synthesis and solution NMR spectroscopy. *Methods* 148: 39–47
- Vauléon S, Ivanov SA, Gwiazda S, Müller S (2005) Site -specific fluorescent and affinity labelling of RNA by using a small engineered twin ribozyme. *ChemBioChem* 6(12): 2158–2162
- Wang Y, Chen Y, Hu Y, Fang X (2020) Site-specific covalent labeling of large RNAs with nanoparticles empowered by expanded genetic alphabet transcription. *Proc Natl Acad Sci USA* 117(37): 22823–22832
- Wawrzyniak-Turek K, Höbartner C (2014) Deoxyribozyme-mediated ligation for incorporating EPR spin labels and reporter groups into RNA. *Method Enzymol* 549: 85–104
- Weinrich T, Gränz M, Grünewald C, Prisner TF, Göbel MW (2017) Synthesis of a cytidine phosphoramidite with protected nitroxide spin label for EPR experiments with RNA. *Eur J Org Chem* 2017(3): 491–496
- Welz R, Bossmann K, Klug C, Schmidt C, Fritz HJ, Müller S (2003) Site -directed alteration of RNA sequence mediated by an engineered twin ribozyme. *Angew Chem Int Ed* 42(21): 2424–2427
- Wen C, Wang G, Yang L, Chen T, Liu H, Gong W (2024) Structural basis for C2'-methoxy recognition by DNA polymerases and function improvement. *J Mol Biol* 436(20): 168744 <https://doi.org/10.1016/j.jmb.2024.168744>
- Westhof E, Leontis NB (2021) An RNA-centric historical narrative around the Protein Data Bank. *J Biol Chem* 296: 100555. <https://doi.org/10.1016/j.jbc.2021.100555>
- Williamson A, Leiros HKS (2019) Structural intermediates of a DNA–ligase complex illuminate the role of the catalytic metal ion and mechanism of phosphodiester bond formation. *Nucleic Acids Res* 47(14): 7147–7162
- Xu B, Zhu Y, Cao C, Chen H, Jin Q, Li G, Ma J, Yang SL, Zhao J, Zhu J, Ding Y, Fang X, Jin Y, Kwok C, Ren A, Wan Y, Wang Z, Xue Y, Zhang H, Zhang Q, Zhou Y (2022) Recent advances in RNA structure. *Sci China Life Sci* 65(7): 1285–1324