

Original Paper

Large-Scale *In Vitro* Transcription, RNA Purification and Chemical Probing Analysis

Fariha Kanwal^a Ting Chen^a Yunlon Zhang^a Altaf Simair^a Cai Rujie^a
Najam us Sahar Sadaf Zaidi^c Xinhang Guo^a Xiaolong Wei^a Geoffrey Siegel^b
Changrui Lu^a

^aCollege of Chemistry, Chemical Engineering and Biotechnology, DongHua University, Shanghai, China,

^bDepartment of Orthopaedic Surgery, Musculoskeletal Oncology Division, University of Michigan Medical School, Ann Arbor, MI, USA, ^cDepartment of Industrial Biotechnology, Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad, Pakistan

Key Words

RNA • Regulatory RNA • RNA analysis • *In vitro* transcription • RNA purification • Size-exclusion FPLC • SHAPE • RNA modification • Chemical probing

Abstract

Background/Aims: RNA elements such as catalytic RNA, riboswitch, microRNA, and long non coding RNA (lncRNA) play central roles in many cellular processes. Studying diverse RNA functions require large quantities of RNA for precise structure analysis. Current RNA structure and function studies can benefit from improved RNA quantity and quality, simpler separation procedure and enhanced accuracy of structural analysis. **Methods:** Here we present an optimized protocol for analyzing the structure of any RNA, including *in vitro* transcription, size-exclusion chromatography (SEC) based denaturing purification and improved secondary structure analysis by chemical probing. **Results:** We observed that higher Mg²⁺, nucleoside triphosphate (NTP) concentrations and longer reaction duration can improve the RNA yield from *in vitro* transcription, specifically for longer and more complicated constructs. Our improved SEC-based denaturing RNA purification effectively halved the experiment duration and labor without introducing any contaminant. Finally, this study increased the accuracy and signal-to-noise ratio (SNR) of selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) chemical probing for analyzing RNA structure. **Conclusion:** Part or all of our modified method can improve almost any RNA-related study from protein-RNA interaction analysis to crystallography.

© 2018 The Author(s)
Published by S. Karger AG, Basel

Introduction

Since the discovery of regulatory RNA elements in the 1990s, RNA has emerged a central role in recent studies of gene regulation. These novel regulators include but are not limited to catalytic RNA, riboswitch, microRNA, and lncRNA. Some evidence even suggests that RNA jump started life on Earth and gave rise to all cellular processes today [1, 2]. Function this diverse requires precise understanding of the structural elements of RNA and we believe the manner in which RNA folds, is a key component and warrants further investigation.

Complete biochemical analyses of RNA-related topics often include structural determination and dynamic studies that stress upon both the quantity and quality of RNA samples. Despite advancements in overexpression of RNA in cells [3, 4], *in vitro* transcription is still the predominant method due to its simplicity and broad applicability [5-8]. However, the quality and quantity of RNA produced in an *in vitro* transcription reaction can vary depending upon many factors including the concentration and ratio of individual components, incubation time and temperature. Additionally, since ribozymes are frequently included to generate homogeneous 3' and 5' ends [9-11], their cleavage efficiency upon completion of transcription also warrants further optimization to increase yield and purity.

After the transcription and ribozyme cleavage, the mixture typically requires lengthy separation to produce full-length mature RNA [5, 9, 11]. Traditionally, the purification requires denaturing polyacrylamide gel electrophoresis (urea-PAGE) to achieve separation of the target RNA from RNAP, DNA template, uncleaved precursor RNAs, abortive transcripts and NTPs. Although effective, this process has several major drawbacks. First, it requires minimum two days to complete, making it the rate limiting step. Second, the elution step frequently leaves some RNA trapped in the gel, lowering the total yield. Thirdly, the eluate often contains acrylamide, lowering the purity of the final product [8, 12-15]. The alternative approaches of urea-PAGE includes protein-assisted affinity purification [16, 17] and fast performance liquid chromatography (FPLC) based methods [18-22]. Affinity tag purification methods typically require a ribozyme cleavable tag (i.e. tandem MS2 tags) that binds to his-tagged MS2 proteins. Followed by additional SEC, this reliable technique can produce milligrams of crystallographic quality RNA of small to medium size. Larger target RNA (200+nt) significantly reduces the efficiency of this technique. Alternatively, recent SEC-based methods effectively separate target RNA with unwanted products and reactants under native conditions. However, this technique currently requires lengthy pretreatment steps to remove RNAP and desalt the transcription mix. A newer variation of the FPLC-based method uses weak-anion-exchange chromatography without removing the RNAP. Despite providing a single-step, rapid purification of target transcript at high efficiency, this method cannot separate multiple ribozyme cleavage products with the target RNA.

After isolating the target RNA, the next goal is to comprehend its structure-function relationship. A number of techniques including but not limited to crystallography, NMR and various chemical/inline probing assays analyze RNA structural and functional properties. Among them, SHAPE chemistry allows quantification of local nucleotide flexibility at single nucleotide resolution in any RNA. SHAPE chemistry exploits the reactivity of the RNA ribose 2'-OH group using small hydroxyl-selective electrophilic reagents. In contrast to other chemical probing techniques, SHAPE reactivities detect and quantify local nucleotide flexibility and dynamics, ignoring base identity [23-30]. Such reactivity profile correlates with secondary/tertiary structure elements, ligand binding pockets and protein recognition sites [31-35]. Current SHAPE protocol, optimized against RNA transcripts less than 200 nucleotides, lacks the SNR for longer RNA templates, possibly due to relatively low processivity of the reverse-transcriptase [27, 36]. Recent studies circumvented this issue on larger RNA templates by performing multiple SHAPE assays with different primers to cover the entire length of RNA [35, 37, 38]. While this approach is effective at obtaining sufficient SNR, it requires extensive and careful normalization of SHAPE reactivity to avoid artifacts from multiple independent reverse-transcription (RT) reactions. Currently, SHAPE experiments produce

weak signals those typically need highly sensitive equipment such as capillary sequencers or expensive laser fluorescent scanners.

To combat these difficulties in analyzing RNA, we present here a complete protocol optimization of RNA analysis, including *in vitro* transcription, RNA purification and SHAPE chemical probing (Fig 1). First, our modified *in vitro* transcription method optimized both the transcription and ribozyme cleavage reaction, increasing the target RNA (up to 450 nt) yield by almost twofold. Next, our SEC-based denaturing purification method reduced both the number of steps and total time of purification while maintaining high yield and purity. Lastly, we modified the current SHAPE protocol to enhance the accuracy and SNR of SHAPE reactivities for larger RNA constructs. Overall, our new protocol for producing, isolating and chemically probing RNA substantially increases RNA yield, purity and confidence in structural analysis, consequently making the protocol more robust and reliable to a wider range of RNA molecules.

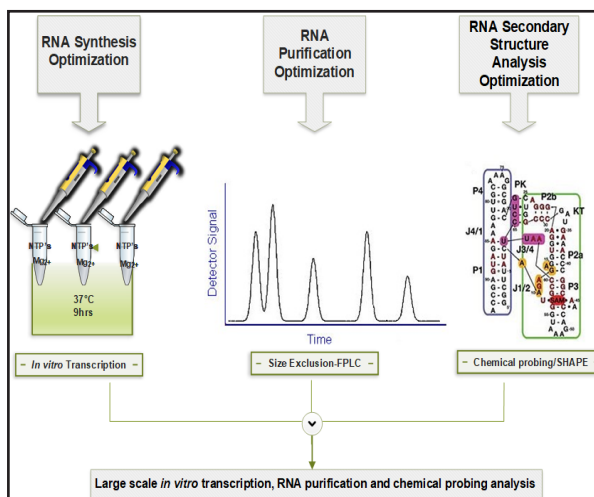


Fig. 1. Optimization schema for large-scale *in vitro* transcription, RNA purification and chemical probing analysis.

Materials and Methods

General

Plasmid DNA was obtained by growing *E. coli* strain DH5 α at 37°C on standard LB agar plates or in aerated LB liquid medium with proper antibiotics. The oligonucleotides for this work were all obtained from Sangon Biotech. The enzymes and RT kits were purchased from Invitrogen (ThermoFisher), and standard chemicals were obtained from Ambion, Inc. and Sigma-Aldrich. RNA purification was conducted on a GE Akta Purifier 100 FPLC system (GE Healthcare) with Superdex-75 and Superdex-200 filtration media packed into 890/10 mm and XK 16/1000 mm columns (GE Healthcare). The buffer used for purification was Tris-ethylenediaminetetraacetic acid (EDTA) containing 8 mM urea and 100 mM NaCl.

Construction and preparation of DNA template

The design of RNA *in vitro* transcription constructs followed standard protocol [11, 39]. The aptamer domain of the *B.subtilis yitJ* S-box riboswitch (S-box; 119 nt) was inserted into a pUC57 plasmid under the control of a T7 RNAP promoter; the hepatitis delta virus (HDV) ribozyme was positioned at the 3'-end and the hammerhead (HH) ribozyme at the 5'-end [9, 10]. We used EcoRI and BamHI restriction sites.

The S-box RNA for chemical probing experiments was constructed to be flanked by a 5'-T7 RNAP promoter and a 3'-RT start site using overlapping PCR and inserted into pUC57 vector [30, 40]. After sequence verification, the SHAPE cassette DNA templates were PCR amplified using terminal primers and used directly to produce RNAs for SHAPE analysis by *in vitro* transcription as explained [39, 41].

The DNA templates for transcription were generated by PCR [1 mL; containing 20 mM Tris (pH 8.4), 50 mM KCl, 2.5 mM Mg²⁺, 200 μ M each dNTP, 500 nM each forward and reverse primer, 5 pM template, and 0.025 units/ μ L of Taq polymerase; (denaturation at 94°C, 20 s; annealing 60°C, 25 s; and elongation 72°C, 30 s; 35 cycles).

In vitro transcription of RNA

RNA was prepared by a standard T7 RNAP run-off transcription reaction using PCR product from the previous step as a template and purified by urea-PAGE as described [5, 8]. The RNA yield from *in vitro* transcription was optimized for each individual DNA template in 25 μ L trial reactions by varying the concentration of Mg^{2+} , NTPs [22] and incubation time. A typical large-scale 10 mL transcription reaction mixture contained 30 mM Tris (pH 8.1 at 37°C), 15 mM Mg^{2+} , 10 mM dithiothreitol (DTT), 2 mM spermidine, 0.01% (v/v) Triton X-100, 4 mM each NTP, 1 mL of PCR-generated DNA template, and 0.1 mg/mL of T7 RNAP [6, 8]. After 2.5 h of incubation at 37°C, pyrophosphate, which forms during *in vitro* transcription reaction, was pelleted down by centrifugation, and additional Mg^{2+} added to the reaction. The reaction continued till 5 h. To concentrate the reaction mixture, we used Millipore centrifugal filter units with appropriate MWCO. The transcription reaction screening was composed of one variable component at a time with the rest of components fixed. The tested Mg^{2+} concentrations included 5 mM, 15 mM, 25 mM, 35 mM, 45 mM, 55 mM, 65 mM, 75 mM, 85 mM and 95 mM while NTPs concentrations were 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM and 10 mM. The incubation time for transcription at 37°C was tested at 5 h, 6 h, 7 h, 8 h, 9 h and 10 h. The results were evaluated through image quantification (BioRad) of target RNA band on 12% TAE urea-PAGE.

Size-exclusion FPLC based RNA purification

RNA sample preparation for chromatographic purification

Transcription mixtures for the S-box RNA were heated to 90°C for 5 min, immediately transferred to ice for 5 min and then pelleted down (centrifuged at 16,000g for 10 min). The supernatant was collected and mixed as 1:4 with Tris-EDTA buffer containing 8 M urea and 100 mM NaCl before loading onto size-exclusion FPLC (SEC) columns. Sample pre-load treatment was operated to remove the T7 RNAP, residual pyrophosphates and aggregates formed during *in vitro* transcription.

Denaturing chromatographic purification

The glass columns (890/10 mm) and (xk16/1000 mm) were packed according to manufacturer's specifications with gel filtration matrix (Superdex-75, Superdex-200) and equilibrated in a cold room (4°C) with several column volumes (CV) of specified buffer. The transcription mixture from the previous step was loaded directly onto the equilibrated SEC column, and the chromatography was performed at 1 mL/min flow rate, collecting 2 mL fractions. Fractions were analyzed by denaturing urea-PAGE. S-box RNA containing fractions were combined and concentrated to 0.5 mL using Millipore centrifugal filter units (EMD Millipore), washed twice with the appropriate buffer for chemical probing analysis. All experiments were performed at 4°C. The efficiency of proposed purification system was confirmed through SHAPE experiments for S-box RNA.

Optimizing RNA chemical probing (SHAPE)

The S-box RNA was first optimized at concentrations of 0.2 μ M, 0.4 μ M, 0.8 μ M, 1 μ M, 1.5 μ M, 2.0 μ M, 3.0 μ M and 4.0 μ M through RT by employing 1 μ L reverse transcriptase; analyzed through a BioRad FX scanner and capillary electrophoresis.

Structure-sensitive RNA modification and Primer Extension

The S-box RNA was subsequently diluted into RNA folding buffer [111 mM Hepes (pH 8.0), 6.7 mM Mg^{2+} , 111 mM NaCl] to 2.0 μ M. Immediately prior to chemical probing, the diluted RNA was refolded as previously described [41] during the annealing process. SHAPE chemical probing was performed as described previously [29, 30]. Refolded 10 μ L of the 2.0 μ M S-box RNA was pre-equilibrated using a gradient PCR machine to 65°C for 10 min and subsequently mixed with 0.1 μ L, 0.5 μ L and 1 μ L of 130 mM N-methylisatoic anhydride (NMIA) in anhydrous dimethyl sulfoxide (DMSO) for best modification optimization by initiating 2'-ribose hydroxyl alkylation. The reactions proceeded to completion by incubating for five half-lives of NMIA; later quenched with 900 μ L of precipitation buffer containing 80% EtOH, 45 μ M NaCl, 0.45 μ M EDTA, and 2 μ L glycoBlue (15 mg/mL). The RNA was precipitated by incubation to -80°C for 30 min, followed by centrifugation at 16,000g to 4°C for 45 min. The pellet was air-dried and re-suspended in 10 μ L of 0.5 \times TE

buffer [39]. The primer extension protocol was similar to that described previously [29]. A fluorescently labeled DNA primer (6-FAM-GAACCGGACCGAAGCCCG; 0.5 μ L, 50 μ M) was annealed to the recovered RNA (10 μ L) by incubation at 65°C for 2 min, then 35°C for 5 min, followed by snap cooling to 4°C for another 2 min. The resulting 10.5 μ L of RNA-primer mixture was mixed with 6 μ L of RT buffer [167 mM Tris (pH 8.3), 250 mM KCl, 10 mM Mg²⁺, 1.67 mM of each dNTP], preheated to 52°C for 1 min, and then incubated with 1 μ L of Superscript III (Invitrogen, 200 units) at 52°C for 20 min. The reactions were quenched, and the RNA template was destroyed by addition of 1 μ L of 5 M NaOH, followed by heating to 90°C for 4 min. Acid stop mix [4:25 (v/v) mixture of 1 M unbuffered Tris-HCl and stop solution (85% formamide, 0.5 \times TBE, 50 mM EDTA, pH 8.0); 29 μ L] was added to each reaction; the reaction was incubated for an additional 5 min to 90°C and then cooled to -20°C. No internal tracking dye was used in the reaction to avoid interference with fragment analysis. Each set of primer extension reactions was first evaluated by 12% TAE urea-PAGE followed by sequencing of band intensities, performed at Sangon Biotech. Each SHAPE reaction product was analyzed using primers labeled with distinct fluorophores as described by [27, 35, 42].

Results

Increasing RNA yield through in vitro transcription optimization

Typical RNA structural and functional studies require milligrams of RNA from an *in vitro* transcription assay, and its yield depends on Mg²⁺ concentration, NTP concentrations and incubation time [18, 19, 21, 43]. We opted not to use pyrophosphatase in our protocol due to its high cost and negligible impact on the transcription yield for large-scale experiments (data not shown). This study systematically varied these parameters to maximize the transcription efficiency. We used a previously characterized S-box RNA (119 nt) as the target RNA. This target RNA is flanked by two ribozymes that self-cleaves after transcription to give the final product as described [9, 10, 39].

Systematic variation of Mg²⁺, NTP concentrations and reaction duration revealed that the optimum value lies outside the standard condition. The target RNA yield peaks at 180% of stock condition, around 45 mM Mg²⁺ (Fig 2A). This increase in yield also coincides with the decrease in the uncleaved precursor transcript (top band), which suggests that the RNA self-cleavage reaction requires higher Mg²⁺ concentration. Further increase in Mg²⁺ beyond 45 mM leads to diminished RNA production (Fig 2A, lane 5-10).

We then independently varied the individual NTP concentration centered on the stock concentration of 4 mM each. In general, the RNA yield increases from 1 mM up to 10 mM NTP, plateauing around 8 mM each (Fig 2B, lane 8). At concentrations beyond 4 mM, we observed increased precursor RNA as well as increased target RNA (Fig 2B, lane 5-10). Quantification of the gel image shows that the target RNA yield increased 100% at 8 mM individual NTPs concentration.

Lastly, we extended the reaction time and observed a steady increase of product until 9 hours. At 37°C reaction temperature, the reaction continued to produce increased RNA product comparing with the standard 5 hours reaction time (Fig 2C). Image quantification of the target RNA yield shows 20% increase compared to the standard condition with a subsequent decrease afterwards.

Combining all three experiments above, we tested the cumulative effect of all three new conditions. Fig. 2D shows the comparison of our optimized transcription protocol with the standard condition through 12% urea-PAGE and its image quantification. In lane 1, before optimization, the *in vitro* transcription product appears much weaker on the gel than the optimized condition in lane 2. Quantification reveals that the optimized conditions (45 mM Mg²⁺, 8 mM individual NTP and 9 hours reaction time) increase the target RNA production by twofold. We further test this condition on various other RNA constructs ranging from 200 to 700 nucleotides and obtained similar results (data not shown).

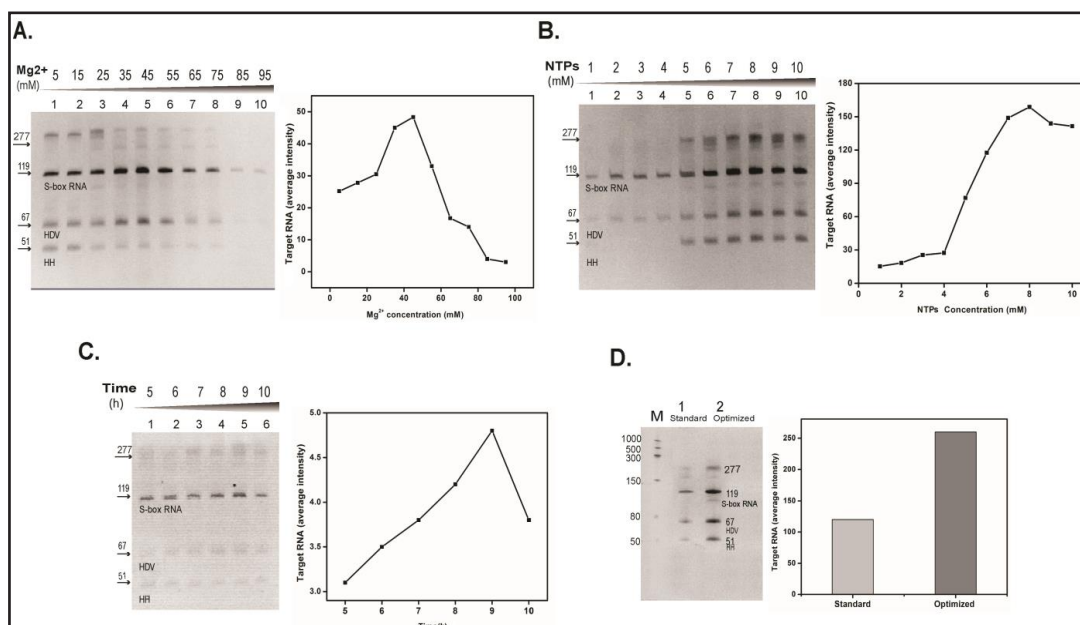


Fig. 2. The S-box RNA (119 nt) *in vitro* transcription optimization. (A) Optimizing Mg^{2+} concentration (mM). Left: Various Mg^{2+} concentrations (lanes 1-10) for *in vitro* transcription reaction separated by 12% TAE urea-PAGE. Arrows indicate target RNA and ribozymes. S-box RNA intensity gradually increased from 5 mM to 45 mM (lane 1-5), attaining the maximum target RNA yield at 45 mM, followed by gradual decrease (lane 6-10). Right: Quantified transcription profile revealed 80% increase in S-box RNA synthesis at 45 mM Mg^{2+} concentration. (B) Effect of individual NTP concentrations (1 mM to 10 mM). Left: Transcription efficiency plateaued around 8 mM individual NTP concentration visualized by 12% TAE urea-PAGE. Right: Quantified transcription profile revealed almost 100% increase in S-box RNA production at 8 mM individual NTP concentrations. (C) Optimizing incubation time (h) at 37°C. Left: Optimized incubation hours (lanes 1-6) for *in vitro* transcription reactions separated by 12% TAE urea-PAGE. A gradual increase was observed in S-box RNA intensity (lane 1-5), attaining the maximum target RNA yield (lane 5) followed by gradual decrease in target RNA synthesis (lane 6). Right: Quantified transcription profile revealed 20% increase in S-box RNA synthesis at 9 h incubation time. (D) Comparison of combined pre- and post-optimization steps depicted through 12% TAE urea-PAGE. Lane M is RNA marker. Left: The S-box RNA *in vitro* transcription profile by following standard method (lane 1) and cumulative effect of optimized protocol (lane 2). Right: Quantification of optimized target RNA yield through bar chart depicted nearly twofold increase in average target RNA intensity. The target RNA and ribozymes are labeled.

Optimizing SE-FPLC based purification of transcription products

Precise analysis of RNA requires removal of unwanted components such as the DNA template, T7 RNAP, unincorporated NTPs, small abortive RNA transcripts and various ribozyme fragments from the transcription mix. Since those components vary in size, we aim to efficiently separate them regardless of their shape and charge properties. In order to avoid introducing any acrylamide or other contaminants, we chose to use size-exclusion chromatography under denaturing conditions. First, we boiled the transcription mix and kept the supernatant. This step effectively removed all RNAP from the mix as we detected no RNAP in any subsequent steps through SDS gel (data not shown). Using Superdex-75 and -200 (16/890) prep grade columns, we obtained an elution profile of target RNA with distinct peaks. For a simple RNA transcription reaction where only one target RNA is expected, figure 3A chromatogram shows two major peaks after each injection, 100 μ L and 500 μ L respectively. The first peak contains the target S-box RNA within one CV of elution, and the second peak contained unincorporated NTPs and/or small abortive transcripts. Target RNA was pooled out in total 8 fractions containing 2 mL each while residual/deactivated T7 RNAP and template

DNA were not detected in any fractions collected (data not shown). This is possible because the majority of RNAP were removed during the first denaturing and precipitation step, and the remaining RNAP exited with the template in the void fraction. The inset in figure 3A shows SE-FPLC purified S-box RNA fractions (lanes 1-8) visualized through 12% TAE urea-PAGE. Fig. 3B shows the chromatogram of a full-scale purification with 5 mL sample loading (3% CV). Our result shows that within the limits of the SEC column, the elution profiles are consistent. The fractions containing the desired RNA were combined, concentrated and buffer exchanged to appropriate conditions using Millipore centrifugal filter units. Our purification method yields > 99% pure target RNA without NTPs, abortive transcripts or acrylamide contaminants. We also shortened the purification process to eight hours or less, avoiding lengthy gel separation and extraction that typically takes two days to complete,

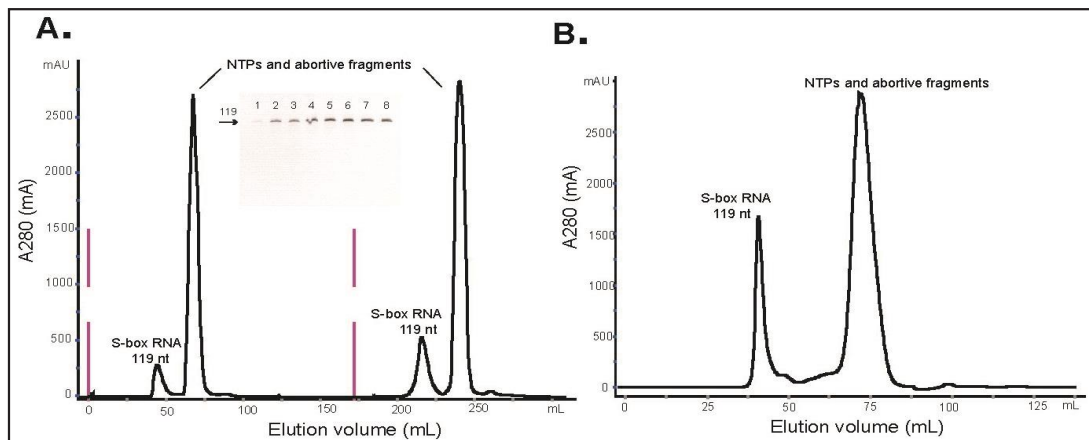


Fig. 3. SE-FPLC based RNA purification. (A) Purification of S-box RNA (from a simple transcription mixture). SE-FPLC was performed at 1 mL/min at 4°C. Elution profile on the left was obtained from 100 μ L sample load volume, and the right one with 500 μ L sample load volume. The inset in (A) shows SE-FPLC purified S-box RNA fractions (lanes 1-8) visualized through 12% TAE urea-PAGE. (B) A scale-up purification profile of S-box RNA through SE-FPLC by loading 5 mL of transcription mixture sample.

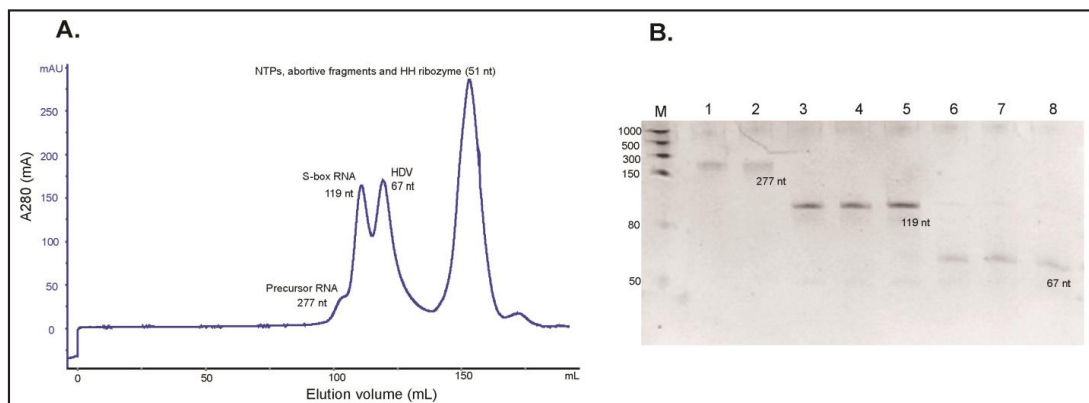


Fig. 4. SE-FPLC based RNA purification from a more complex transcription mixture. (A) Purification of S-box RNA (119 nt) from RNA crystallization construct. Elution profile, obtained through Superdex-200 (xk16/1000) column with sample load volume of 100 μ L has three distinct peaks. The shoulder of the first peak contained the unspliced precursor RNA transcript (277 nt). First peak carried over S-box RNA (119 nt) spliced from both ribozymes, and the HDV ribozyme (67 nt) pooled out in second peak. Whereas, NTPs and small abortive transcription products including HH ribozyme (51 nt) eluted out in third large peak. (B) A 12% TAE urea-PAGE stained with ethidium bromide confirmed the identity of each purified fraction. Lane M is RNA marker, lane (1-2) represents unspliced precursor S-box RNA (277 nt), lane (3-5) contains purified S-box RNA (119 nt) and lane (6-8) indicates the identity of 67 nt long HDV ribozyme.

all using standard FPLC system and off-the-shelf SEC columns and parts. Experiment duration can be further compressed using pre-packed columns at higher flow rates. Since our SEC-based protocol does not involve urea-PAGE, we can also improve yield by not relying on gel elution. As a result, our RNA yield has a sevenfold improvement (31 mg/mL) as compared to the conventional PAGE purification method (4.2 mg/mL) during parallel experiments.

In addition to purify simple *in vitro* transcription reaction, we also tested our protocol against more complex RNA crystallization constructs. We used a previously characterized

S-box RNA construct containing a 5'- HH and 3'- HDV ribozyme using a Superdex-200 (xk16/1000) SEC column. Chromatography with 100 μ L sample volume was performed at 1 mL/min (4°C), collecting 2 mL fractions. Fig. 4A shows an elution profile with three distinct peaks, the shoulder before the first distinct peak contained the unspliced precursor RNA transcript (277 nt), and the first peak contained the target S-box RNA (119 nt) spliced from both ribozymes. The second peak contained the HDV ribozyme (67 nt) while the third peak contained smaller fragments such as HH ribozyme (51 nt), NTPs and other small abortive products. Our subsequent urea-PAGE on the first two peaks confirmed the identity of the purified target RNA with a fivefold improvement (14.8 mg/mL) as compared to conventional PAGE purification method (2.7 mg/mL) during parallel experiments.

SHAPE optimization for RNA secondary structure analysis

After obtaining pure target RNA in large quantities, the next logical process is to analyze its structure and function. Here we aim to optimize the quantitative RNA structure analysis at single nucleotide resolution through SHAPE. In our study, we focused on the RT step since we have seen major advancements in the commercial reverse transcription development in recent years. We focused at 8 RT experiments with the S-box RNA concentrations of 0.2 μ M, 0.4 μ M, 0.8 μ M, 1 μ M, 1.5 μ M, 2 μ M, 3 μ M and 4 μ M without any chemical modification on the RNA. Fig. 5A shows the urea-PAGE result of the reverse transcription using 6-FAM-labeled primer. Quantification of the band intensities (Fig 5B) shows that the resulting cDNA maxed out at 2 μ M RNA concentration. Adding more than 2 μ M RNA offers no additional increase as the RTase has been saturated. Compared with the standard protocol of 0.2 μ M, we observed a 10.5-fold increase in cDNA yield from reverse transcriptase. The linear increase in band intensity up to 2 μ M suggests that using current generation of superscript reverse transcriptase, the enzyme was in excess and more template RNA can increase the signal linearly up to 2 μ M.

Since we changed the RNA concentration in the RT step, the SHAPE reagent also needs correction to adjust for the RNA quantity to ensure at most one modification per RNA molecule. We therefore fixed the RNA concentration at 2 μ M and explored multiple NMIA concentrations and noticed that reported concentration of 13 mM or higher introduced significant end bias (data not shown). Lowering the NMIA concentration ten-fold to 1.3 mM resolved the issue. Fig. 6A shows the unmodified raw sequencing trace of the cDNA from modification and subsequent reverse transcription during parallel experiments. The red

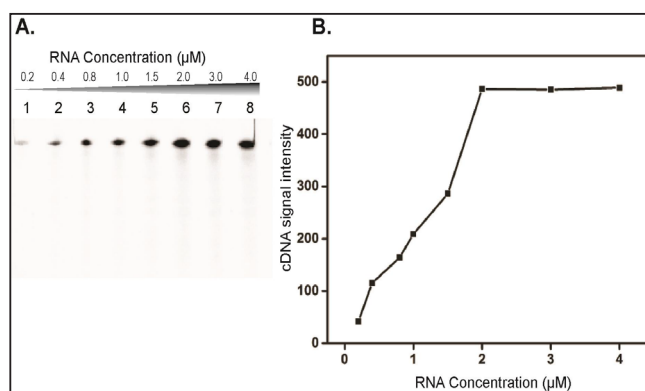


Fig. 5. Optimizing S-box RNA RT step before chemical probing. (A) The urea-PAGE illustration of S-box RNA RT, analyzed through a BioRad FX scanner. Lane 1 shows the cDNA signal at 0.2 μ M RNA concentration, whereas subsequent lanes (2-8) shows gradually improved cDNA signal with the increasing RNA template concentrations. (B) The confirmation of optimized RNA concentration results for RT through image quantification.

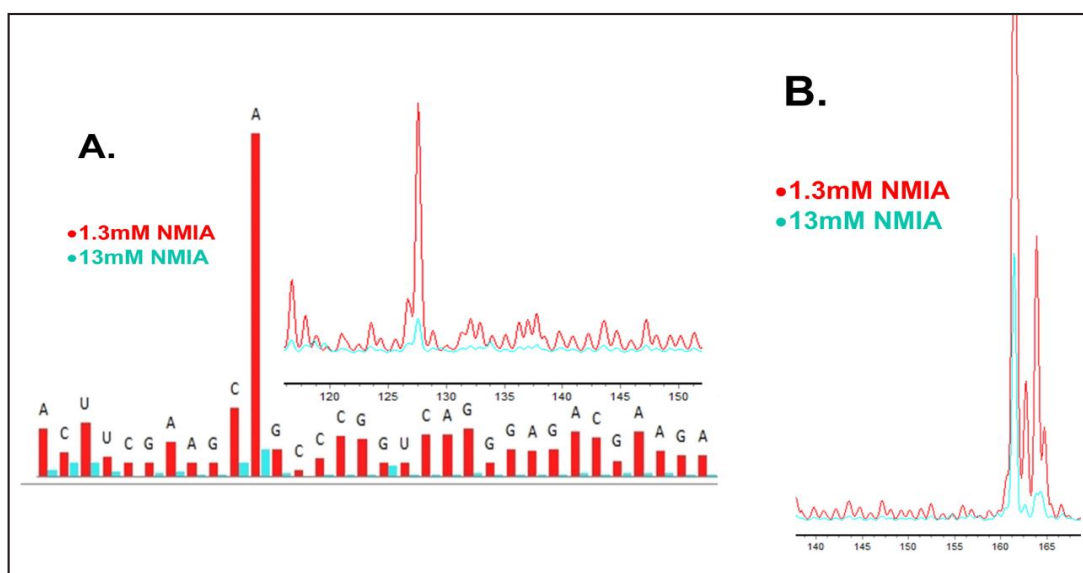


Fig. 6. SHAPE analysis using updated RNA concentration coupled with different NMIA ratios. (A) NMIA reactivity of S-box RNA P2 domain corresponding to nucleotide positions (118-149) at 1.3 mM (red, post-optimized) and 13 mM (cyan, pre-optimized), Inset: RAW trace of the SHAPE reaction from capillary electrophoresis of the above-mentioned region. (B) Raw trace signal corresponding to the full length cDNA pre- and post-optimization. Coloring is consistent throughout this Fig.

trace from the 1.3 mM NMIA concentration has higher SNR than 13 mM NMIA (cyan) (Fig 6A; inset). Pre- and post- optimization for NMIA concentration corresponding to the S-box RNA P2 domain (118-149 nt) shows an average of 13-fold increase in effective signals at 1.3 mM NMIA concentration whereas 13 mM NMIA concentration resulted in marginally detectable signal using the same equipment and parameters. Similarly, lower (1.3 mM) NIMA also generated more full-length cDNA product manifested by much higher signal in Fig. 6B. This result signifies less modification on the RNA (preferably one per RNA chain), reducing signal bias towards the 3' of the target RNA [29]. We further test these conditions on various other riboswitches and obtained similar results (data not shown). Our data indicate [NMIA (mM)/RNA(μ M)=13/20] is the best working ratio to evaluate S-box RNA secondary structure through SHAPE.

Discussion

Current structural and biochemical studies have demonstrated that RNA actively participates in all aspects of cellular processes. Fast, efficient RNA preparation and analysis greatly facilitate RNA structure and functional studies. This paper provides an improved guide to analyze the structure of any RNA, from generating the RNA to secondary structure analysis by SHAPE.

During transcription optimization, we observed that higher than reported Mg^{2+} NTP, concentrations and longer duration can improve the RNA yield from *in vitro* transcription, specifically for longer and more complicated constructs. The Tris-HCl buffer (40 mM, pH 8.1 at 37°C), 5 mM DTT, 0.01% Triton-X100 and 1 mM spermidine were selected as the optimal transcription reaction conditions [44]. Raising the concentrations of Mg^{2+} up to 45 mM increased the amount of transcribed target RNA. Since we opted not to use pyrophosphatase, magnesium pyrophosphate precipitates out of the solution as a result of NTP incorporation and drive the reaction forward. Additionally, HH and HDV ribozymes both require magnesium to catalyze their self-cleavage [45-49]. Consequently, we found that higher initial concentration offsets the loss of magnesium ions to the pyrophosphate precipitates and facilitates ribozyme

cleavage. Similarly, during our trials, we observed a linear increase of transcription efficiency and no apparent adverse effects with increased NTP concentrations up to 8 mM (32 mM total), two to sixteen-fold that of previously reported ranges between 0.5 to 4 mM [11, 18-20, 22, 50, 51]. We speculate that this discrepancy results from the much larger volumes of our reactions (5 to 15 mL) comparing 10 to 50 μ L reported reaction volumes. The increased reaction volume may decrease effective local reactant concentrations, reducing effective collision between active reactants. Therefore, higher initial reactant concentrations of NTPs and Mg^{2+} both positively affect *in vitro* transcription yield, consistent with previous reports [18, 19, 21, 43].

Using the standard urea-PAGE, isolating the target RNA typically takes two or more days and may contain unwanted acrylamide [22]. Alternative approaches such as ion exchange [18, 19] require specific conditions for different RNA molecules. Previous SEC-based studies [20, 21, 52] require either toxic phenol/chloroform extraction or addition of proteinase K to remove RNAP. Our improved method based on denaturing gel-filtration effectively halved the experiment duration and the number of steps without introducing any contaminant. In addition, the method described above allows ribozyme modifications for crystallographic studies and is insensitive to target RNA conformation. Overall, this improved method should apply to most medium to large RNAs for biophysical or biochemical studies. Additional scaling-up only require similar modifications (i.e. slightly higher reactants concentration) and higher capacity off-the-shelf SEC components.

Finally, this study increased the SNR of SHAPE chemical probing for analyzing RNA structure. SHAPE technique has seen generations of improvements since first introduction [30], such as new probing reagent [42], better time resolution [53, 54] and *in vitro* probing techniques [26, 55]. This study provides a further optimization that benefits all SHAPE techniques. We focused on two SHAPE reaction components, RNA and NMIA concentrations. Increasing RNA concentrations up to 2 μ M with the constant amount of reverse transcriptase offers a marked increase in cDNA synthesis. To limit only one modification per RNA for unbiased SHAPE result [30, 56], additional RNA warrants altered concentrations of modifying reagent NMIA to yield 2'-O-adducts at flexible nucleotide positions. Initially, we speculate that more NMIA is required to maintain a constant ratio between RNA and NMIA as reported [29]. However, the ten-fold RNA increase required much lower (also ten-fold) NMIA concentration. We speculate that increased RNA concentration also increases the probability of RNA-NMIA collision, out-competing the parallel NMIA hydrolysis reaction. Therefore, further increasing RNA concentration in SHAPE may require lower NMIA to avoid 3' bias caused by multiple modifications. Our data suggest that minimizing secondary NMIA addition to the target RNA may require maintaining a constant RNA-NMIA reaction rate. Therefore, higher RNA concentration warrants lower NMIA concentration (rate = k [RNA]^m [NMIA]ⁿ). The linear decrease of NMIA concentration, we observed is consistent with previous observation that longer RNAs need a similar linear decrease of NMIA concentration [29]. Since longer RNAs have more sites available for NMIA attack, lowering reaction duration and/or NMIA concentration will significantly improve modification signals and eliminate end bias [29, 57]. Consequently, individual experiments with different RNA lengths may require an additional tweak of NMIA. We recommend keeping the RNA concentration constant at 2 μ M for optimal result, and in case lower RNA amount is used, slightly higher NMIA may be desired.

We consider this fast, efficient RNA preparation and analysis guide as a worthwhile approach for RNA researchers at almost all steps studying RNA structure/function. Our improved *in vitro* transcription conditions noticeably amplify the magnitude of RNA. A modified gel-filtration based purification system separates the target RNA with transcription mixture components, being useful as a comparatively rapid, less-laborious and contamination-free system. Finally, an optimized method increases the accuracy and SNR of SHAPE chemical probing for analyzing RNA structure. In short, our improved RNA study guide can benefit the research world for any RNA-related study from protein-RNA interaction analysis to crystallography.

Acknowledgements

We thank Dr. Chen, Peiran, and Dr. Muhammad Aqeel for helpful discussions. Funding was provided by the National Natural Science Foundation of China (No.31300603), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (No.2012-28), Fundamental Research Funds for the Central Universities (No.15D110527, No.15D110508, No.13D110522, No.15D110568, No.15D310523), the National College Student Innovation Experiment Program (No.14T10501, 17D210502), General Financial Grant from the China Postdoctoral Science Foundation (2015M571455). We also acknowledge the China Scholarship Council (2014GXY252) for sponsoring the PhD fellowship.

Disclosure Statement

The authors declare no conflicts of interest.

References

- 1 Sharp PA: The Centrality of RNA. *Cell* 2009;136:577-580.
- 2 Lilley DMJ: The origins of RNA catalysis in ribozymes. *Trends Biochem Sci* 2003;28:495-501.
- 3 Ponchon L, Dardel F: Recombinant RNA technology: the tRNA scaffold. *Nat Methods* 2007;4:571-576.
- 4 Ponchon L, Beauvais G, Nonin-Lecomte S, Dardel F: A generic protocol for the expression and purification of recombinant RNA in *Escherichia coli* using a tRNA scaffold. *Nat Protoc* 2009;4:947-959.
- 5 Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC: Oligoribonucleotide synthesis using T7 RNA-polymerase and synthetic DNA templates. *Nucleic Acids Res* 1987;15:8783-8798.
- 6 Puglisi JD, Wyatt JR: Biochemical and NMR studies of RNA conformation with an emphasis on RNA pseudoknots. *Methods Enzymol* 1995;261:323-350.
- 7 Ponchon L, Dardel F: Large scale expression and purification of recombinant RNA in *Escherichia coli*. *Methods* 2011;54:267-273.
- 8 Wyatt J, Chastain M, Puglisi J: Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques* 1991;11:764-769.
- 9 Price SR, Ito N, Oubridge C, Avis JM, Nagai K: Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol* 1995;249:398-408.
- 10 FerreDamare AR, Doudna JA: Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of *in vitro* transcribed RNA. *Nucleic Acids Res* 1996;24:977-978.
- 11 Ke A, Doudna JA: Crystallization of RNA and RNA-protein complexes. *Methods* 2004;34:408-414.
- 12 Loening U: The fractionation of high-molecular-weight ribonucleic acid by polyacrylamide-gel electrophoresis. *J Biochem* 1967;102:251.
- 13 Uhlenbeck OC: Keeping RNA Happy. *RNA* 1995;1:4-6.
- 14 Cheong HK, Hwang E, Lee C, Choi BS, Cheong C: Rapid preparation of RNA samples for NMR spectroscopy and X-ray crystallography. *Nucleic Acids Res* 2004;32:10.
- 15 Summer H, Grämer R, Dröge P: Denaturing urea polyacrylamide gel electrophoresis (Urea PAGE). *JoVE* 2009;32:1485.
- 16 Batey RT, Kieft JS: Improved native affinity purification of RNA. *RNA* 2007;13:1384-1389.
- 17 Kieft JS, Batey RT: A general method for rapid and nondenaturing purification of RNAs. *RNA* 2004;10:988-995.
- 18 Koubek J, Lin KF, Chen YR, Cheng RP, Huang JJT: Strong anion-exchange fast performance liquid chromatography as a versatile tool for preparation and purification of RNA produced by *in vitro* transcription. *RNA* 2013;19:1449-1459.
- 19 Easton LE, Shibata Y, Lukavsky PJ: Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA* 2010;16:647-653.

- 20 McKenna SA, Kim I, Puglisi EV, Lindhout DA, Aitken CE, Marshall RA, Puglisi JD: Purification and characterization of transcribed RNAs using gel filtration chromatography. *Nat Protoc* 2007;2:3270-3277.
- 21 Kim I, McKenna SA, Puglisi EV, Puglisi JD: Rapid purification of RNAs using fast performance liquid chromatography (FPLC). *RNA* 2007;13:289-294.
- 22 Lukavsky PJ, Puglisi JD: Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides. *RNA* 2004;10:889-893.
- 23 Rice GM, Leonard CW, Weeks KM: RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA* 2014;20:846-854.
- 24 McGinnis JL, Dunkle JA, Cate JH, Weeks KM: The mechanisms of RNA SHAPE chemistry. *J Am Chem Soc* 2012;134:6617-6624.
- 25 Steen KA, Siegfried NA, Weeks KM: Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease (RNase-detected SHAPE) for direct analysis of covalent adducts and of nucleotide flexibility in RNA. *Nat Protoc* 2011;6:1683-1694.
- 26 Weeks KM: Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol* 2010;20:295-304.
- 27 Deigan KE, Li TW, Mathews DH, Weeks KM: Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A* 2009;106:97-102.
- 28 Gherghe CM, Mortimer SA, Krahn JM, Thompson NL, Weeks KM: Slow conformational dynamics at C2'-endo nucleotides in RNA. *J Am Chem Soc* 2008;130:8884-8885.
- 29 Wilkinson KA, Merino EJ, Weeks KM: Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* 2006;1:1610-1616.
- 30 Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM: RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 2005;127:4223-4231.
- 31 Archer EJ, Simpson MA, Watts NJ, O'Kane R, Wang B, Erie DA, McPherson A, Weeks KM: Long-range architecture in a viral RNA genome. *Biochemistry* 2013;52:3182-3190.
- 32 Gherghe C, Lombo T, Leonard CW, Datta SAK, Bess JW Jr., Gorelick RJ, Rein A, Weeks KM: Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proc Natl Acad Sci U S A* 2010;107:19248-19253.
- 33 Duncan CD, Weeks KM: Nonhierarchical ribonucleoprotein assembly suggests a strain-propagation model for protein-facilitated RNA folding. *Biochemistry* 2010;49:5418-5425.
- 34 Wilkinson KA, Vasa SM, Deigan KE, Mortimer SA, Giddings MC, Weeks KM: Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *RNA* 2009;15:1314-1321.
- 35 Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM: High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* 2008;6:e96.
- 36 Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, Burch CL, Weeks KM: Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 2009;460:711-716.
- 37 Liu F, Somarowthu S, Pyle AM: Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat Chem Biol* 2017;13:282.
- 38 Smola MJ, Christy TW, Inoue K, Nicholson CO, Friedersdorf M, Keene JD, Lee DM, Calabrese JM, Weeks KM: SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc Natl Acad Sci U S A* 2016;113:10322-10327.
- 39 Lu C, Ding F, Chowdhury A, Pradhan V, Tomsic J, Holmes WM, Henkin TM, Ke A: SAM Recognition and Conformational Switching Mechanism in the *Bacillus subtilis* yitJ S Box/SAM-I Riboswitch. *J Mol Biol* 2010;404:803-818.
- 40 Lu C, Smith AM, Ding F, Chowdhury A, Henkin TM, Ke A: Variable sequences outside the SAM-binding core critically influence the conformational dynamics of the SAM-III/SMK box riboswitch. *J Mol Biol* 2011;409:786-799.
- 41 Lu C, Smith AM, Fuchs RT, Ding F, Rajashankar K, Henkin TM, Ke A: Crystal structures of the SAM-III/S-MK riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nat Struct Mol Biol* 2008;15:1076-1083.
- 42 Mortimer SA, Weeks KM: A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* 2007;129:4144-4145.

Fariha et al.: An Improved Guide for Large-Scale *in Vitro* Transcription, RNA Purification and Chemical Probing

- 43 Sherlin LD, Bullock TL, Nissan TA, Perona JJ, Lariviere FJ, Uhlenbeck OC, Scaringe SA: Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization. *RNA* 2001;7:1671-1678.
- 44 Chamberlin M, Ring J: Characterization of T7-specific ribonucleic acid polymerase I. General properties of the enzymatic reaction and the template specificity of the enzyme. *J Biol Chem* 1973;248:2235-2244.
- 45 Heus HA, Pardi A: Nuclear magnetic resonance studies of the hammerhead ribozyme domain: secondary structure formation and magnesium ion dependence. *J Mol Biol* 1991;217:113-124.
- 46 Perreault JP, Labuda D, Usman N, Yang JH, Cedergren R: Relationship between 2'-hydroxyls and magnesium binding in the hammerhead RNA domain: a model for ribozyme catalysis. *Biochemistry* 1991;30:4020-4025.
- 47 Lafontaine DA, Ananvoranich S, Perreault J-P: Presence of a coordinated metal ion in a trans-acting antigenomic delta ribozyme. *Nucleic Acids Res* 1999;27:3236-3243.
- 48 Veeraraghavan N, Ganguly A, Golden BL, Bevilacqua PC, Hammes-Schiffer S: Mechanistic strategies in the HDV ribozyme: chelated and diffuse metal ion interactions and active site protonation. *J. Phys Chem B* 2011;115:8346-8357.
- 49 Tanaka Y, Tagaya M, Hori T, Sakamoto T, Kurihara Y, Katahira M, Uesugi S: Cleavage reaction of HDV ribozymes in the presence of Mg²⁺ is accompanied by a conformational change. *Genes Cells* 2002;7:567-579.
- 50 Milligan JF, Uhlenbeck OC: Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* 1989;180:51-62.
- 51 Nilsen TW, Rio DC, Ares M: High-yield synthesis of RNA using T7 RNA polymerase and plasmid DNA or oligonucleotide templates. *Cold Spring Harb protoc* DOI:10.1101/pdb.prot078535.
- 52 Chillón I, Marcia M, Legiewicz M, Liu F, Somarowthu S, Pyle AM: Native purification and analysis of long RNAs. *Methods Enzymol* 2015;558:3-37.
- 53 Mortimer SA, Weeks KM: Time-resolved RNA SHAPE chemistry. *J Am Chem Soc* 2008;130:16178-16180.
- 54 Mortimer SA, Weeks KM: Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat Protoc* 2009;4:1413.
- 55 Knapp G: Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol* 1989;180:192-212.
- 56 Wilkinson KA, Merino EJ, Weeks KM: RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA Asp transcripts. *J Am Chem Soc* 2005;127:4659-4667.
- 57 Seetin MG, Kladow W, Bida JP, Das R: Massively parallel RNA chemical mapping with a reduced bias MAP-seq protocol. *Methods Mol Biol* 2014;1086:95-117.