PLOS ONE

RESEARCH ARTICLE

Enzymatic incorporation of an isotope-labeled adenine into RNA for the study of conformational dynamics by NMR

Hannes Feyrer[1], **Cenk Onur Gurdap**[1], **Maja Marušič**[1,2], **Judith Schlagnitweit**[1,3], **Katja Petzold**[1]*

1 Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden, **2** Slovenian NMR Center, National Institute of Chemistry, Ljubljana, Slovenia, **3** Centre de RMN à Très Hauts Champs de Lyon, UMR5082 CNRS/ENS-Lyon/Université Claude Bernard Lyon 1, Villeurbanne, France

* katja.petzold@ki.se

OPEN ACCESS

Citation: Feyrer H, Gurdap CO, Marušič M, Schlagnitweit J, Petzold K (2022) Enzymatic incorporation of an isotope-labeled adenine into RNA for the study of conformational dynamics by NMR. PLoS ONE 17(7): e0264662. https://doi.org/10.1371/journal.pone.0264662

Editor: Patrick van der Wel, Rijksuniversiteit Groningen, NETHERLANDS

Received: February 9, 2022

Accepted: June 8, 2022

Published: July 8, 2022

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: JS: Marie Sklodowska-Curie IF (EU H2020, MSCA-IF Project No. 747446). European Research Executive Agency https://ec.europa.eu/research/mariecurieactions/ KP: Swedish Research Council (grant number 2014-04303, and NT-2018-15), https://www.vr.se/english.html the Swedish Foundation for Strategic Research (project number ICA14-0023 and FFL15-0178), https://strategiska.

Abstract

Solution NMR spectroscopy is a well-established tool with unique advantages for structural studies of RNA molecules. However, for large RNA sequences, the NMR resonances often overlap severely. A reliable way to perform resonance assignment and allow further analysis despite spectral crowding is the use of site-specific isotope labeling in sample preparation. While solid-phase oligonucleotide synthesis has several advantages, RNA length and availability of isotope-labeled building blocks are persistent issues. Purely enzymatic methods represent an alternative and have been presented in the literature. In this study, we report on a method in which we exploit the preference of T7 RNA polymerase for nucleotide monophosphates over triphosphates for the 5' position, which allows 5'-labeling of RNA. Successive ligation to an unlabeled RNA strand generates a site-specifically labeled RNA. We show the successful production of such an RNA sample for NMR studies, report on experimental details and expected yields, and present the surprising finding of a previously hidden set of peaks which reveals conformational exchange in the RNA structure. This study highlights the feasibility of site-specific isotope-labeling of RNA with enzymatic methods.

Introduction

Nuclear magnetic resonance spectroscopy (NMR) is a versatile tool to study the structure and dynamics of RNA at high resolution [1–4]. Unfortunately, the chemical shift dispersion, a valuable descriptor of structure, is rather low due to high homogeneity of the helical secondary structure, and the presence of only four different building blocks. Therefore, resonances of larger RNAs often overlap severely and make resonance assignment and downstream analysis, e.g. conformational dynamics, more challenging or even impossible. Besides solutions provided by NMR itself (e.g. multidimensional experiments or selective magnetization transfer [5–7]), the issue can be tackled on the sample preparation side. Site-specific isotope labeling is one way to overcome the signal overlap that has been addressed extensively in the literature [8, 9].

se/en/ Harald och Greta Jeansson Stiftelse (JS20140009), https://jeanssonsstiftelser.se/om-jeanssons-stiftelser/ Eva och Oscar Ahréns Stiftelse, https://www.ahrenstiftelse.se Åke Wiberg Stiftelse (467080968 and M14-0109), https://ake-wiberg.se Cancerfonden (CAN 2015/388, CAN 2018/715), https://www.cancerfonden.se the Karolinska Institute (KID 2016-00062), https://staff.ki.se/kid-funding the Ragnar Söderberg Stiftelse (M91/14), https://ragnar.soderbergs.org Knut och Alice Wallenberg foundation [project grant KAW 2016.0087] https://kaw.wallenberg.org The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

The most common method for site-specific incorporation of ^{13}C or ^{15}N-labeled nucleotides is solid-phase oligonucleotide synthesis (SPOS). Since these building blocks are often synthesized in-house, they can be modified, atom-specifically labeled or both [9, 10]. While SPOS has unique strengths, the most important one being simple incorporation of chemical modifications or isotope-labeled residues, the yield drops rapidly for larger constructs (>50 nucleotides (nt)). Furthermore, modified or isotope-labeled building blocks are often not commercially available or come at a significant cost.

Methods for site-specific incorporation of isotope-labeled nucleotides with *in vitro* transcription (IVT) have been presented, and can be divided in chemo-enzymatic approaches [11–13] and pause-restart methods [14]. The former relies on chemical synthesis of the isotope-labeled nucleotide analog, which is then ligated to the 3' terminus of an unlabeled strand. Another unlabeled strand is ligated to the now 3'-labeled RNA molecule [12]. Liu and coworkers developed the position-selective labeling of RNA (PLOR) method, which is a sophisticated adaption of pause-restart peptide synthesis methods to RNA synthesis on agarose beads [14]. Both these methods were shown to be versatile and robust, at the only caveat of needing either chemical synthesis of labeled nucleotide analog or a robotic platform for automated PLOR synthesis and modified DNA templates.

Instead of single nucleotides, shorter fragments can be labeled site-specifically and then be ligated to a larger unlabeled RNA fragment(s), which still reduces signal overlap significantly. Duss et al. presented a cut-and-paste protocol, based on *in-vitro*-transcribed labeled and unlabeled strands. These strands are then cleaved with RNase H and internal ribozymes, so that labeled fragments can be ligated to unlabeled ones, which yields a long RNA, with specific isotope-labeled segments [15, 16].

Purely enzymatic methods to incorporate site-specific NMR labels and modifications have been exploiting the fact that T7 RNA polymerase (T7RNAP) favors nucleotide monophosphates (NMP) over nucleotide triphosphates (NTP) in the first position, even though the enzyme will use NTPs if no NMPs are present [17, 18]. The lack of the pyrophosphate as a leaving group makes it impossible for the NMP to be incorporated into any other position than the 5' end. This approach was used by Brown and coworkers to incorporate protonated GMP in the first position of a 232 kDa RNA, while all other nucleotides remained deuterated, and thus invisible in NMR spectroscopy [19]. Lebars and colleagues incorporated a 6-thioguanosine into the 5' of an RNA, which was later ligated with an unmodified RNA strand to yield a site-specifically modified RNA. This thioguanosine was later click-labeled with a proxyl-derivate radical, allowing for paramagnetic NMR studies [20].

This article describes a similar workflow, which incorporates a site-specific isotope-labeled nucleotide into an RNA molecule using only established methods based on T7 IVT. Our approach is similar to the method published by Lebars and coworkers and therefore strengthens the general validity of such site-specific RNA labeling methods. Importantly, compared to most other studies, we use adenosine-monophosphate (AMP) instead of guanosine-monophosphate (GMP) [19, 20]. Initiating transcription with an adenosine is known to hamper transcription initiation and result in lower yield [21, 22] and thus poses a greater challenge. We thoroughly report on the sample preparation details, quantitative yields and highlight limitations and bottlenecks of the methods. Additionally, this work provides the feasibility of site-directed RNA isotope labeling to the wider NMR and RNA communities and shows how useful it can be to circumvent experimental issues with resonance overlap.

In brief, we utilize T7 RNA polymerase to incorporate a $^{13}C/^{15}N$-labeled AMP in the 5' position of a 16 nt RNA fragment, to then site-specifically label a longer RNA after ligation of a second RNA strand with T4 DNA ligase. We present the successful incorporation of a single $^{13}C/^{15}N$-labeled adenosine into a 46 nt RNA fragment stemming from helix-44 in the human

ribosome [23], which was not possible to be resonance-assigned using uniform labeling in our hands. The isotope label aids in the unambiguous assignment of the A31 residue, and further analysis of the labeled nucleotide with solution-state NMR spectroscopy gave unexpected insights into the conformational dynamics of the 46-mer.

Material and methods

All RNA molecules were prepared using T7 IVT from a single-stranded template (double-stranded in the promoter region), which were 2'-OMe modified at the last two nucleotides of the template strand [24, 25]. Transcription products of similar length were purified using ion-exchange (IE) HPLC before further processing.

RNA sequences (Fig 1)

Full-length RNA (A31 in bold):

5 ′- GGU UUA GUG AGG CCC UCG GAU UUC GAG AAG **A**CG GUC GAA CUU GAC C -3′

5'-RNA fragment:

5′- GGU UUA GUG AGG CCC UCG GAU UUC GAG AAG -3′

3'-RNA fragment:

3′- ACG GUC GAA CUU GAC C -3′

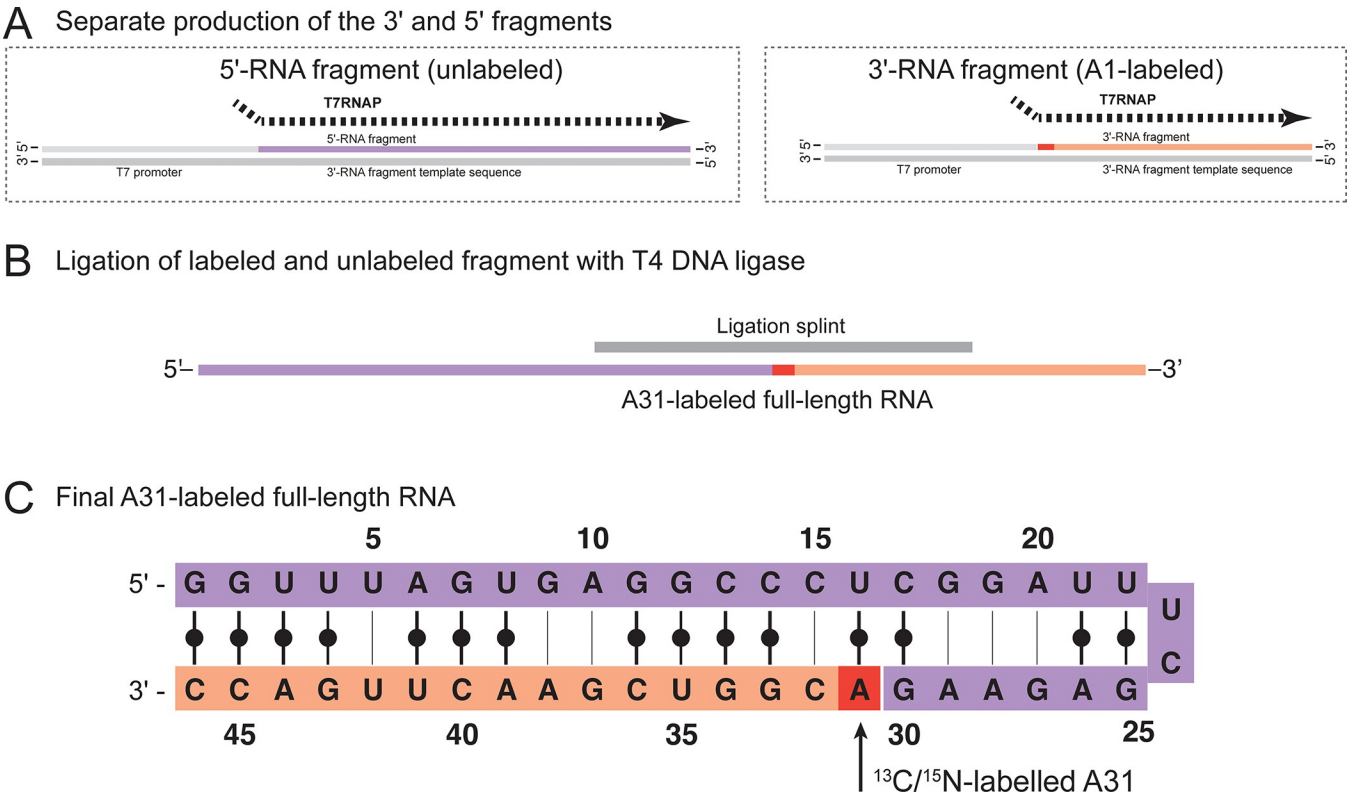

Fig 1. Scheme illustrating the workflow of A31-labeling of target RNA (46 nt). A: Separate production of two RNA fragments. The 3'-RNA fragment (orange) starts with the labeled nucleotide A31, which is supplied as 13C/15N-AMP into the in vitro transcription reaction. B: Ligation of unlabeled 5'-RNA (purple) and A1-labeled 3'-RNA fragment (orange) using T4 DNA ligase to produce the A31-labeled full-length RNA construct. C: Final A31-labeled RNA construct comprised of unlabeled 5'-RNA (purple) and A1-labeled 3'-RNA (orange).

https://doi.org/10.1371/journal.pone.0264662.g001

Template preparation

DNA templates for 5'-RNA fragment and 3'-RNA fragment (see Fig 1) were ordered from IDT as custom DNA oligos with the 'Standard Desalting' purification option. Received oligos were diluted to 100 μM in MilliQ H_2O and mixed by pipetting.

Template sequences with promoter complement indicated in italics are shown below:

Full-length DNA template:

5′-mGmGT CAA GTT CGA CCG TCT TCT CGA AAT CCG AGG GCC TCA CTA AAC C*TA TAG TGA GTC GTA TTA* -3′

5'-DNA fragment template:

5′- mCmUT CTC GAA ATC CGA GGG CCT CAC TAA ACC *TAT AGT GAG TCG TAT TA* -3'

3'-DNA fragment template:

5′- mGmGT CAA GTT CGA CCG T*TA TAG TGA GTC GTA TTA* -3′

T7 promoter: 5′- TAA TAC GAC TCA CTA TA -3′

T7 promoter and respective template oligo were annealed by mixing both DNA strands to a final concentration of 25 μM each and heating to 95˚C for 5 minutes and successive cooling at room temperature for 30 minutes.

T7 In vitro transcription

The IVT conditions for all 3 constructs were optimized by rational variation in small scale before production and purification in large scale. Reactions were prepared at 50 μL in parallel and incubated for 1.5 hours at 37˚C. 1 μL of each IVT reaction was quenched in 9 μL gel loading dye (5 mM EDTA, 300 μM bromophenol blue in formamide) and heated to 95˚C for 2 minutes. 1 μL of the solution was loaded onto a 20% denaturing PAGE. Concentration ranges screened were: $MgCl_2$: 10–50 mM, Tris-Cl: 60–200 mM, DTT: 10–60 mM, Spermidine: 2–20 mM, NMP: 0–20 mM, NTPs: 3–8 mM and DMSO: 0–20%.

Conditions giving the strongest target band were then scaled up to 10 mL for purification and NMR sample preparation (see Table 1). Reagents were added in order as shown in the Table 1. All reagents were obtained from Sigma-Aldrich at molecular biology grade. T7RNAP has been expressed in *E.coli* and purified by His-tag affinity purification and size-exclusion chromatography by the protein science facility (PSF) at the Karolinska Institute.

Denaturing polyacrylamide gel electrophoresis

All analytical gels were performed in 20% polyacrylamide gels (acrylamide:bisacrylamide 19:1) using 8x12 cm plates in a mini-PROTEAN system (BioRad). Acrylamide solution was prepared in 1X TBE buffer (89 mM Tris, 89 mM Boric acid, 2 mM EDTA, pH 8.3) and 8 M urea.

Table 1. Optimized IVT conditions for RNA samples.

Reagent	5'-RNA	3'-RNA	Full-length RNA
Tris-Cl pH 8	80 mM	100 mM	150 mM
$MgCl_2$	20 mM	30 mM	16 mM
DTT	60 mM	60 mM	10 mM
Spermidine	5 mM	2 mM	20 mM
AMP	-	10 mM	-
NTPs	3 mM	3 mM	4 mM
DNA	0.4 μM	0.4 μM	0.4 μM
T7RNAP	200 μg/mL	300 μg/mL	300 μg/mL
DMSO	20% (v/v)	5% (v/v)	5% (v/v)

https://doi.org/10.1371/journal.pone.0264662.t001

5 mL acrylamide solution were polymerized in a 15 mL centrifuge tube by adding 4 μL N,N,N,N-tetramethylethylenediamine (TEMED) and 40 μL 10% ammonium persulfate (APS) solution (v/w). The solution was mixed by inverting before pouring between the prepared glass plates (0.75 mm spacer). Gels were run in 1X TBE buffer at 350 V for typically 60–75 minutes. The wells were cleaned from urea by flushing with running buffer using a syringe before loading the sample.

1 μL sample was diluted in 9 μL loading solution (5 mM EDTA, 300 μM bromophenol blue in formamide) and heated to 95˚C for 2 minutes. Typically, 1 μL of the diluted sample was loaded onto the gel. Gels were stained with SYBRGold (Thermo) according to the manufacturer's instructions and imaged in an Amersham ImageQuant 800 UV at 498 nm excitation wavelength.

HPLC purification

HPLC purification was performed following previously published methods [26, 27]. Large scale reactions (transcription or ligation) were quenched by adding EDTA to 50 mM final concentration and filtered using a 0.2 μm cellulose acetate syringe filter.

The Dionex Ultimate 3000 UHPLC system (Thermo) was equipped as follows: DNAPac PA200 22x250 Preparative column, LPG-3400RS Pump, TCC-3000RS Column thermostat, AFC-3000 Fraction collector, VWD-3100 Detector. Buffer A: 20 mM sodium acetate, 20 mM sodium perchlorate, pH 6.5. Buffer B: 20 mM sodium acetate, 600 mM sodium perchlorate, pH 6.5. Buffers were degassed and filtered before use. Column temperature: 75˚C. Flow rate: 8 mL/min. Injection loop: 5 mL. Pump sequences: 0–10 min: 0% buffer B, 10–40 min: elution gradient, 40–50 min: 100% buffer B, 50–60 min: 0% buffer B. Elution gradients: 5'-RNA (30 nt): 22–30% buffer B, 3'-RNA (16 nt): 15–25% buffer B, Full-length RNA (46 nt): 15–45% buffer B.

Collected fractions were tested for RNA of interest by loading 0.1 μL on a 20% denaturing PAGE.

RNA ligation

Ligation splints were obtained from IDT as standard DNA oligos, with the 'Standard Desalting' purification option. Received oligos were diluted to 100 μM in MilliQ H_2O and mixed by pipetting.

Ligation splint 21 nt: `5′- AGT TCG ACC GTC TTC TCG AAA -3′`

Ligation splint 26 nt: `5′- CAA GTT CGA CCG TCT TCT CGA AAT CC -3′`

RNA ligations with T4 DNA ligase were optimized in 20 μL scale before scale-up of labeled samples by concentration variation of PEG 4000 (0–20%), DMSO (0–40%), reaction time (1, 3, 5, 24 and 48 h) and temperature (4, 22 and 37˚C). Final conditions were: 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, 15 μM 5'-RNA fragment, 15 μM 3'-RNA fragment, 22.5 μM ligation splint, 2.5% PEG 4000 (w/v), 10% DMSO (v/v), 10 u/μL T4 DNA ligase (stock 400 u/μL, NEB) for 48 hours at room temperature. Especially the combination of both PEG and DMSO with the long reaction time of 48 h seemed to make a large difference in the success of the reaction, which is where our conditions deviate from commercial suppliers. No difference in yield was found between the ligation splints. Ligation reactions have been purified by IE HPLC as described above, using a gradient of 15–45% buffer B.

DNase I digest

The concentrated HPLC fractions were DNase I-digested in 6 mL at 1 u/μL in 1X DNase buffer (Thermo) for 15 minutes at 37˚C and purified by ion-exchange HPLC as described above.

NMR spectroscopy

For folding, NMR samples have been diluted to 10–50 μM in NMR buffer (15 mM sodium phosphate, 25 mM NaCl, 0.1 mM EDTA, pH 6.5) by heating to 95˚C for 5 minutes and snap-cooled in a water/ice/salt mixture for 30 minutes. The sample was then concentrated to 225 μL in a centrifugal filter unit (Amicon Ultra-2, 3 kDa cutoff). 10% D_2O (v/v) was added to reach sample volume of 250 μL and transferred to a Shigemi tube.

NMR experiments were acquired on a 600 MHz Bruker Avance III spectrometer equipped with 5 mm HCNP QXI Cryo-probe at 298 K. Spectra were processed using the Bruker Topspin 3.6 and 4.0.6 software.

$^{1}H,^{13}C$-HSQC spectra for the A1-labeled 3'-RNA fragment (300 μM) were recorded with 128 indirect points and 8 scans. $^{1}H,^{13}C$-HSQC spectra for the A31-labeled full-length RNA (230 μM) were recorded with 256 points in the indirect dimension and 200 scans, summing up to 24 hours measurement time, and with 350 indirect points and 16 scans for the A/U-labeled sample (1.3 mM). 2D F_1-^{13}C-edited [28] nuclear Overhauser effect spectroscopy (NOESY)/exchange spectroscopy (EXSY) and rotating-frame nuclear Overhauser spectroscopy (ROESY) [29] experiments of the A31-labeled full-length RNA were recorded with 114 points in the indirect dimension and 384 scans (NOESY) or 567 scans (ROESY). NOESY mixing time was 175 ms and ROESY spinlock was 8 ms at 10 kHz.

Results

Construct design

We chose the 46 nt construct (Fig 1) because several resonances could not be assigned in a uniformly A/U-labeled sample due to spectral overlap and missing sequential connections, one of them A31. Two new constructs were designed, called 5'-RNA fragment (30 nt) and 3'-RNA fragment (16 nt), which will carry the 5'-adenosine label from AMP-labeled IVT and be ligated to the 5'-RNA fragment, as shown in Fig 1A–1C. Fig 1C shows a secondary structure simulation (MC-Fold), and has not yet been confirmed with experimental data.

The incorporation of a single label at the 5' position of the 16 nt 3'-RNA fragment was successful according to NMR experiments. $^{13}C/^{15}N$ labeled AMP (at 10 mM) was used at 3.3x excess over the other NTPs (3 mM each), as it was optimized in small scale reactions. A denaturing PAGE shows a product of 16 nt after IE HPLC purification (Fig 2A, lane 2). An aromatic $^{1}H,^{13}C$-HSQC of the A1-labeled 3'-RNA fragment showed two signals (C2 and C8), as expected for a signal originating from a single $^{13}C/^{15}N$-labeled adenosine nucleotide (Fig 2B).

RNA production

For the NMR spectra shown in Figs 2 and 3, an IVT reaction of the A1-labeled 3'-RNA fragment at 10 mL scale was performed. After HPLC purification, 130 nmol were obtained. This corresponds to a yield of 4.8%, whereby the limiting reagent is the most abundant nucleotide in the RNA sequence, as described previously [30].

The unlabeled 5'-fragment was produced at larger scale (at a yield of 5.8%) and added to stoichiometric amounts of the A-labeled 3'-fragment in the ligation reaction. The ligation reaction was prepared at a scale of 8.5 mL, which contains all of the produced labeled 3'-RNA fragment at a final concentration of 15 μM. During optimization of the ligation reaction, we found a combination DMSO and PEG-4000 to give the highest ligation yields over the reaction time of 48 hours.

The concentration of the final NMR sample was 230 μM in 250 μL, which equals 58 nmol labeled full-length RNA and implies a ligation yield, including IE HPLC purification, of 45%.

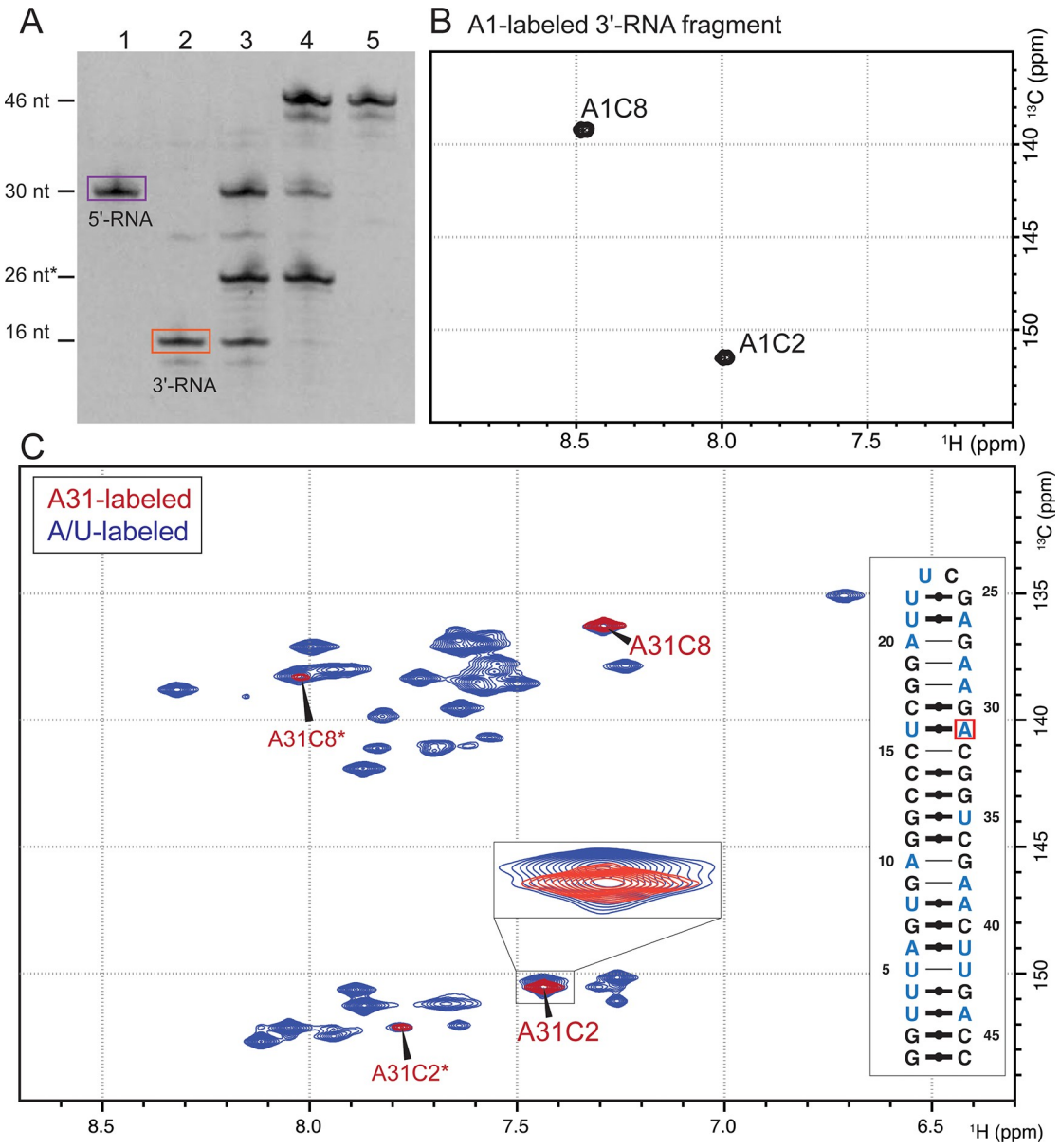

Fig 2. Confirmation of A31-label incorporation into the 46 nt full-length RNA. A: Denaturing PAGE of ligation reaction. Size reference nt* refers to a DNA oligo (ligation splint). Lane 1: Purified 5'-RNA fragment (purple box); Lane 2: Purified 3'-RNA fragment (orange box); Lane 3: Negative control of ligation reaction without T4 DNA ligase. Lane 4: Ligation reaction after 48 h. Lane 5: HPLC-purified full-length RNA. B: ^{1}H,^{13}C-HSQC of A1-labeled 3'-RNA fragment (16 nt) showing the two expected signals for C8 and C2 of A1. C: ^{1}H,^{13}C-HSQC spectra of the aromatic region of A31-labeled full-length RNA (red) and uniformly A/U-labeled full-length RNA (blue). For the A31-labeled RNA, two dominant set of C8/C2 signals appear alongside a weaker set of signals (A31C2* and A31C8*). The zoom shows that A31C2 does indeed overlay directly with a signal of the A/U-labeled RNA, which is partially overlapped with another signal. Furthermore, clear differences between 2B and 2C indicate proper incorporation of the isotope-labeled nucleotide into the RNA. A secondary structure prediction [31] shows which nucleotides are expected to give HSQC signals for A31-labeled RNA (red) or A/U-labeled RNA (blue).

https://doi.org/10.1371/journal.pone.0264662.g002

In comparison, Lebars et al. and Duss et al. reported on 50 and 100% ligation yield respectively, however, using T4 RNA ligases instead of T4 DNA ligase [15, 20]. At this point, it should be noted that co-elution of the full-length RNA with the DNA splint required DNase I treatment and re-purification of the final sample, which could have impacted the yield. PAGE

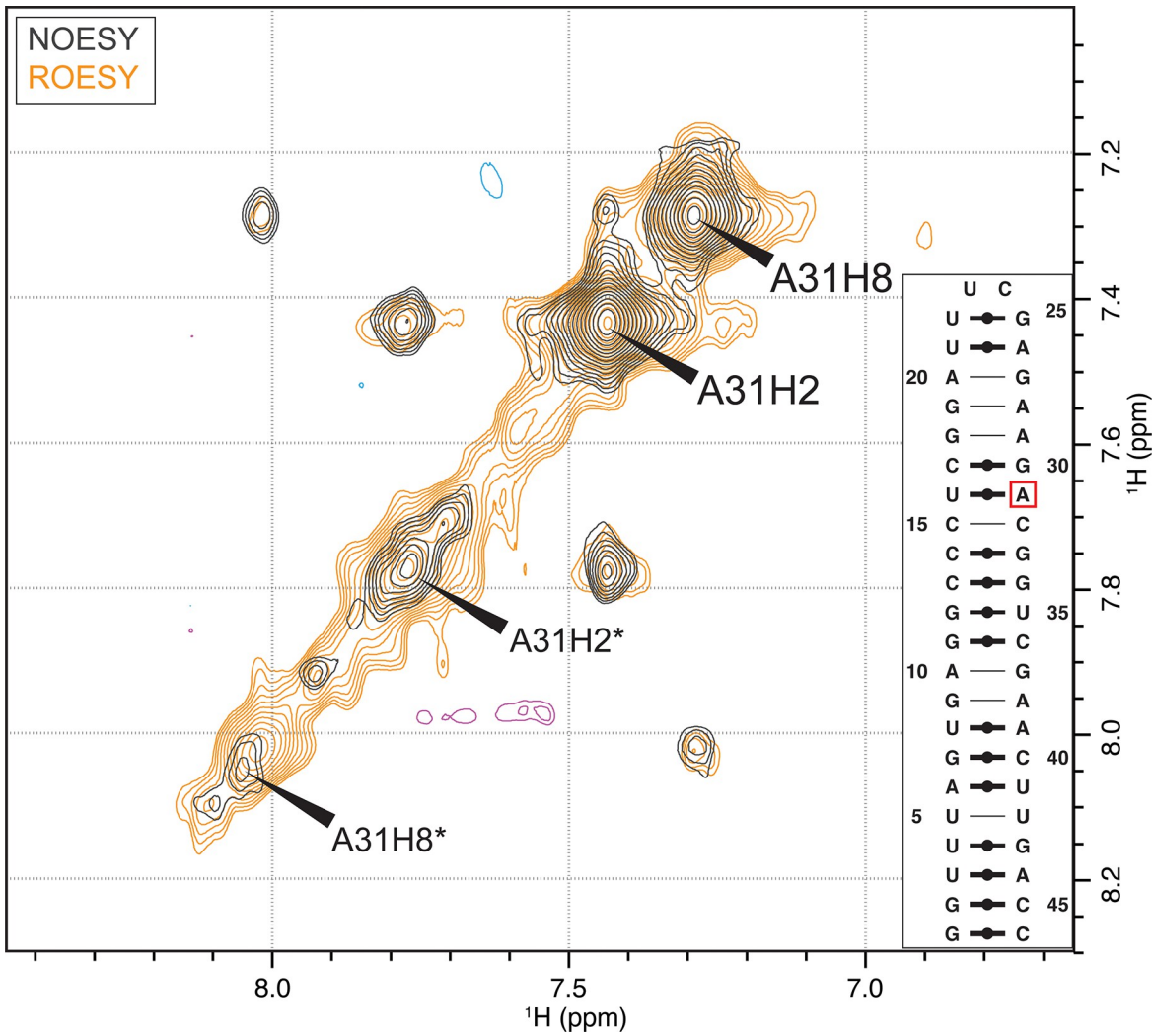

Fig 3. Investigation of the second set of signals in the A31-labeled RNA. ^{13}C-filtered NOESY (gray/blue) and ^{13}C-filtered ROESY (beige/purple) for A31-labeled RNA to probe slow conformational exchange. All cross peaks have the same sign as the diagonal signals in the respective experiment, confirming conformational exchange. A secondary structure prediction [31] is shown in the bottom right box.

https://doi.org/10.1371/journal.pone.0264662.g003

analysis of the crude ligation reaction suggests that the entire 3'-RNA was consumed (see Fig 2A, lane 4), which indicates that most of the product was lost during HPLC purification and buffer-exchange.

Confirmation of site-specific RNA labeling

The ligation of 3'-RNA and 5'-RNA fragments was confirmed by denaturing PAGE and NMR spectroscopy. Denaturing PAGE shows a ligated band, indicating the successful formation of the 46 nt construct (Fig 2A, lane 4). A second, lower band at approx. 44 nt is visible, indicating a shorter product, probably representing an impurity in the preparation of the 3'-RNA fragment (Fig 2A, lane 2).

The aromatic ^{1}H,^{13}C-HSQC spectra of the full-length RNA sample after HPLC purification and snap-cooling, shows four peaks which overlap exactly with the A/U-labeled sample for the fully labeled construct (Fig 2C). Only two signals were expected for a single adenosine in the

RNA strand and Fig 2B shows that only one adenosine is incorporated in the 3'-RNA fragment. However, two sets of signals each for A31C2 and A31C8 were observed. We hypothesized slow conformational exchange to be the reason for the second set of peaks., as the 5' fragment band after HPLC purification is highly pure, as visible in Fig 2A, lane 1.

To probe for slow conformational exchange, we performed 2D F_1-^{13}C-edited [28] NOESY/EXSY and ROESY experiments, where the signs of the cross and diagonal peaks allow the discrimination between true NOE/ROE signals and cross peaks caused by the exchange process [32]. For an RNA molecule of 46 nt, NOE cross and diagonal peaks are positive in both cases. In contrast, the ROE cross and diagonal peaks are of opposite sign in the case of 'true' ROE peaks, but of the same sign in the case of conformational exchange. In both spectra, all cross and diagonal signals are of positive sign, indeed indicating a slow exchange process occurring (Fig 3). Conclusively, the second set of signals does not invalidate the sample preparation procedure but is a feature of the conformational dynamics of the full-length RNA, an information that would have likely not been possible to identify without the single label. At this point, the alternative conformation has not been probed further.

Discussion

By now, methods for chemo-enzymatic [12] or purely enzymatic incorporation of site-specific spin labeling [20], including the use of NMPs for 5'-incorporation of labeled nucleotides [19] have been published. Compared to these studies, our method uses minor variations, for example, in the choice of ligase (T4 DNA ligase), or RNA purification method (ion-exchange HPLC). Furthermore, we incorporated a ^{13}C/^{15}N-labeled nucleotide at the specific site, although other studies used either unlabeled or modified nucleotides [12, 19, 20, 33]. Our approach is similar enough to the published literature to expect a successful incorporation of the labeled nucleotide in the right position, as we have shown here. With our work, we showcase the robustness and replicability of the method, and hopefully encourage more researchers to apply site-specific labeling where other NMR approaches prove difficult.

The optimization of several reaction steps was crucial for the success of the protocol. First, transcription conditions for all RNA constructs needed to be optimized to yield enough RNA for further reactions and NMR analysis. This was achieved by small-scale IVT reactions (50 μL) and variation of $MgCl_2$, Tris-Cl, NMP, NTP, and DMSO concentrations. Ligation reactions required optimization of PEG, DMSO, temperature, and incubation time. We found no change in efficacy between a 21 nt and a 26 nt DNA splint. Due to several HPLC purification steps the final yield is rather low, yet sufficient for the assignment of the unknown nucleotide. Further scale-up of the reaction seems feasible, however, one can expect diminishing returns due to the limited separation capacity of the HPLC columns with larger injection amounts. Labeled nucleotides can, in principle, be recovered and purified from the IVT buffer [34, 35]. Ultimately, the maximal achievable yield will depend on the yield of the IVT and ligation steps, and loss during purification. We decided to thoroughly purify both 3'- RNA and 5'-RNA before ligation to remove transcripts of similar length, as those could still be ligated over a gap [36]. Increased yield could be achieved by shortening purification efforts, e.g., through phenol-choloform extraction and ethanol precipitation, but could ultimately lead to ligation side products stemming from heterogeneous transcription.

The method is expected to work at a higher yield for integration of site-specific guanosine residues, as all native T7RNAP promoters contain at least two guanosine nucleotides at the 5' end of the transcript. Probably for this reason most previous publications used GMP at the site of labeling or modification [19, 20]. Despite a lower yield for a 5'-terminal adenosine, we successfully produced an NMR sample that could answer questions about the assignment and

conformational dynamics. This showed that the method is feasible for starting nucleotides other than GMP, however, the yield is expected to drop further when pyrimidine bases are used [21, 22, 37]. Despite reduced yield, non-G nucleotides should not be completely be disregarded for starting nucleotides as we and others have shown the transcription initiation with adenosine [38] and uridine [3].

Next to isotope-labeled nucleotides, certain modifications or even oligonucleotides could be incorporated by the same protocol, as their 5'-incorporation has been shown before [18, 38, 39]. Such modifications, next to isotopically nucleotides for NMR, open the possibility of site-specific analysis of RNA with methods like fluorescence spectroscopy [40] or EPR spectroscopy [41]. The recently published SMRITI-method for introduction of several modifications during IVT [40] could also be modified to introduce a labeled segment via an engineered elongation complex [42].

Another limitation of the method is the position of the label. While T7 IVT has no upper length limit (within the size limit of solution NMR), short sequences (below 15 nt) are difficult to synthesize. Labels closer to the 3' or 5' termini of the full-length RNA could potentially be prepared in conjuction with a cis-cleaving ribozyme or with solid-phase oligonucleotide synthesis instead.

Alternative routes for ligation include T4 RNA ligase 1 and 2, which are known to be highly efficient, and differ with regard to their sequence and structure specificity [43–45]. Also, a number of ribozymes and DNAzymes have opened new possibilities to ligate RNA in different contexts [46, 47].

Conclusion

In this report, we show another application of purely enzymatic site-specific labeling of RNA by IVT with isotope-labeled nucleotide monophosphate and successive ligation to an unlabeled fragment (Fig 1). Finally, we could show that the single adenine isotope label reveals a second conformation, which is in slow exchange with the ground state conformation. The strength of the method lies in the absence of an upper size-limitation over solid-phase oligonucleotide synthesis and in a small number of robust steps needed (*in vitro* transcription and successive RNA ligation) over chemo-enzymatic methods. The weak point of the method is the sequence-dependency of incorporation efficiency of the labeled nucleotide monophosphate at the 5'-end of the labeled RNA strand. By thoroughly reporting on yields and limitations, we hope to encourage researchers to use this approach to overcome limitations from spectral crowding in NMR studies of large RNA molecules.

Supporting information

S1 Raw image.
(PDF)

Acknowledgments

We thank the Petzoldlab and especially Lorenzo Baronti for inspiring discussions and the Martin Hällberg group for the generous gift of the inorganic phosphatase. We are indebted to the protein science facility (PSF) at the Karolinska Institute for reliable production of the T7 RNA polymerase.

Author Contributions

Conceptualization: Hannes Feyrer, Maja Marušič, Katja Petzold.

Data curation: Hannes Feyrer.

Formal analysis: Hannes Feyrer, Cenk Onur Gurdap.

Funding acquisition: Katja Petzold.

Investigation: Hannes Feyrer, Cenk Onur Gurdap, Maja Marušič, Judith Schlagnitweit.

Methodology: Hannes Feyrer, Cenk Onur Gurdap, Maja Marušič, Judith Schlagnitweit.

Project administration: Katja Petzold.

Supervision: Katja Petzold.

Validation: Judith Schlagnitweit.

Visualization: Hannes Feyrer.

Writing – original draft: Hannes Feyrer, Cenk Onur Gurdap.

Writing – review & editing: Hannes Feyrer, Cenk Onur Gurdap, Maja Marušič, Judith Schlagnitweit, Katja Petzold.

References

1. Liu B., Shi H., and Al-Hashimi H.M. (2021). Developments in solution-state NMR yield broader and deeper views of the dynamic ensembles of nucleic acids. Curr. Opin. Struct. Biol. 70, 16–25. https://doi.org/10.1016/j.sbi.2021.02.007 PMID: 33836446
2. Dethoff E.A., Petzold K., Chugh J., Casiano-Negroni A., and Al-Hashimi H.M. (2012). Visualizing transient low-populated structures of RNA. Nature 491, 724–728. https://doi.org/10.1038/nature11498 PMID: 23041928
3. Baronti L., Guzzetti I., Ebrahimi P., Friebe Sandoz S., Steiner E., Schlagnitweit J., et al. (2020). Base-pair conformational switch modulates miR-34a targeting of Sirt1 mRNA. Nature 583, 139–144. https://doi.org/10.1038/s41586-020-2336-3 PMID: 32461691
4. Shi H., Rangadurai A., Abou Assi H., Roy R., Case D.A., Herschlag D., et al. (2020). Rapid and accurate determination of atomistic RNA dynamic ensemble models using NMR and structure prediction. Nat. Commun. 11, 5531. https://doi.org/10.1038/s41467-020-19371-y PMID: 33139729
5. Farès C., Amata I., and Carlomagno T. (2007). 13C-detection in RNA bases: revealing structure-chemical shift relationships. J. Am. Chem. Soc. 129, 15814–15823. https://doi.org/10.1021/ja0727417 PMID: 18052161
6. Nikonowicz E.P., and Pardi A. (1992). Three-dimensional heteronuclear NMR studies of RNA. Nature 355, 184–186. https://doi.org/10.1038/355184a0 PMID: 1370345
7. Latham M.P., Brown D.J., McCallum S.A., and Pardi A. (2005). NMR methods for studying the structure and dynamics of RNA. Chembiochem 6, 1492–1505. https://doi.org/10.1002/cbic.200500123 PMID: 16138301
8. Skrisovska L., and Allain F.H.-T. (2008). Improved segmental isotope labeling methods for the NMR study of multidomain or large proteins: application to the RRMs of Npl3p and hnRNP L. J. Mol. Biol. 375, 151–164. https://doi.org/10.1016/j.jmb.2007.09.030 PMID: 17936301
9. Olenginski L.T., Taiwo K.M., LeBlanc R.M., and Dayie T.K. (2021). Isotope-Labeled RNA Building Blocks for NMR Structure and Dynamics Studies. Molecules 26. https://doi.org/10.3390/molecules26185581 PMID: 34577051
10. Neuner S., Santner T., Kreutz C., and Micura R. (2015). The "Speedy" Synthesis of Atom-Specific (15) N Imino/Amido-Labeled RNA. Chem. Eur. J 21, 11634–11643. https://doi.org/10.1002/chem.201501275 PMID: 26237536
11. Alvarado L.J., Longhini A.P., LeBlanc R.M., Chen B., Kreutz C., and Dayie T.K. (2014). Chemo-enzymatic synthesis of selectively $^{13}C/^{15}N$-labeled RNA for NMR structural and dynamics studies. Meth. Enzymol. 549, 133–162.
12. Keyhani S., Goldau T., Blümler A., Heckel A., and Schwalbe H. (2018). Chemo-Enzymatic Synthesis of Position-Specifically Modified RNA for Biophysical Studies including Light Control and NMR Spectroscopy. Angew. Chem. Int. Ed 57, 12017–12021. https://doi.org/10.1002/anie.201807125 PMID: 30007102

13. Blümler A., Schwalbe H., and Heckel A. (2022). Solid-Phase-Supported Chemoenzymatic Synthesis of a Light-Activatable tRNA Derivative. Angew. Chem. Int. Ed 61, e202111613. https://doi.org/10.1002/anie.202111613 PMID: 34738704

14. Zhang X., Li M., and Liu Y. (2020). Optimization and characterization of position-selective labelling of RNA (PLOR) for diverse RNA and DNA sequences. RNA Biol. 17, 1009–1017. https://doi.org/10.1080/15476286.2020.1749797 PMID: 32249673

15. Duss O., Maris C., von Schroetter C., and Allain F.H.-T. (2010). A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. Nucleic Acids Res. 38, e188. https://doi.org/10.1093/nar/gkq756 PMID: 20798173

16. Duss O., Diarra Dit Konté N., and Allain F.H.-T. (2015). Cut and paste RNA for nuclear magnetic resonance, paramagnetic resonance enhancement, and electron paramagnetic resonance structural studies. Meth. Enzymol. 565, 537–562.

17. Huang F., He J., Zhang Y., and Guo Y. (2008). Synthesis of biotin-AMP conjugate for 5' biotin labeling of RNA through one-step in vitro transcription. Nat. Protoc. 3, 1848–1861. https://doi.org/10.1038/nprot.2008.185 PMID: 18989262

18. Williamson D., Cann M.J., and Hodgson D.R.W. (2007). Synthesis of 5'-amino-5'-deoxyguanosine-5'-N-phosphoramidate and its enzymatic incorporation at the 5'-termini of RNA molecules. Chem. Commun., 5096–5098. https://doi.org/10.1039/b712066d PMID: 18049765

19. Brown J.D., Kharytonchyk S., Chaudry I., Iyer A.S., Carter H., Becker G., et al. (2020). Structural basis for transcriptional start site control of HIV-1 RNA fate. Science 368, 413–417. https://doi.org/10.1126/science.aaz7959 PMID: 32327595

20. Lebars I., Vileno B., Bourbigot S., Turek P., Wolff P., and Kieffer B. (2014). A fully enzymatic method for site-directed spin labeling of long RNA. Nucleic Acids Res. 42, e117. https://doi.org/10.1093/nar/gku553 PMID: 24981512

21. Kuzmine I., Gottlieb P.A., and Martin C.T. (2003). Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. J. Biol. Chem. 278, 2819–2823. https://doi.org/10.1074/jbc.M208405200 PMID: 12427761

22. Conrad T., Plumbom I., Alcobendas M., Vidal R., and Sauer S. (2020). Maximizing transcription of nucleic acids with efficient T7 promoters. Commun. Biol. 3, 439. https://doi.org/10.1038/s42003-020-01167-x PMID: 32796901

23. Qin D., Liu Q., Devaraj A., and Fredrick K. (2012). Role of helix 44 of 16S rRNA in the fidelity of translation initiation. RNA 18, 485–495. https://doi.org/10.1261/rna.031203.111 PMID: 22279149

24. Chen Z., and Zhang Y. (2005). Dimethyl sulfoxide targets phage RNA polymerases to promote transcription. Biochem. Biophys. Res. Commun. 333, 664–670. https://doi.org/10.1016/j.bbrc.2005.05.166 PMID: 15975554

25. Kao C., Zheng M., and Rüdisser S. (1999). A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. RNA 5, 1268–1272. https://doi.org/10.1017/s1355838299991033 PMID: 10496227

26. Karlsson H., Baronti L., and Petzold K. (2020). A robust and versatile method for production and purification of large-scale RNA samples for structural biology. RNA 26, 1023–1037. https://doi.org/10.1261/rna.075697.120 PMID: 32354720

27. Karlsson H., Feyrer H., Baronti L., and Petzold K. (2021). Production of structured RNA fragments by in vitro transcription and HPLC purification. Curr. Protoc. 1, e159. https://doi.org/10.1002/cpz1.159 PMID: 34138527

28. Zwahlen C., Legault P., Vincent S.J.F., Greenblatt J., Konrat R., and Kay L.E. (1997). Methods for Measurement of Intermolecular NOEs by Multinuclear NMR Spectroscopy: Application to a Bacteriophage λ N-Peptide/boxB RNA Complex. J. Am. Chem. Soc. 119, 6711–6721.

29. Thiele C.M., Petzold K., and Schleucher J. (2009). EASY ROESY: reliable cross-peak integration in adiabatic symmetrized ROESY. Chem. Eur. J 15, 585–588. https://doi.org/10.1002/chem.200802027 PMID: 19065697

30. Feyrer H., Munteanu R., Baronti L., and Petzold K. (2020). One-Pot Production of RNA in High Yield and Purity Through Cleaving Tandem Transcripts. Molecules 25. https://doi.org/10.3390/molecules25051142 PMID: 32143353

31. Parisien M., and Major F. (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature 452, 51–55. https://doi.org/10.1038/nature06684 PMID: 18322526

32. Marušič M., Schlagnitweit J., and Petzold K. (2019). RNA dynamics by NMR spectroscopy. Chembiochem 20, 2685–2710. https://doi.org/10.1002/cbic.201900072 PMID: 30997719

33. Zhang Q., Lv H., Wang L., Chen M., Li F., Liang C., et al. (2016). Recent methods for purification and structure determination of oligonucleotides. Int. J. Mol. Sci. 17. https://doi.org/10.3390/ijms17122134 PMID: 27999357

34. Dagenais P., and Legault P. (2012). Preparative separation of ribonucleoside monophosphates by ion-pair reverse-phase HPLC. Methods Mol. Biol. 941, 247–256. https://doi.org/10.1007/978-1-62703-113-4_18 PMID: 23065566

35. Martino L., and Conte M.R. (2012). Biosynthetic preparation of 13C/15N-labeled rNTPs for high-resolution NMR studies of RNAs. Methods Mol. Biol. 941, 227–245. https://doi.org/10.1007/978-1-62703-113-4_17 PMID: 23065565

36. Goffin C., Bailly V., and Verly W.G. (1987). Nicks 3' or 5' to AP sites or to mispaired bases, and one-nucleotide gaps can be sealed by T4 DNA ligase. Nucleic Acids Res. 15, 8755–8771. https://doi.org/10.1093/nar/15.21.8755 PMID: 3684572

37. Helm M., Brulé H., Giegé R., and Florentz C. (1999). More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. RNA 5, 618–621. https://doi.org/10.1017/s1355838299982328 PMID: 10334331

38. Huang F., Wang G., Coleman T., and Li N. (2003). Synthesis of adenosine derivatives as transcription initiators and preparation of 5' fluorescein- and biotin-labeled RNA through one-step in vitro transcription. RNA 9, 1562–1570. https://doi.org/10.1261/rna.5106403 PMID: 14624011

39. Pitulle C., Kleineidam R.G., Sproat B., and Krupp G. (1992). Initiator oligonucleotides for the combination of chemical and enzymatic RNA synthesis. Gene 112, 101–105. https://doi.org/10.1016/0378-1119(92)90309-d PMID: 1372580

40. Wang S., Chen D., Gao L., and Liu Y. (2022). Short Oligonucleotides Facilitate Co-transcriptional Labeling of RNA at Specific Positions. J. Am. Chem. Soc. https://doi.org/10.1021/jacs.2c00020 PMID: 35293210

41. Collauto A., von Bülow S., Gophane D.B., Saha S., Stelzl L.S., Hummer G., et al. (2020). Compaction of RNA duplexes in the cell*. Angew. Chem. Int. Ed 59, 23025–23029.

42. Daube S.S., and von Hippel P.H. (1992). Functional transcription elongation complexes from synthetic RNA-DNA bubble duplexes. Science 258, 1320–1324. https://doi.org/10.1126/science.1280856 PMID: 1280856

43. Yin S., Ho C.K., and Shuman S. (2003). Structure-function analysis of T4 RNA ligase 2. J. Biol. Chem. 278, 17601–17608. https://doi.org/10.1074/jbc.M300817200 PMID: 12611899

44. Ho C.K., Wang L.K., Lima C.D., and Shuman S. (2004). Structure and mechanism of RNA ligase. Structure 12, 327–339. https://doi.org/10.1016/j.str.2004.01.011 PMID: 14962393

45. Chen H., Cheng K., Liu X., An R., Komiyama M., and Liang X. (2020). Preferential production of RNA rings by T4 RNA ligase 2 without any splint through rational design of precursor strand. Nucleic Acids Res. 48, e54. https://doi.org/10.1093/nar/gkaa181 PMID: 32232357

46. Micura R., and Höbartner C. (2020). Fundamental studies of functional nucleic acids: aptamers, riboswitches, ribozymes and DNAzymes. Chem. Soc. Rev. 49, 7331–7353. https://doi.org/10.1039/d0cs00617c PMID: 32944725

47. Scheitl C.P.M., Lange S., and Höbartner C. (2020). New deoxyribozymes for the native ligation of RNA. Molecules 25. https://doi.org/10.3390/molecules25163650 PMID: 32796587