

Mapping RNA Structure In Vitro with SHAPE Chemistry and Next-Generation Sequencing (SHAPE-Seq)

Kyle E. Watters and Julius B. Lucks

Abstract

Mapping RNA structure with selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) chemistry has proven to be a versatile method for characterizing RNA structure in a variety of contexts. SHAPE reagents covalently modify RNAs in a structure-dependent manner to create adducts at the 2'-OH group of the ribose backbone at nucleotides that are structurally flexible. The positions of these adducts are detected using reverse transcriptase (RT) primer extension, which stops one nucleotide before the modification, to create a pool of cDNAs whose lengths reflect the location of SHAPE modification. Quantification of the cDNA pools is used to estimate the “reactivity” of each nucleotide in an RNA molecule to the SHAPE reagent. High reactivities indicate nucleotides that are structurally flexible, while low reactivities indicate nucleotides that are inflexible. These SHAPE reactivities can then be used to infer RNA structures by restraining RNA structure prediction algorithms. Here, we provide a state-of-the-art protocol describing how to perform in vitro RNA structure probing with SHAPE chemistry using next-generation sequencing to quantify cDNA pools and estimate reactivities (SHAPE-Seq). The use of next-generation sequencing allows for higher throughput, more consistent data analysis, and multiplexing capabilities. The technique described herein, SHAPE-Seq v2.0, uses a universal reverse transcription priming site that is ligated to the RNA after SHAPE modification. The introduced priming site allows for the structural analysis of an RNA independent of its sequence.

Key words SHAPE, SHAPE-Seq, RNA, RNA structure probing, RNA structure mapping, Next-generation sequencing, RNA structure, RNA folding

1 Introduction

Mapping RNA structure with chemical probes has become a powerful technique for uncovering RNA structure–function relationships in a broad array of contexts [1]. Chemical probing experiments use reagents that covalently modify RNAs in a structure-dependent fashion, allowing structural properties of an RNA under study to be inferred once the locations of the modifications are determined. Although chemical probing structural information is lower resolution than that achievable with biophysical methods such as X-ray crystallography and NMR [1, 2], the experimental speed, flexibility,

and accessibility of RNA chemical probing experiments have made them amenable to an increasing number of innovative RNA structural biology studies. In recent years, chemical probing techniques have become powerful and have been extended to investigate a variety of topics including: long-range RNA–RNA interactions [3], ribosomal assembly [4], genome-wide RNA structures inside cells [5–7], viral genome organization [8], and high-resolution structure prediction and modeling [9, 10].

RNA structure probing experiments consist of several distinct steps: preparation and folding of an RNA of interest, structure-dependent covalent modification of the RNA at the nucleotide level, and determination of the modification locations [1]. Since many chemical probes preferentially modify nucleotides that are unstructured, a higher frequency of modification in specific regions can be used to infer the presence of single stranded regions, loops, or bulges (Fig. 1) [14–16]. While there are a wide variety of chemicals that can be used to probe RNA structure [1, 14, 16, 17], here we will focus on the SHAPE class of chemical probes [1, 18, 19]. SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) reagents react with the 2'-OH group of the ribose backbone of an RNA to form covalent adducts at nucleotides that are structurally flexible [15, 20]. The positions of these adducts are then detected with reverse transcriptase (RT) primer extension, which stops one nucleotide before the modification, to create a pool of cDNAs whose lengths reflect the location of SHAPE modification (+ channel) [21]. A control RT primer extension on an unmodified RNA (– channel) is also performed to identify locations where the RT has a natural propensity to abort extension and “drop off”. The (+) and (–) channel cDNAs can then be used to estimate a “reactivity” for each nucleotide in the RNA, where “reactivity” represents the likelihood that a given nucleotide within an RNA sequence will be modified by the SHAPE reagent [22, 23]. High SHAPE reactivities correspond to nucleotides that are unstructured, and are more likely to occur in single-stranded regions, loops, or bulges (Fig. 1). Conversely, low reactivities can correspond to constrained nucleotides located in double-stranded helices, bound to a protein or ligand, or involved in a noncanonical base pair, stacking, or tertiary interaction [24]. In addition to qualitative interpretation, SHAPE reactivity data can also be used quantitatively to restrain secondary structure prediction algorithms to generate RNA structure models that are more consistent with experimental measurements [25, 26].

To estimate SHAPE reactivities, the location and frequency of SHAPE modifications and natural RT drop off sites need to be determined from the cDNA pools. Originally, this was done using gel electrophoresis with radiolabeled primers [21]. A sequencing ladder was used to map the location of the 3' end of the cDNAs in the (+) and (–) channels, and a comparison of the band intensities

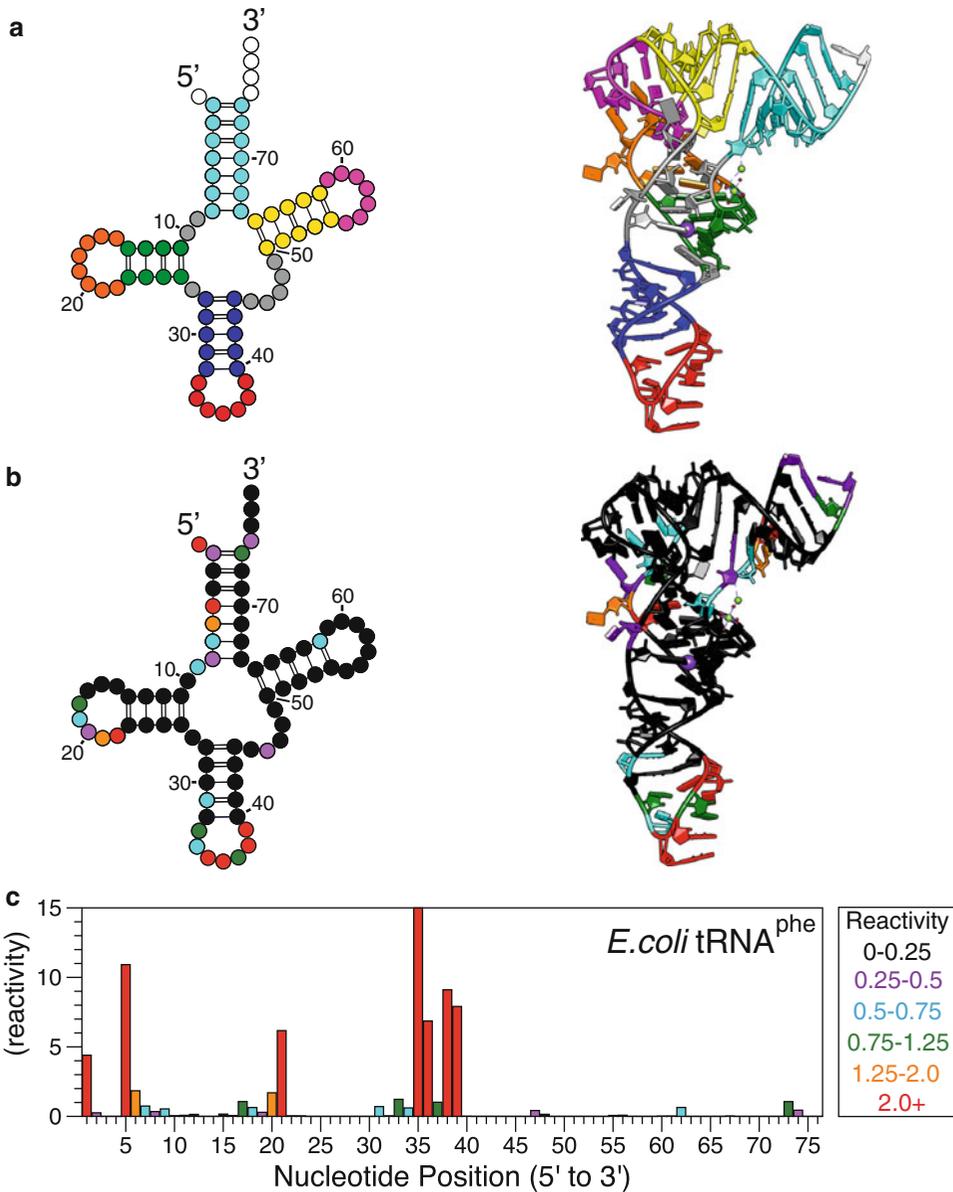


Fig. 1 Secondary and tertiary structures of *E. coli* tRNA^{phe}. **(a)** Secondary structure model of *E. coli* tRNA^{phe} (left) colored to show features of the tertiary structure of tRNA^{phe} (right; PDB: 3LOU) [11]. Note that the secondary structure image does not capture the tertiary interaction between the hairpin 1 (orange) and hairpin 3 (magenta) loops. **(b)** Secondary and tertiary structures of tRNA^{phe} colored to show SHAPE-Seq reactivity intensity (ρ) according to the bar chart in **(c)**. Also note that the nucleotides involved in the tertiary interaction between hairpins 1 and 3 have low reactivities even though they occur in loops. **(c)** Bar chart showing a representative reactivity spectrum from SHAPE-Seq v2.0 for the unmodified tRNA^{phe} from *E. coli*, colored to show degrees of reactivity intensity. Reactivity data from Loughrey et al. [12] (RMDB ID [13]: TRNAPH_1M7_0002)

between the channels provided the relative frequency of modification at each nucleotide [20]. Fluorescently labeled primers were introduced later as an alternate readout method, which improved throughput with the use of capillary electrophoresis (CE) [18] and simplified data analysis with the ShapeFinder software, which assists with peak intensity quantification and reactivity calculation [27]. However, both methods are limited in that they cannot be multiplexed and suffer from the noise associated with integrating analog signals to determine the abundance of each cDNA length. Ultimately, these problems were solved by coupling the chemical modification step to next-generation sequencing (NGS) to determine the modification and natural RT drop off positions, creating the SHAPE-Seq method [28, 29]. In SHAPE-Seq, and related techniques, cDNA sequences are bioinformatically aligned to determine the precise location of (+) and (-) channel cDNA ends, providing a “digital” read out of this important information.

NGS provides a number of advantages to mapping RNA structure with chemical probes. The digital output of raw counts of modification and RT drop off positions for each nucleotide allows for a more accurate determination of reactivities without having to integrate peak or band intensities as with gel or CE-based methods. This allows for convenient implementation of algorithms that can automatically correct for signal decay due to the unidirectional nature of the RT process [22, 23]. In addition, with NGS any incorrect sequences within the cDNA sequencing reads can be filtered out, such as those that occur from off-target priming during the RT step. Finally, the automation and multiplexing afforded by the increasingly powerful NGS platforms allows for higher throughput, increased accuracy, and the ability to perform experiments on a mixture of RNAs [12, 28].

Here we describe SHAPE-Seq v2.0, which combines the innovations of the original SHAPE-Seq (v1.0) protocol [28, 29] with a number of advances to increase the flexibility and accuracy of the SHAPE-Seq technique. In SHAPE-Seq v2.0, a ssDNA linker is ligated to the 3' end of the RNA after chemical modification to provide an RT priming site (Fig. 2) [12]. The introduced priming site is “universal”, requiring only one RT primer sequence for any SHAPE-Seq v2.0 experiment, regardless of the RNA(s) being studied. Universal priming also removes the requirement of older versions of SHAPE-Seq to contain extra RNA sequences, or structure cassettes, to act as an RT priming site, which had the potential to interfere with RNA folding (Fig. 2) [12, 28]. SHAPE-Seq v2.0 also uses internal barcodes to increase the multiplexing capabilities of SHAPE-Seq and allows more experiments to be sequenced within a single NGS lane [12]. Lastly, an updated set of parameters have been determined with SHAPE-Seq v2.0 for use with RNAstructure [25] to generate experimentally restrained secondary structure models of the probed RNAs [12].

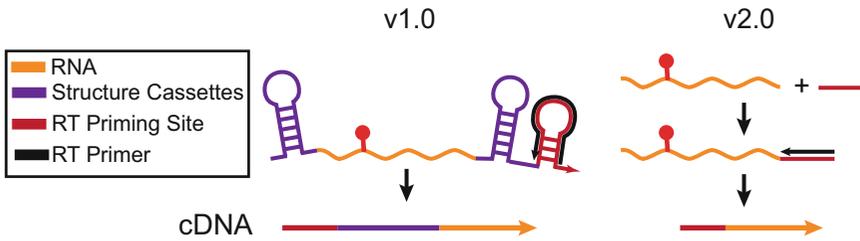


Fig. 2 Comparison of the original SHAPE/SHAPE-Seq vs. SHAPE-Seq v2.0 reverse transcription priming strategies. The structure cassettes are RNA hairpins added to both ends of an RNA of interest and are designed to fold independently of the desired RNA sequence under study. They were originally designed to mask signal noise that occurs at the 5' and 3' ends when capillary or gel electrophoresis is used to quantify the (+) and (-) channel cDNA distributions. The original SHAPE-Seq v1.0 strategy [28] used these structure cassette flanking sequences to provide a reverse transcription priming site. In SHAPE-Seq v2.0, a linker sequence is added to the RNA post-modification, which serves as a priming site for reverse transcription without the need for the structure cassettes [12]. Reproduced from Loughrey et al. 2014 by permission of Oxford University Press [12]

The SHAPE-Seq v2.0 protocol proceeds through the following steps: chemical modification with a SHAPE reagent, linker ligation, reverse transcription, ssDNA adapter ligation and PCR (for NGS library preparation), NGS, and data analysis (Fig. 3) [12]. After library preparation, NGS is performed using paired-end sequencing. Each paired-end sequence contains two reads. One read determines where the RT priming site was, while the other read indicates where the RT stop occurred (the modification site) within a given RNA sequence or set of RNA targets [29]. Data analysis is performed using Spats (<http://github.com/LucksLab/spats>), which separates sequencing reads into (+) and (-) channels before bioinformatically aligning the reads in each channel to the target sequence(s) to determine the (+) and (-) channel RT stop distributions. Using these distributions, Spats then applies a maximum likelihood estimation procedure to determine the reactivity values for each nucleotide in each RNA of the experiment. The reactivities are reported as θ values, where θ is the probability that a modification occurs at a particular nucleotide within a given RNA sequence [22, 23]. After normalization, the final result is a set of reactivities for each RNA analyzed that indicates the “flexibility” of each nucleotide. These reactivities can then be qualitatively interpreted or used to restrain secondary structure folding algorithms to generate experimentally guided RNA structure models [12, 25].

2 Materials

Prepare all solutions and buffers with RNase-free water. Necessary components and equipment are listed for each section of the protocol for clarity, thus, some components are repeated if needed

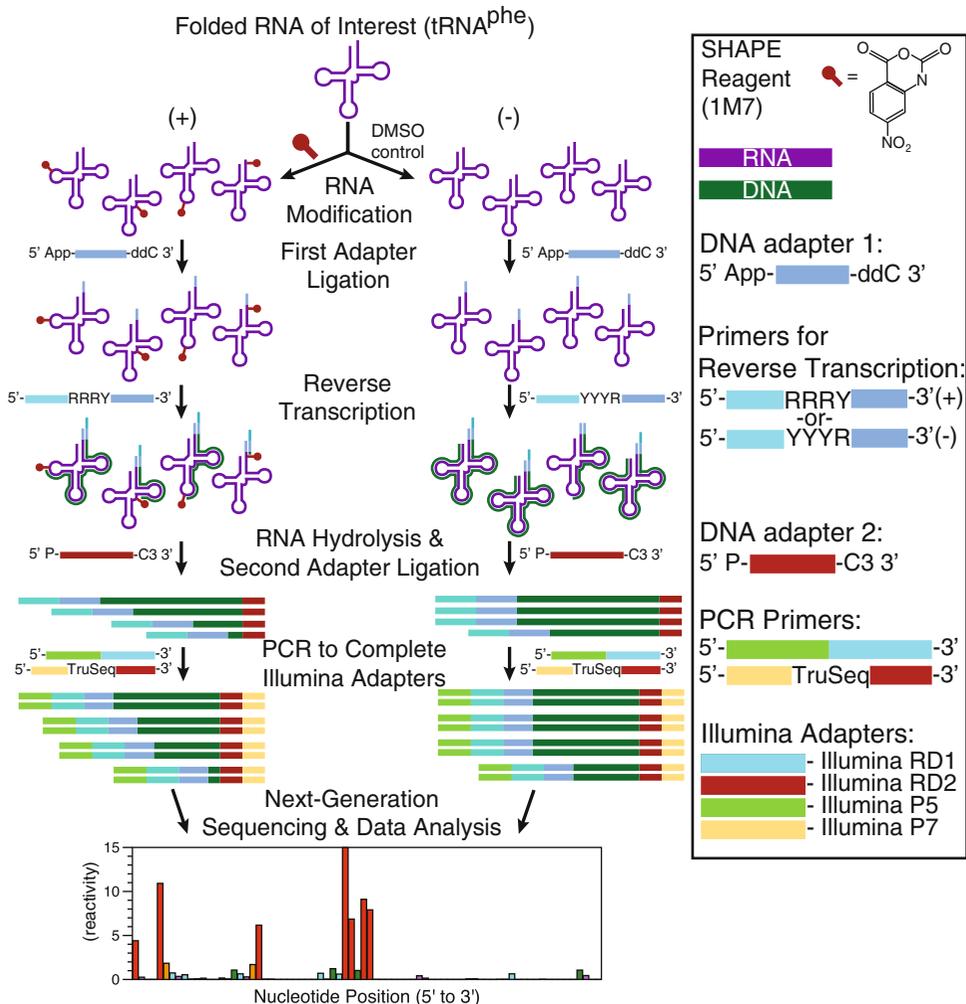


Fig. 3 SHAPE-Seq v2.0 method overview with tRNA^{phe} from *E. coli* as an example. In SHAPE-Seq v2.0, RNAs are modified with a SHAPE reagent (+), such as 1M7 [18], in a structure-dependent manner. Less structured nucleotides are more likely to be modified. DNA adapter 1 is then ligated to the RNA to provide a reverse transcription (RT) priming site. Since reverse transcriptase is blocked by SHAPE modifications, RT is used to determine the RNA modification positions by creating a pool of complementary DNAs (cDNAs) whose length distribution reflects the distribution of modification positions. Control reactions (-) are performed to account for the propensity of natural RT drop off from factors other than SHAPE modification. RT primer tails contain an indexing handle that distinguishes between the (+) and (-) channels and a portion of one of the required Illumina sequencing adapters (see Note 14). The other Illumina sequencing adapter is added to the 3' end of each cDNA through a single-stranded DNA ligation. A limited number of PCR cycles are used to both amplify the library and add the rest of the required adapters prior to sequencing. A freely available bioinformatic pipeline, Spats (<http://github.com/LucksLab/spats>), is then used to align sequencing reads, correct for biases due to RT-based signal decay, and calculate reactivity values

in multiple sections. Also, perform all steps in an RNase-free area. We suggest cleaning pipettes and surfaces with RNaseZap (Life Technologies), or a similar RNase-removing solution. This method assumes that a purified RNA (or set of RNAs) of interest has already been generated using *in vitro* transcription followed by gel purification, or similar methods.

2.1 RNA Modification Components

1. 20 pmol purified RNA of interest (*see Note 1*) with a free 2'-OH on the 3' end (*see Note 2*).
2. 1 M HEPES buffer, pH 8.0: Add 4.7 g HEPES to 20 mL H₂O, adjust pH to 8.0 with NaOH.
3. 5 M NaCl.
4. 1 M MgCl₂.
5. Anhydrous dimethyl sulfoxide (DMSO).
6. 3.33× folding buffer: 333 mM HEPES, 333 mM NaCl, 33 mM MgCl₂, pH 8.0. Combine 333 μL 1 M HEPES, 66.6 μL 5 M NaCl, 33 μL 1 M MgCl₂, and 576.4 μL H₂O (*see Note 3*).
7. 1-methyl-7-nitroisatoic anhydride (1M7) (*see Note 4*): Make a 65 mM solution by weighing out 1 mg of 1M7 and dissolving in 69.3 μL anhydrous DMSO (*see Note 5*).
8. Thermal cycler.
9. 0.5 mL microcentrifuge tubes or thin-walled PCR tubes (*see Note 6*).

2.2 RNA Ligation Components

1. SuperaseIN RNase Inhibitor (Life Technologies).
2. T4 RNA ligase 2, truncated KQ (New England BioLabs).
3. 10× T4 RNA ligase buffer (*see Note 7*): 50 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, pH 7.5. To 600 μL H₂O, add 7.88 mg Tris-HCl, 10 μL of 1 M MgCl₂ stock solution, and 10 μL of 100 mM DTT solution. Adjust to pH 7.5 with NaOH and bring to 1 mL with H₂O. Store at -20 °C.
4. 50% PEG 8000 solution (*see Note 7*): Slowly dissolve 1 g of PEG 8000 in 1 mL RNase-free H₂O. May require heat. Store at -20 °C.
5. IDT miRNA cloning linker 2 (DNA adapter 1): /5rApp/CACTCGGGCACCAAGGA/3ddC/. /5rApp/ indicates an adenylation modification and /3ddC/ is a dideoxycytidine modification. Rehydrate to a final concentration of 20 μM (Integrated DNA Technologies).
6. Ethanol (EtOH), absolute.
7. Glycogen, 20 mg/mL.
8. 3 M NaOAc, pH 5.5.

9. 70% EtOH solution: Dilute 7 mL absolute EtOH by adding 3 mL RNase-free H₂O.
10. Refrigerated microcentrifuge.

2.3 Reverse Transcription (RT) Components

1. RT primers: Both sequences are DNA oligonucleotides. Neither requires any special purification. The (+) sample RT primer sequence is: 5'-CTTTCCTACACGACGCTCTTCGATCTRRRY_{xx}TCCTTGGTGCCCGAGTG-3' and the (-) sample RT primer sequence is: CTTTCCTACACGACGCTCTTCGATCTYYR_{xx}TCCTTGGTGCCCGAGTG (*see Note 8*). The “xx” sequences can be replaced with internal barcodes or deleted if desired (*see Note 9*). Make dilutions of 0.5 μM for both (*see Note 1*).
2. 10 mM deoxyribonucleotide triphosphates (dNTPs): Solution of 10 mM each ATP, CTP, GTP, and TTP.
3. SuperScript III (SSIII) reverse transcriptase (Life Technologies).
4. SSIII storage buffer: 20 mM Tris-HCl, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 0.01% Triton, 50% glycerol, pH 7.5. To 3 mL of H₂O, add 31.5 mg Tris-HCl, 58.4 mg NaCl, 1.54 mg DTT, 2.92 mg EDTA, 10 μL 1% Triton X-100, and 5 mL glycerol. Adjust to pH 7.5 with NaOH. Add H₂O to 10 mL. Store at -20 °C.
5. 100 mM dithiothreitol (DTT) (*see Note 10*). Store at -20 °C.
6. 5× SSIII first strand buffer (*see Note 10*): 250 mM Tris-HCl, 375 mM KCl, 15 mM MgCl₂, pH 8.3. In 500 μL H₂O, dissolve 39.4 mg Tris-HCl, 28.0 mg KCl, and 1.43 mg MgCl₂. Adjust to pH 8.3 with NaOH. Add H₂O to 1 mL final volume. Store at -20 °C.
7. RT master mix: Mix 4 volumes of 5× SSIII First Strand Buffer, 1 volume of 100 mM DTT, and 1 volume of 10 mM dNTPs. Store at -20 °C, 100 μL aliquots recommended.
8. 0.5× SSIII: Combine equal parts SSIII storage buffer and SSIII reverse transcriptase. Mix well and store at -20 °C until needed.
9. 4 M NaOH solution.
10. 1 M HCl solution.
11. Thermal cycler.
12. Ethanol, absolute.
13. Glycogen, 20 mg/mL.
14. 3 M NaOAc, pH 5.5.
15. 70% EtOH solution: Dilute 7 mL absolute EtOH by adding 3 mL RNase-free H₂O.
16. Refrigerated microcentrifuge.

2.4 ssDNA Ligation and Quality Analysis (QA) Components

1. DNA adapter 2: PAGE-purified DNA oligonucleotide with sequence 5'-/5Phos/AGATCGGAAGAGCACACGTCTGACTCCAGTCAC/3SpC3/-3'. /5Phos/ indicates a phosphate modification and /3SpC3/ is a three-carbon spacer modification. Dilute to 100 μ M for use.
2. CircLigase I ssDNA ligase (Epicentre).
3. 10 \times CircLigase I reaction buffer (*see Note 11*): 0.5 M MOPS, 0.1 M KCl, 50 mM MgCl₂, and 10 mM DTT, pH 7.5. In 500 μ L H₂O, dissolve 104.6 mg MOPS salts, 7.46 mg KCl, 4.76 mg MgCl₂, and 1.54 mg DTT. Adjust pH to 7.5 with NaOH. Add H₂O to a final volume of 1 mL. Store at -20 $^{\circ}$ C.
4. 1 mM adenosine triphosphate solution (ATP) (*see Note 11*).
5. 50 mM MnCl₂ solution (*see Note 11*).
6. Glycogen, 20 mg/mL.
7. 3 M NaOAc, pH 5.5.
8. Ethanol, absolute.
9. Refrigerated microcentrifuge.
10. Agencourt AMPure XP beads (Beckman Coulter).
11. Magnetic stand for 96-well plate or microcentrifuge tubes.
12. TE Buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 7.5. In 5 mL H₂O, dissolve 15.8 mg Tris-HCl and 2.92 mg EDTA. Adjust pH to 7.5 with NaOH, then add H₂O to a final volume of 10 mL.
13. Phusion DNA polymerase, supplied with 5 \times reaction buffer (New England BioLabs).
14. 10 mM dNTPs: Solution of 10 mM each ATP, CTP, GTP, and TTP.
15. 1.5 mL microcentrifuge tubes.
16. QA_R primers: Two fluorescently labeled DNA oligonucleotides are used. No special purification is necessary. The (+) primer is 5'-VIC-GTGACTGGAGTTCAGACGTGTGCTC-3' and the (-) primer is 5' NED-GTGACTGGAGTTCAGACGTGTGCTC-3' (Life Technologies) (*see Note 12*).
17. QA_F: DNA oligonucleotide with sequence 5'-CCCTACACGACGCTCTTCCGATC-3'. No purification is necessary. Make a 1 μ M dilution.
18. GeneScan 500 LIZ standard (Applied Biosystems).
19. Deionized formamide.
20. ABI 3730xl DNA Analyzer, or similar DNA analyzer (Applied Biosystems).
21. ShapeFinder software [27] or other program for viewing capillary electrophoresis traces.
22. (Optional) BioAnalyzer and dsDNA high sensitivity chips (*see Note 13*) (Applied Biosystems).

2.5 Library Construction Components

1. Thin-walled PCR tubes or 96-well reaction plate.
2. Thermal cycler.
3. Phusion DNA polymerase, supplied with 5× reaction buffer.
4. 10 mM dNTPs: Solution of 10 mM each ATP, CTP, GTP, and TTP.
5. PE_F: HPLC-purified DNA oligonucleotide with sequence 5'-AATGATACGGCGACCAC CGAGATCTACACTCTTTC CCTACACGACGCTCTTCCGATCT-3'. Dilute to 100 μM. (Oligonucleotide sequence © 2007–2013 Illumina, Inc. All rights reserved.)
6. Indexing PCR primers: PAGE-purified DNA oligonucleotide with sequences 5'-CAAGCAGAAGACGGCATAACGAGAT **XXXXXX**GTGACTGGAGTTCAGACGTGTGCTC-3'. Replace "XXXXXX" with indexes for Illumina sequencing, as shown in Table 1 for indexes 1–6. Dilute to 100 μM for use (*see Note 14*). (Oligonucleotide sequences © 2007–2013 Illumina, Inc. All rights reserved.)
7. Exonuclease I (ExoI).
8. Agencourt AMPure XP beads.
9. Magnetic stand for 96-well plate or microcentrifuge tubes.
10. TE Buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 7.5. In 5 mL H₂O, dissolve 15.8 mg Tris-HCl and 2.92 mg EDTA. Adjust pH to 7.5 with NaOH, then add H₂O to a final volume of 10 mL.
11. Qubit Fluorometer High Sensitivity dsDNA kit (recommended; Life Technologies) or NanoDrop.

Table 1
Example oligonucleotides for Illumina barcoding, indexes 1–6

Illumina Index #1	CAAGCAGAAGACGGCATAACGAGATCGTGATGTGACT GGAGTTCAGACGTGTGCTC
Illumina Index #2	CAAGCAGAAGACGGCATAACGAGATACATCGGTGAC TGGAGTTCAGACGTGTGCTC
Illumina Index #3	CAAGCAGAAGACGGCATAACGAGATGCCTAAGTGACT GGAGTTCAGACGTGTGCTC
Illumina Index #4	CAAGCAGAAGACGGCATAACGAGATTGGTCAGTGACT GGAGTTCAGACGTGTGCTC
Illumina Index #5	CAAGCAGAAGACGGCATAACGAGATCACTGTGTGACT GGAGTTCAGACGTGTGCTC
Illumina Index #6	CAAGCAGAAGACGGCATAACGAGATATTGGCGTGACT GGAGTTCAGACGTGTGCTC

Sequences are reverse complements of TrueSeq adapter sequences (Oligonucleotide sequences © 2007–2013 Illumina, Inc. All rights reserved.)

2.6 NGS and Data Analysis

1. Illumina Sequencing Platform (MiSeq or HiSeq).
2. Unix, Linux, or Mac OS X equipped system.
3. Spats, v1.0.0, installed (*see Note 15*).

3 Methods

3.1 RNA Modification

1. Dilute 20 pmol of RNA in 12 μL RNase-free H_2O in a 0.5 mL microcentrifuge tube or thin-walled PCR tube.
2. Prepare fresh 65 mM 1 M7 solution. To a fresh tube, add 1 μL of recently prepared 65 mM 1 M7 and mark as “+”. To another fresh tube, add 1 μL of anhydrous DMSO (*see Note 16*) and mark as “-”. Set both tubes aside until **step 6** (*see Note 17*).
3. Incubate the tube from **step 1** at 95 °C for 2 min to denature the RNA (*see Note 18*).
4. Snap cool the RNA by incubating on ice for 1 min, then add 6 μL of 3.3 \times folding buffer and mix well. There should now be 18 μL of diluted RNA.
5. Incubate the RNA at 37 °C for 20 min (*see Note 19*).
6. Add 9 μL of the RNA solution to the “-” tube and the other 9 μL to the “+” tube and mix well. Incubate the “+” and “-” tubes for 2 min (*see Note 4*) at 37 °C (*see Note 20*), then transfer to ice (*see Note 21*). Discard the now empty tube containing the original RNA dilution.

3.2 RNA Ligation

1. Warm 50% PEG 8000 to room temperature to ease pipetting (*see Note 22*).
2. Add the following reagents in order to each “+” and “-” tube of RNA from Subheading 3.1, **step 6** according to Table 2 (*see Note 23*).
3. Vortex to mix well and incubate overnight (≥ 8 h) at room temperature.
4. If using PCR tubes, transfer to 0.5 mL microcentrifuge tubes. Add 130 μL RNase-free H_2O , 1 μL 20 mg/mL glycogen, 15 μL 3 M NaOAc solution, and 450 μL of ice-cold absolute EtOH to both tubes and mix well.
5. Incubate at -80 °C for 30 min, then spin on a chilled microcentrifuge at maximum speed for 30 min to pellet the RNA. This step can be paused by leaving the precipitations in the freezer until ready to continue.
6. Aspirate the supernatant (*see Note 24*), then add 200 μL of 70% EtOH and gently shake the tube to wash the pellet. Respin briefly (30 s to 2 min) and aspirate the EtOH wash. Repeat the brief spin and remove any remaining traces of EtOH (*see Note 25*), then dissolve each pellet in 10 μL of RNase-free H_2O .

Table 2
DNA adapter 1 ligation reaction components

10 μ L	Modified RNA from Subheading 3.1, step 6
0.5 μ L	SupraseIn RNase inhibitor
6 μ L	50% PEG 8000
2 μ L	10 \times T4 RNA ligase buffer
1 μ L	IDT miRNA cloning linker 2 (DNA adapter 1; 20 μ M)
0.5 μ L	T4 RNA ligase, truncated KQ (200 U/ μ L)
20 μ L	Total volume

3.3 Reverse Transcription

1. Add 3 μ L of 0.5 μ M (+) RT primer to the “+” tube and 3 μ L of 0.5 μ M (–) RT primer to the “–” tube (*see Note 26*).
2. Incubate both tubes at 95 $^{\circ}$ C for 2 min, then 65 $^{\circ}$ C for 5 min to denature the RNA. During this step, mix the SSIII master mix by combining 12 μ L of RT master mix with 2 μ L 0.5 \times SSIII and leave on ice until the next step.
3. Snap cool the “–” and “+” tubes by incubating on ice for 30 s, then add 7 μ L of SSIII master mix to each tube, mixing well and being sure to knock down any condensation from the tube walls.
4. Incubate at 45 $^{\circ}$ C for 1 min, then incubate at 52 $^{\circ}$ C for 25 min.
5. Incubate at 65 $^{\circ}$ C for 5 min.
6. Add 1 μ L of 4 M NaOH solution to each tube and incubate at 95 $^{\circ}$ C for 5 min to hydrolyze the RNA.
7. Allow the tubes to cool to room temperature (*see Note 27*), then add 2 μ L of 1 M HCl solution to partially neutralize the base (*see Note 28*).
8. Add 69 μ L of ice-cold absolute EtOH and incubate at –80 $^{\circ}$ C for 15 min, followed by centrifugation at 15 min at maximum speed in a chilled microcentrifuge. This will pellet the cDNA. This step can be paused by leaving the precipitations in the freezer until ready to continue.
9. Aspirate the supernatant (*see Note 24*), then add 500 μ L of 70% EtOH and gently shake the tube to wash the pellet (*see Note 29*). Respin briefly (30 s to 2 min) and aspirate the EtOH wash. Repeat the brief spin and remove any remaining traces of EtOH, then dissolve each pellet in 22.5 μ L of RNase-free H₂O.

3.4 ssDNA Ligation and Quality Analysis

1. Add the reagents from Table 3 in order to each “+” and “-” tube of RNA from Subheading 3.3, step 9 and mix well.
2. Incubate at 60 °C for 2 h and then 80 °C for 10 min to heat deactivate CircLigase I.
3. Add 70 μL of H_2O , 1 μL of 20 mg/mL glycogen, 10 μL 3 M NaOAc, and 300 μL ice-cold absolute EtOH to each tube.
4. Incubate at -80 °C for 30 min, then spin on a chilled microcentrifuge at maximum speed for 30 min to pellet the DNA (*see Note 30*). This step can be paused by leaving the precipitations in the freezer until ready to continue.
5. Aspirate the supernatant (*see Note 24*), then respin briefly (30 s to 2 min) and aspirate any remaining traces of EtOH. Dissolve each pellet in 20 μL of nuclease-free H_2O .
6. Purify the ssDNA libraries with Agencourt AMPure XP beads (*see Note 31*). To do this, make sure the beads are well suspended by swirling the bottle, then add 36 μL of the beads to each “+” and “-” tube and mix well. Incubate at room temperature for 5 min. Carefully place the tubes on a magnetic stand and let the beads separate for 2 min. Use 5 min if using a 96-well format (*see Note 32*). Aspirate the supernatant. While the tubes are still on the stand, gently wash with 200 μL of 70% EtOH. Incubate for 30 s before aspirating and discarding the EtOH (*see Note 33*). Repeat the wash and aspiration, then let the beads air dry for 3 min (*see Note 34*). Remove tubes from the magnetic stand and resuspend the beads in 20 μL of TE buffer. Then, place the tubes/plate back on the magnetic stand and let the beads separate for 1 min. Transfer the supernatant, which contains the ssDNA libraries, to fresh, labeled tubes. Store libraries at -20 °C when not in use.
7. To perform quality analysis two PCRs will be set up, one for the “+” sample and one for the “-” sample (*see Note 35*). To do

Table 3
DNA adapter 2 ligation reaction components

22.5 μL	cDNA from Subheading 3.3, step 9
3 μL	10 \times CircLigase I ligation buffer
1.5 μL	MnCl ₂ solution (50 mM)
1.5 μL	ATP solution (1 mM)
0.5 μL	DNA adapter 2 (100 μM)
1 μL	CircLigase I ssDNA ligase (100 U/ μL)
30 μL	Total

this, first set up the “+” reaction by adding the following reagents from Table 4 in order and mixing well. Similarly, set up a “-” reaction by replacing the (+) QA_R primer with the (-) QA_R primer and using the “-” ssDNA library instead of the “+” library. Mix well.

8. Run the following thermal cycler protocol: 98 °C for 30 s, 98 °C for 10 s, 65 °C for 30 s, 72 °C for 30 s, repeat the previous three steps 11 times, 72 °C for 5 min.
9. Combine the “+” and “-” samples in a 1.5 mL microcentrifuge tube and add 50 μ L H₂O, 10 μ L 3 M NaOAc, and 300 μ L ice-cold absolute EtOH.
10. Incubate at -80 °C for 15 min, then spin on a chilled microcentrifuge at maximum speed for 15 min to pellet the DNA (*see Note 36*).
11. Aspirate the supernatant (*see Note 24*), then respin briefly (30 s to 2 min) and aspirate any remaining traces of EtOH. Dissolve each pellet in 10 μ L of deionized formamide, incubating at 95 °C to aid dissolution.
12. Add 0.25 μ L GeneScan 500 LIZ standard to the dissolved QA DNA and run on ABI 3730xl DNA Analyzer (or similar capillary electrophoresis machine).
13. Open the resulting .fsa files generated by the DNA Analyzer with ShapeFinder [27] (*see Note 37*) and use the LIZ standard to identify peak lengths (*see Note 38*). There should be a full-length peak clearly visible at a position equal to the number of nucleotides of the RNA+80 nt (for the Illumina adapters) (Fig. 4). Peaks at 80 nt are indicative of side product dimers that form between unextended RT primer and DNA adapter 2 and are amplified by PCR. A good library trace should show

Table 4
Reaction components for generating QA libraries

14.75 μ L	Nuclease-free H ₂ O
5 μ L	5 \times Phusion buffer
0.5 μ L	dNTPs solution (10 mM)
1.5 μ L	QA_F primer (1 μ M)
1.5 μ L	(+) QA_R primer (1 μ M)
1.5 μ L	“+” ssDNA library from Subheading 3.4, step 6
0.25 μ L	Phusion DNA polymerase (2 U/ μ L)
25 μ L	Total

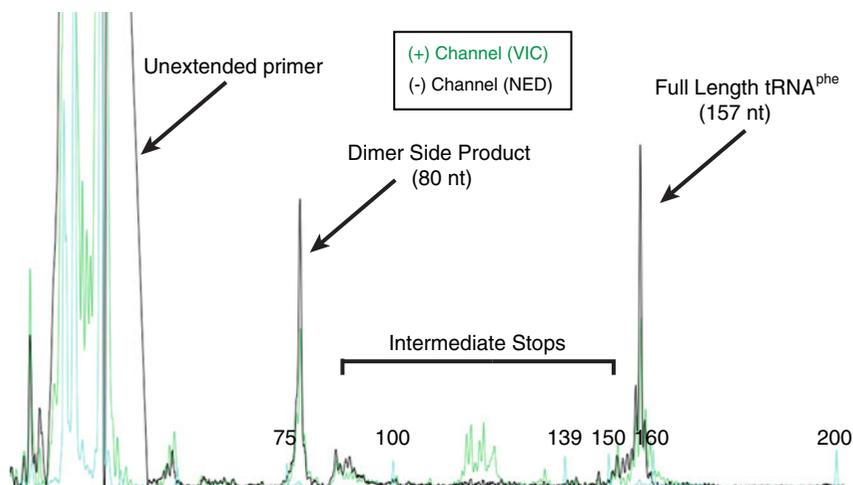


Fig. 4 Quality analysis (QA) example with an *E. coli* tRNA^{Phe} dsDNA library viewed with ShapeFinder [27]. The “+” and “-” samples are visualized with the VIC (green) and NED (black) fluorophores, respectively. Unextended reverse transcription primer (Subheading 3.3) can ligate to DNA adapter 2 (Subheading 3.4) and create an unwanted dimer side product 80 nucleotides (nt) long after PCR amplification. The fully extended cDNA is observed as a large peak at 157 nt for the RNA length (77 nt for tRNA^{Phe}) plus 80 nt for the partially added Illumina adapter sequences. Intermediate stops between 80 nt and 157 nt suggest either premature reverse transcriptase drop off or modified positions (“+” sample only). Unextended, fluorescently labeled QA_R primers from the QA library construction step are visible as short fragments on the *left* as indicated. This QA trace falls within our guidelines for a “good” library and was sequenced to produce the data shown in Figs. 1 and 3. Peaks for the LIZ standard (*cyan*) are annotated according to their size distribution (*see* Note 38)

considerable peak height and a good full length–side product ratio, with minor peaks between the full length and side product peaks. The minor peaks where the (+) channel exceeds the (-) channel are indicative of RT stops due to SHAPE modifications (*see* Note 39). If both the (+) and (-) channels look good, continue to Subheading 3.5, otherwise this procedure should be repeated.

3.5 Library Construction

1. To build the SHAPE-Seq libraries that will be sequenced on the Illumina platform, mix two 50 μ L PCRs, one for the “+” sample and one for the “-” sample. Combine the reagents from Table 5, in order, for the “+” and “-” samples separately using the same Illumina index for both. Mix well.
2. Run the following thermal cycler protocol: 98 $^{\circ}$ C for 30 s, 98 $^{\circ}$ C for 10 s, 65 $^{\circ}$ C for 30 s, 72 $^{\circ}$ C for 30 s, repeat the previous three steps eight times, 72 $^{\circ}$ C for 5 min (*see* Note 40).
3. Incubate at 4 $^{\circ}$ C for 5 min (*see* Note 41), then add 5 U of Exonuclease I to each reaction and incubate at 37 $^{\circ}$ C for 30 min. This will remove excess primer from the libraries (*see* Note 42).

Table 5
Reaction components to create libraries for Illumina sequencing

35.5 μL	Nuclease-free H_2O
10 μL	5 \times Phusion buffer
0.5 μL	dNTPs solution (10 mM)
0.25 μL	PE_F (100 μM)
0.25 μL	Illumina Index Primer (100 μM)
3 μL	ssDNA library
0.5 μL	Phusion DNA polymerase (2 U/ μL)
50 μL	Total

4. Bead purify the dsDNA libraries according to Subheading 3.4, **step 6**, except use 90 μL beads instead of 36 μL . Elute in 20 μL of TE buffer (*see Note 43*).
5. Use the Qubit dsDNA High Sensitivity kit or NanoDrop to measure the mass concentration of each library. Any measurable quantity can be sequenced, but a minimal target concentration of 2 ng/ μL is sought after (*see Note 44*).

3.6 NGS and Data Analysis

1. Determine which SHAPE-Seq dsDNA libraries are to be sequenced (*see Note 45*). Then, use the quality analysis traces from Subheading 3.4, **step 13** to determine the average lengths of the “+” and “-” dsDNA libraries (*see Note 46*). Use the formula
$$\frac{1000 \times x}{(L + 64) \times 607.4 + 157.9}$$
 to calculate the average micromolar concentration (μM) of each library, where L is the average length of the dsDNA library and x is the mass concentration (ng/ μL) measured in Subheading 3.5, **step 5** (*see Note 47*). Do this for both the “+” and “-” dsDNA libraries separately, and for all dsDNA libraries being sequenced together (multiplexed).
2. Use the molar concentrations determined in **step 1** to balance all of the dsDNA libraries being sequenced such that they all have the same fmol of library (typically 10–30 fmol) in the final mixture, which should have a total final volume between 10 and 30 μL . Pipette all of the libraries together and measure the mass concentration of the final mixture to be sequenced with the Qubit dsDNA High Sensitivity kit or NanoDrop.
3. Sequence the mixture on the Illumina platform, using 2×35 bp paired end reads. The concentration of the mixture and the average length will be required to properly dilute the sample for sequencing. The MiSeq works well for 20 library pairs (“+”

- and “-”) or fewer, while the HiSeq is better for a larger number of SHAPE-Seq libraries (*see Note 48*). Both platforms will require at least 5% PhiX DNA included as a control (*see Note 49*). Reads longer than 35 bp are not necessary (*see Note 50*).
4. Obtain the sequencing data from a local storage drive, sequencing facility, or BaseSpace (www.basespace.com) on a Linux, Unix, or Mac OS X capable computer and unzip the files to extract the *.fastq.gz files. There should be two files present for each Illumina index sequenced, one for read 1 (R1) and one for read 2 (R2) of the paired end reads. Then, use gunzip to decompress these files to the *.fastq format (*see Note 51*).
 5. Create a fasta (.fa) formatted targets file Include the linker sequence and internal barcodes (if present) (*see Note 52*). For each *.fastq pair (R1 and R2), there should only be one *.fa file, that contains all of the target RNAs within that library.
 6. Run `adapter_trimmer.py` with the following command: `adapter_trimmer.py <R1_seq.fastq> <R2_seq.fastq> <targets.fa>`, where `<R1_seq.fastq>` and `<R2_seq.fastq>` are the Illumina *.fastq files for R1 and R2, respectively, and `<targets.fa>` is the fasta-formatted targets file created in **step 5** (*see Note 53*). By default, the script will use the adapter sequences and read lengths described above. However, some values must be set using the option flags if differing from the default (*see Note 54*). The output of `adapter_trimmer.py` is two files: `combined_R1.fastq` and `combined_R2.fastq`, which have been processed to remove any Illumina adapter sequences, if present (*see Note 55*).
 7. Feed the output of `adapter_trimmer.py` to Spats using the following command: `spats <targets.fa> RRRY and YYYYR combined_R1.fastq combined_R2.fastq`, where `<targets.fa>` is the targets file created in **step 5** (*see Note 56*). The RRRY and YYYYR inputs indicate that the “+” sample has the “RRRY” handle and the “-” sample has the “YYYYR” handle (*see Note 8*). The output of Spats is a directory containing the split raw reads (+/-; *.fq), raw alignments (*.sam), and calculated reactivities (reactivities.out). Note that the calculated reactivities for all of the RNAs are concatenated together in the reactivities.out text file (*see Note 57*). Reactivities are reported as θ values for each nucleotide in a column labeled “theta”.
 8. Normalize the output θ_i values to ρ_i values by multiplying all of the θ_i values by one less than the original RNA length. Do not include the linker or adapter sequences in this length. These ρ reactivities are the final output SHAPE-Seq v2.0, where values between 0 and 0.5 are considered “low,” 0.5–1.25 are “moderate,” and >1.25 are “high” (*see Note 58*). These ρ_i values

can be plotted to obtain reactivity maps (Fig. 1c) or used as restraints for secondary structure prediction with RNAstructure [25] (*see Note 59*), using an m -value of 1.1 and a b -value of -0.3 [12] (*see Note 60*).

4 Notes

1. Less than 20 pmol of RNA can be used in this protocol. However, it has been optimized for 20 pmol and if less is used, more side product will be present later and will cut back on the number of usable sequencing reads. A lower concentration of RT primer can be used to partially alleviate this problem.
2. A modification of the 3' terminal 2'-OH on the RNA of interest will prevent the RNA–DNA ligation in Subheading 3.2 from working properly. If the RNA of interest is purified directly from an *in vitro* transcription using T7 RNA polymerase, there are likely to be 1–3 non-templated nucleotides added to the 3' end. The additional 3' bases will cause an overhang mismatch during the sequence alignment step with Bowtie, which can reduce the number of usable sequencing reads. There are a few ways this problem can be dealt with: either use a ribozyme that cleaves off the 3' end, such as the hepatitis δ ribozyme (with end healing to regenerate the 3'-OH) [12, 30], introduce a methoxy-modification near the 3' end of the DNA template to reduce non-templated base addition [31], or add the extra non-templated bases to the targets file when performing the sequence alignment with Bowtie and assume the extra bases do not affect the RNA structure of interest.
3. Other folding buffers can be used, but should be prepared at a 3.33 \times concentration.
4. 1M7 is not commercially available and must be synthesized. There are two protocols currently available to do so [29, 32]. If an alternate is desired, one commercially available option is *N*-methylisatoic anhydride (NMIA; Life Technologies) [21]. However, if using NMIA, *increase* the modification time in Subheading 3.1, **step 6** from 1 to 45 min. Similarly, if benzoyl cyanide [33] is used, *reduce* the modification time from 1 min to <5 s.
5. The 1M7 solution should be prepared fresh whenever possible. However, 1M7 will retain most of its reactivity for approximately 5–6 days when stored properly in a low-moisture environment in a well-sealed tube. Degraded 1M7 will have undergone the same color change as if it was hydrolyzed as mentioned in **Note 21**.
6. Using a thermal cycler with the capability to handle 0.5 mL microcentrifuge tubes is the easiest method and will not require

tube transfers. If using thin-walled PCR tubes, you will need to transfer to a 0.5 mL microcentrifuge tube to perform EtOH precipitations and transfer back to a new PCR tube.

7. 10× T4 RNA ligase buffer and 50% PEG 8000 solution are supplied with T4 RNA Ligase, truncated KQ when purchased from New England BioLabs.
8. The basic composition of the primer is as follows: (1) CTTTCCCTACACGACGCTCTTCCGATCT is the 3' half of one Illumina adapter required for sequencing, the other half is added with PCR during library construction in Subheading 3.5. (Oligonucleotide sequence © 2007–2013 Illumina, Inc. All Rights Reserved.) (2) RRRY/YYYY (R for purine, Y or pyrimidine) is a degenerate pool of indexing handles used to distinguish the “+” and “-” samples during reactivity calculation. If the libraries are properly balanced during sequencing, the first four reads will also provide randomness to help calibrate the Illumina sequencing platform. (3) “xx” is a placeholder for an optional internal barcode (*see Note 9*). (4) TCCTTGGTGCCCGAGTG is the reverse complement to DNA adapter 1, providing an RT priming site. *See Loughrey et al. Fig. S4 for more detailed information [12].*
9. For large numbers of experiments where fewer Illumina indexes are being used (to reduce the cost of buying many expensive oligonucleotides) or a lot of barcoding options are required, an internal barcode can be included at the “xx” position. These barcodes, if used with different lengths, can also be used to shift the sequencing reading frame to improve randomness when sequencing the DNA adapter 1 sequence common to all of the experiments. For example, adding “aa” as a barcode and “c” as another, two experiments containing the same RNA sequence can be distinguished by uniquely aligning to different targets with Bowtie, one that has the “aa” present and the other that has “c” present. Note that if any barcodes are used, they need to be able to uniquely align with Bowtie. Thus, using “aa” and “a” as two different barcodes will not work, as the “a” barcode would still align to the “aa” barcode sequence. However, if “a” were replaced with any other base it would uniquely align.
10. 100 mM DTT and 5× SSIII first strand buffer are both supplied in excess with SuperScript III when purchased from Life Technologies.
11. 10× CircLigase I reaction buffer, 1 mM ATP, and 50 mM MnCl₂ are supplied in excess with CircLigase I when purchased from Epicentre.
12. Because of the proprietary nature of the VIC and NED fluorophores, they can only be ordered from Life Technologies.

13. If using the Applied Biosystems BioAnalyzer to perform library quality analysis instead of capillary electrophoresis with fluorescently labeled DNA, as outlined in Subheading 3.4, you will need the high sensitivity DNA chips to properly analyze the dsDNA libraries (*see Note 46*). Note that the resolution is lower than with the capillary electrophoresis method. Also, you will not need to acquire **items 16–21** in Subheading 2.4.
14. The Illumina TruSeq indexing system has a number of pre-designed sequences as shown in Table 1. For additional indexing sequences beyond indexes 1–6, replace the “XXXXXX” with the six nucleotide indexing sequence as provided by Illumina for the TruSeq system.
15. First, the FastX toolkit and LibGTextUtils (http://hannonlab.cshl.edu/fastx_toolkit/download.html) need to be installed. The Spats pipeline also requires the alignment program Bowtie (<http://sourceforge.net/projects/bowtie-bio/files/bowtie/0.12.8/>) and the Boost software libraries (www.boost.org). The `adapter_trimmer.py` script mentioned in Subheading 3.6 is included with the Spats source code. Detailed installation instructions for Spats can be found at <http://luckslab.github.io/spats/installation.html>.
16. Use a syringe to puncture the septum of the anhydrous DMSO bottle and measure the volume of DMSO to add. It is not a problem if the DMSO/1M7 solution is briefly exposed to air afterward.
17. If using thin-walled PCR tubes and need to perform tube transfers at later steps for EtOH precipitations, etc. continue to label new tubes as “+” and “-”. Always use a fresh tube when transferring.
18. If incubating at 95 °C would irreversibly denature or destroy an RNA-containing complex and/or the RNA of interest is already folded in a different buffer, skip to Subheading 3.1, **step 5**. An alternate RNA folding protocol could be inserted here instead. If so, perform the alternative folding method and then proceed to Subheading 3.1, **step 5**.
19. If the relevant temperature and folding time for analyzing the RNA of interest is not 37 °C, this step can be easily altered. However, we suggest incubating for at least a few minutes at a constant temperature to allow the temperature to come to equilibrium before modification.
20. If using a folding temperature other than 37 °C, perform the modification at the same temperature. Note that the reaction rate of 1M7 is dependent on temperature, requiring more incubation time if using a lower temperature.
21. The 1M7 reaction is self-quenching in water, and will turn from yellow to orange when the reaction is complete.

22. Pipette the PEG solution slowly and carefully, it is very viscous and it is easy to add the wrong volume by accident. However, a slight inaccuracy in the PEG volume will not likely negatively impact the ligation reaction.
23. If a different folding buffer other than the one described in Subheading 2.1 was used, EtOH precipitate the modified RNAs from Subheading 3.1, **step 6** by adding 90 μL RNase-free H_2O , 10 μL 3 M NaOAc (pH 5.5), 1 μL 20 mg/mL glycogen, and 300 μL of ice-cold absolute EtOH and store at -80°C for 30 min. Then spin on a chilled microcentrifuge for 30 min at maximum speed and aspirate all of the EtOH, making sure that no EtOH remains. Dissolve the pellet in 10 μL of 10% DMSO in H_2O before mixing the ligation reaction in Subheading 3.2, **step 2**.
24. When aspirating large volumes of EtOH, doubling pipette tips can reduce the chance of disrupting the pellet. To do this, start with a larger volume pipette (such as 1 mL) and attach a tip. Then, plunge that tip into a low-volume tip with a fine point (such as a 10 μL tip). That way you can use the large pipette to handle the volume of EtOH present, but the fine tip to carefully withdraw EtOH from around the small pellet.
25. The glycogen should clearly distinguish the pellet, which frequently looks like a white streak at the bottom of the tube.
26. If using internal barcodes (*see Note 9*), add the appropriate matching barcode primer set. It is required that both the “+” and “-” samples have the same internal barcode.
27. Be careful when handling the tubes. If they are not cooled first, the pressure inside will cause them to pop open, potentially ejecting some of the sample.
28. Adding too much acid will prevent the EtOH precipitation in the next step from working properly. It is not required to fully neutralize the base.
29. The pellet after this step is frequently easy to see from some salts precipitating. However, not seeing a pellet does not necessarily mean that the reverse transcription or precipitation failed, it could simply indicate that fewer salts precipitated. The amount of cDNA product expected should not be easily visible by eye without glycogen.
30. After precipitation following the ligation step, the pellet can be seen as a rounded white dot on the bottom of the tube. Because glycogen has been added, the pellet should be clearly visible. Not seeing a pellet after this step could be an indicator that the reverse transcription did not generate much cDNA, but does not necessarily indicate complete failure. Observing a brown pellet is indicative of insufficient removal of the base following the RNA hydrolysis step, which causes the CircLigase reaction

to fail. Note, however, that extended centrifugation (much greater than 30 min) may cause a mild brown discoloration to occur in the pellet in samples that were properly ligated.

31. While the Agencourt Ampure XP beads are generally thought of as a method for purifying dsDNA, they also work very well for ssDNA. We have noticed that ssDNAs under 50 nucleotides do not bind to the beads very well, but above 50 nucleotides ssDNAs bind well. The roughly 50 nucleotide cutoff helps remove unligated adapters and unextended reverse transcription primer.
32. The bead separation will sometimes proceed slowly. If so, allow more time for the beads to completely separate. Longer incubation times are fine here, and not allowing complete separation will introduce recovery losses.
33. Allowing the beads to sit in the EtOH wash for too long can introduce recovery losses.
34. Do not overdry the beads. We have observed some recovery issues when this occurs.
35. If using the BioAnalyzer to perform quality analysis, skip this step and continue to Subheading 3.5.
36. The pellet after this step should be a clearly visible round white dot. The fluorophores present can sometimes give the pellet a pinkish hue, which is normal.
37. Other capillary electrophoresis software programs are suitable as well. We use ShapeFinder since it provides an easy-to-view window and is freely available. However, none of the specific analysis tools inherent to ShapeFinder are required. One software alternative is the freely available Peak Scanner software (Life Technologies).
38. The GeneScan 500 LIZ standard has peaks at: 35, 50, 75, 100, 139, 150, 160, 200, 250, 300, 340, 350, 400, 450, 490, and 500 nucleotides.
39. The smaller the peak at 80 nucleotides, the better the library quality. The side product is sequenced and takes away from the usable sequencing reads for alignment and reactivity calculation. While some side product is always expected, it is typically at a low enough level that data analysis is not a problem, although up to 50% side product is not unusual. However, if the library is completely dominated by side product, few reads will be usable and more material will need to be sequenced to get an adequate number of usable reads. The amount of side product visible is typically related to how well the ligation and reverse transcription steps were performed. Some RNA sequences could, however, ligate or reverse-transcribe poorly due to interference caused by secondary structures.

40. As a standard, nine cycles of PCR amplification are used to minimize potential bias introduced to the data from PCR. However, we have recently shown that up to 20 cycles can be used instead (if higher concentrations are needed) with little to no apparent bias introduced [12].
41. This cooling step is very important. If skipped, the dsDNA libraries could be partly heat-denatured and digested by ExoI when added. ExoI is not immediately heat deactivated, so skipping this cooling step will likely result in complete digestion of the dsDNA libraries.
42. This step is not explicitly required, but does help improve accuracy when measuring library concentration. Errors in quantification can affect library balancing when sequencing, especially for low concentration dsDNA libraries.
43. When eluting, be careful not to over-pipette. Frequently, traces of the surfactants present in the 5× Phusion buffer can make the final dsDNA solution appear somewhat bubbly. These bubbles will not affect the quantification or sequencing, however.
44. Libraries with less than 2 ng/μL are likely to be mostly dimer side product due to poor reverse transcription. These libraries can be sequenced, but will likely require more sequencing depth to overcome the amount of side product present. A library concentration of 10 ng/μL or higher is considered optimal for nine cycles of PCR. Be mindful that increasing the number of cycles to increase the library concentration will also increase the amount of side product and not help overcome problems with too much side product.
45. Because the raw number of reads required per “+” and “-” library pair is around 1–2 M, many pairs of libraries can be run together in a single MiSeq or HiSeq run. We recommend obtaining this many reads to overcome data losses due to sequencing the dimer side product. Many different libraries can be prepared together or separately to be sequenced all at once. However, when multiplexing many libraries together, each library pair needs to be able to uniquely align with Bowtie. Using all different TruSeq indexes (*see Note 14*) for each library pair will generate separate sequencing data files, thus ensuring unique alignments. If multiple library pairs share the same TruSeq index, be sure that their constituent RNA sequences are all unique, or are all barcoded to be unique (*see Note 9*).
46. If using the BioAnalyzer for quality analysis, run 1 μL of each of the dsDNA libraries on a high sensitivity DNA chip and use the resulting electropherograms to determine what the average fragment lengths of the “+” and “-” libraries are.

47. If the BioAnalyzer was used to determine the average lengths of the dsDNA libraries, use this formula instead:
- $$\frac{1000 \times x}{607.4 \times L + 157.9}$$
48. The number of libraries sequenced and the sequencing platform should be chosen such that each library pair should have 1–2 M sequencing reads for downstream data analysis. This is a conservative recommendation, and fewer reads can be used, though we recommend 1–2 M when starting out with this protocol. This gives roughly 20–25 libraries for the MiSeq or 100–150 libraries per HiSeq lane.
49. To provide a certain degree of randomness for the Illumina sequencing platform, some fraction of the sequenced library should be PhiX control DNA (provided with sequencing kit). We suggest using 5 % PhiX, although higher concentrations of the PhiX control will only improve sequencing quality. Due to differences in the phasing correction between the MiSeq and HiSeq, using a higher %PhiX with the HiSeq platform may be advisable. The choice of sequencing with MiSeq or HiSeq will not alter the final data [12].
50. Because SHAPE-Seq v2.0 aligns the sequencing data to a file of known RNA targets, only a short stretch of bases are required to confidently align the sequencing reads. Longer reads would only be necessary if the RNA sequence was unknown and needed to be assembled de novo.
51. `gunzip` can be run on the Linux/Unix command line or Mac OS X terminal with the command: `gunzip *.fastq.gz`. The fastq files should be named according to the provided sample name (when samples were sequenced) and TruSeq index.
52. As an example of constructing the targets file, consider an RNA of interest that contains a polyA at the 3' end: ...AAAAAAA. To include the linker sequence, add “CACTCGGGCACC AAGGAC” to the 3' end of the RNA of interest sequence. Without any internal barcodes the targets file should read:
- ```
>RNA_of_interest
...AAAAAAACACTCGGGCACCAAGGAC
```
- If an “aa” barcode was included in the reverse transcription primer, add the reverse complement to the 3' end, resulting in the following for a barcoded target:
- ```
>RNA_of_interest
...AAAAAAACACTCGGGCACCAAGGACtt
```
53. Detailed documentation can be found at <http://github.com/LucksLab/spats>. Run time scales with the number of reads in a given fastq file pair. Under 5 M reads, the analysis should take less than an hour. For 25 M reads, the adapter trimming could

take multiple hours, mostly due to the stepwise trimming portion of `adapter_trimmer.py` (see **Note 55**). The screen output of the program will indicate how long each step is taking and provide an idea of how long the entire algorithm will take to complete.

54. There are a few options that can be added to the `adapter_trimmer.py` command to specify variables important to the analysis. Use the “-h” or “--help” options to reveal all of them. Most of the options will be unnecessary for most users of this protocol. Two options may be useful. The first is “--read-len <N>”, where N is the $2 \times N$ bp paired-end read length, to be included if something other than 2×35 is used (which is the default value). The other is “--trim-match <N>”, where N , in this case, is the number of bases used to search for Illumina adapter sequence (see **Note 55**). The “trim-match” variable only needs to be increased from the default value, 6, if any of the RNAs of interest contain the beginning of an Illumina adapter sequence. It can be manually changed if this is the case or, more easily, optimally determined using the `targets_analyzer.py` script included in the Spats source code from <http://github.com/LucksLab/spats>.
55. `adapter_trimmer.py` works by first determining if a sequencing read pair aligns to any of the targets provided. If not, the read pair is searched to determine if it contains a portion of the Illumina adapter sequence based on the *trim-match* variable (see **Note 54**). If no adapter sequence is found, base-by-base sequential trimming is used to determine if only a few bases of adapter are present in the sequencing read (shorter than the *trim-match* length). By exhaustively searching all of the sequencing reads in this manner, all adapter sequences are removed that perfectly match any length of the adapter sequence.
56. Spats runs much more quickly than `adapter_trimmer.py` (see **Note 53**) and should take less than an hour to calculate reactivities for 25 M reads.
57. The `reactivities.out` file should show a range of reactivities (θ) for all nucleotides in an RNA. Typically, a fair number of values are “0”, but the total number of 0s should generally not exceed 80% or so. Reads aligned to every nucleotide in both the “+” and “-” samples should be observed, although the exact number of reads will vary from nucleotide to nucleotide. Because reactivities are calculated using fragment distributions [22, 23], it is possible to see positions where the “+” sample alignments exceed the “-” sample, but the reactivity is “0”. Also, aligned reads tend to accumulate at the 3' end of the RNA due to abortive reverse transcription. Thus, the reactivities of the last four nucleotides in an RNA sequence may show unusual patterns. If no reads are aligning in either the “+” or “-” sam-

ples, the targets file should be the first place to look for issues. Common problems include miscopied RNA sequence and incorrect reverse complementarity for the linker/internal barcode sequences. Comparing the raw fastq files in Read 1 to the 3' ends of the targets is generally the best place to look for inconsistencies between the targets file and the sequencing data. If the targets file contains all the correct sequence, extensive sequencing errors (having degenerate bases appear, for example) could cause poor alignment. If these errors occur at the 3' end of either the Read 1 or Read 2 reads, they can be trimmed off using `fastx_trimmer` from the FastX toolkit without affecting data calculation. Lastly, poor alignment of one sample could be due to an underabundance of that sample, typically caused by addition of the wrong RT primer, pipetting errors when preparing the library mix for sequencing, or incorrect balancing calculations/average length determination.

58. A repository of chemical probing data, including SHAPE-Seq v2.0, can be found at the RNA Mapping Database (rmdb.stanford.edu/repository/) [13]. We highly recommend that users of this method deposit their data in this database after publication for easy community access. See Loughrey et al. for locations of example SHAPE-Seq v2.0 data [12].
59. Bar charts of the reactivity values, ρ , can serve as “fingerprints” for an RNA structure. Clusters of highly reactive positions are strong indicators of loop regions that are very flexible. Likewise, clusters of very low reactivity positions can indicate base-paired nucleotides. One typical pattern to observe contains clusters of low reactivity, followed by high then low again, which suggest a stem-loop. Note that tertiary interactions and noncanonical base pairs can also affect nucleotide reactivity and make data interpretation more difficult. Further, these types of interactions are usually not included in RNA folding algorithms and can lead to oversimplified views of predicted RNA structures. Thus, we typically focus on the reactivity map as the final form of SHAPE-Seq data, supplemented with a number of possible interpretations without relying on a single minimum free energy structure.
60. RNAstructure [25] has long supported SHAPE reactivity values as restraints for secondary structure prediction. RNAstructure can be accessed via web server or downloaded as a GUI or command line executable (<http://rna.urmc.rochester.edu/RNAstructure.html>). To include restraints, a reactivities file must first be generated. To do this, create a text file with two tab-separated columns, where the first column has numbers from 1 to the exact length of the RNA being folded and the second has the reactivity (ρ) value. Use the value “-999” for bases to ignore or if there is no data present for that base. Save this file as *.shape instead of *.txt, then

input the *.shape file and the values 1.1 for m and -0.3 for b to the SHAPE Constraints subsection of the Optional Data section of the RNAstructure webserver (<http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html>) when folding the RNA. The output will be restrained with the SHAPE-Seq reactivities. The *Fold* algorithm of RNAstructure adds an extra pseudo-free energy term of the form $\Delta G_{\text{SHAPE}} = m \ln(\rho + 1) + b$ during total free energy calculation. The values m and b were heuristically determined to be 1.1 and -0.3 , respectively, for SHAPE-Seq v2.0 using a panel of well-characterized RNAs [12]. Alternatively, *ShapeKnots*, which includes pseudoknots, can be used in place of *Fold* [34].

Acknowledgements

We thank Alex Settle for assistance on experimental procedures, Peter Schweitzer and the Cornell Life Sciences Core facility for sequencing support, and David Loughrey and James Chappell for helpful comments in reviewing this manuscript. This work was supported by the National Science Foundation Graduate Research Fellowship Program (grant number DGE-1144153 to K.E.W.); the Cornell University Center for Life Sciences Enterprises, a New York State Center for Advanced Technology supported by New York State and industrial partners (grant number C110124 to J.B.L.); and a New Innovator Award through the National Institute of General Medical Sciences of the National Institutes of Health (grant number DP2GM110838 to J.B.L.). K.E.W. is a Fleming Scholar in the Robert F. Smith School of Chemical and Biomolecular Engineering at Cornell University. J.B.L. is an Alfred P. Sloan Research Fellow.

References

1. Weeks KM (2010) Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol* 20:295–304
2. Butcher SE, Pyle AM (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc Chem Res* 44:1302–1311
3. Duncan CDS, Weeks KM (2008) SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA. *Biochemistry* 47:8504–8513
4. McGinnis JL, Weeks KM (2014) Ribosome RNA assembly intermediates visualized in living cells. *Biochemistry* 53:3237–3247
5. Rouskin S, Zubradt M, Washietl S et al (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505:701–705
6. Ding Y, Tang Y, Kwok CK et al (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505:696–700
7. Talkish J, May G, Lin Y et al (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* 20:713–720
8. Siegfried NA, Busan S, Rice GM et al (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 11:959–965
9. Homan PJ, Tandon A, Rice GM et al (2014) RNA tertiary structure analysis by 2'-hydroxyl molecular interference. *Biochemistry* 53:6825–6833

10. Cheng C, Chou FC, Kladwang W et al (2015) Consistent global structures of complex RNA states through multidimensional chemical mapping. *eLife* 4, e07600
11. Byrne RT, Konevega AL, Rodnina MV, Antson AA (2010) The crystal structure of unmodified tRNAPhe from *Escherichia coli*. *Nucleic Acids Res* 38:4154–4162
12. Loughrey D, Watters KE, Settle AH, Lucks JB (2014) SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res* 42:e165
13. Cordero P, Lucks JB, Das R (2012) An RNA mapping DataBase for curating RNA structure mapping experiments. *Bioinformatics* 28:3006–3008
14. Peattie DA, Gilbert W (1980) Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci U S A* 77:4679–4682
15. McGinnis JL, Dunkle JA, Cate JHD, Weeks KM (2012) The mechanisms of RNA SHAPE chemistry. *J Am Chem Soc* 134:6617–6624
16. Culver GM, Noller HF (1998) Directed hydroxyl radical probing of 16S ribosomal RNA in ribosomes containing Fe(II) tethered to ribosomal protein S20. *RNA* 4:1471–1480
17. Brunel C, Romby P (2000) Probing RNA structure and RNA-ligand complexes with chemical probes. *Methods Enzymol* 318:3–21
18. Mortimer SA, Weeks KM (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* 129:4144–4145
19. Spitale RC, Flynn RA, Torre EA et al (2014) RNA structural analysis by evolving SHAPE chemistry. *Wiley Interdiscip Rev RNA* 5:867–881
20. Wilkinson KA, Merino EJ, Weeks KM (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* 1:1610–1616
21. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 127:4223–4231
22. Aviran S, Trapnell C, Lucks JB et al (2011) Modeling and automation of sequencing-based characterization of RNA structure. *Proc Natl Acad Sci U S A* 108:11069–11074
23. Aviran S, Lucks JB, Pachter L (2011) RNA structure characterization from chemical mapping experiments. 49th annual Allerton Conference on communication, control, and computing, pp 1743–1750. doi:10.1109/Allerton.2011.6120379
24. Steen K-A, Rice GM, Weeks KM (2012) Fingerprinting noncanonical and tertiary RNA structures by differential SHAPE reactivity. *J Am Chem Soc* 134:13160–13163
25. Mathews DH, Disney MD, Childs JL et al (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* 101:7287–7292
26. Low JT, Weeks KM (2010) SHAPE-directed RNA secondary structure prediction. *Methods* 52:150–158
27. Vasa SM, Guex N, Wilkinson KA et al (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* 14:1979–1990
28. Lucks JB, Mortimer SA, Trapnell C et al (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc Natl Acad Sci U S A* 108:11063–11068
29. Mortimer SA, Trapnell C, Aviran S et al (2012) SHAPE-Seq: high-throughput RNA structure analysis. *Curr Protoc Chem Biol* 4:275–297
30. Avis JM, Conn GL, Walker SC (2012) Cis-acting ribozymes for the production of RNA in vitro transcripts with defined 5' and 3' ends. *Methods Mol Biol* 941:83–98
31. Kao C, Zheng M, Rüdiger 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
32. Turner R, Shefer K, Ares M (2013) Safer one-pot synthesis of the "SHAPE" reagent 1-methyl-7-nitroisatoic anhydride (1m7). *RNA* 19:1857–1863
33. Mortimer SA, Weeks KM (2009) Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat Protoc* 4:1413–1421
34. Hajdin CE, Bellaousov S, Huggins W et al (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc Natl Acad Sci U S A* 110:5498–5503