# Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies

Steven Busan,[†] Chase A. Weidmann,[†] Arnab Sengupta, and Kevin M. Weeks*[ID]

Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290, United States

S Supporting Information

ABSTRACT: Chemical probing is an important tool for characterizing the complex folded structures of RNA molecules, many of which play key cellular roles. Electrophilic SHAPE reagents create adducts at the 2′-hydroxyl position on the RNA backbone of flexible ribonucleotides with relatively little dependence on nucleotide identity. Strategies for adduct detection such as mutational profiling (MaP) allow accurate, automated calculation of relative adduct frequencies for each nucleotide in a given RNA or group of RNAs. A number of alternative reagents and adduct detection strategies have been proposed, especially for use in living cells. Here we evaluate five SHAPE reagents: three previously well-validated reagents

1M7 (1-methyl-7-nitroisatoic anhydride), 1M6 (1-methyl-6-nitroisatoic anhydride), and NMIA (*N*-methylisatoic anhydride), one more recently proposed NAI (2-methylnicotinic acid imidazolide), and one novel reagent 5NIA (5-nitroisatoic anhydride). We clarify the importance of carefully designed software in reading out SHAPE experiments using massively parallel sequencing approaches. We examine SHAPE modification in living cells in diverse cell lines, compare MaP and reverse transcription–truncation as SHAPE adduct detection strategies, make recommendations for SHAPE reagent choice, and outline areas for future development.
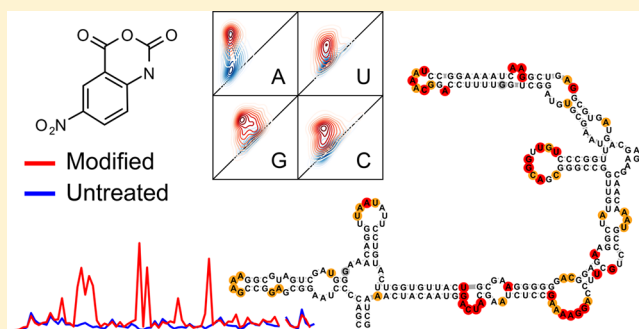
Many RNA molecules fold into complex structures that drive key processes, such as catalysis,[1] viral packaging, protein binding,[2] splicing,[3] polyadenylation,[4] transcription termination,[5] transcript degradation,[6] and translation efficiency.[7] For decades, chemical probing has been an important strategy for mapping structural features of nucleic acids. The pursuit of reagents and a technology that report on the local structure of all four ribonucleotides in a concise experiment led to the development of SHAPE (for selective 2′-hydroxyl acylation analyzed by primer extension), a straightforward approach using an electrophilic reagent to modify the backbone of conformationally dynamic nucleotides, read out by truncations during reverse transcription (RT) to identify adduct locations and quantify their relative abundances.[8] SHAPE has been applied to study diverse RNAs, including small riboswitches;[9−11] bacteriophage, bacterial, and eukaryotic mRNAs;[12−16] and the RNA genomes of complete human viruses.[17−20]

Multiple efforts then focused on increasing the throughput and scope of SHAPE experiments by replacing or supplementing the adduct detection step with methods integrated with massively parallel sequencing. The sequencing reads generated from adduct-containing RNAs are then quantified with automated software to generate per-nucleotide reactivities. Most of these methods rely on counting truncations that occur when RT enzymes terminate at RNA adducts.[21−24] These RT-truncation methods require follow-up steps to ligate and process the initial cDNA products into a sequenceable library

format, potentially affecting the information recorded during RT truncation. Mutational profiling (MaP), by contrast, detects internal mutations in cDNA generated when an RT enzyme reads through RNA adducts during relaxed-fidelity DNA synthesis.[25] Downstream library construction steps have little to no effect on the information recorded during the MaP read-through step. MaP has enabled rapid structural analysis of many RNAs relevant to basic biology and human health including HIV,[25,26] the *Xist* long noncoding RNA,[27] a satellite tobacco mosaic virus genome in packaged and unpackaged states,[28] rRNAs from divergent bacteria,[29] hundreds of *Escherichia coli* transcripts,[30] human *SERPINA1* mRNAs,[31] and bacterial and human protein-binding targets.[27,32,33]

The SHAPE reagents 1-methyl-7-nitroisatoic anhydride (1M7), 1-methyl-6-nitroisatoic anhydride (1M6), and *N*-methyl-nitroisatoic anhydride (NMIA) have been extensively validated by examining their reactivity on multiple RNAs of known structure.[34,25,35] Recent efforts have focused on improving SHAPE modification and adduct detection efficiency, especially in living cells.[36,27,37] Additional SHAPE reagents with long half-lives have been proposed such as 2-methyl-3-furoic acid imidazolide (FAI) and 2-methylnicotinic acid imidazolide (NAI).[38] New methods for enhancing adduct

detection have also been introduced, including biotin enrichment approaches coupled with the SHAPE reagent N-propanone isatoic anhydride (NPIA)[23] and the clickable reagent 2-methylnicotinic acid imidazolide-azide (NAI-N$_3$).[24] These new reagents and adduct detection methods have not been as extensively validated on large RNAs of known and complex structures as have 1M7 and MaP. Here we report results of MaP experiments using multiple SHAPE reagents across different cell types and compare MaP and RT truncation as SHAPE adduct detection strategies. We include a new reagent with favorable properties, 5- nitroisatoic anhydride (5NIA). This study clarifies the importance of carefully designed software in reading out SHAPE experiments using massively parallel sequencing approaches, provides guidelines for SHAPE reagent choice, and identifies areas for future development.

■ METHODS

Cell Culture. Mouse embryonic stem cells (SM33) were cultured on tissue culture plates coated with 0.1% gelatin and grown in high-glucose Dulbecco's modified Eagle's medium (DMEM) with sodium pyruvate supplemented with 15% fetal bovine serum (FBS), 0.1 mM non-essential amino acids (Gibco), 2 mM L-glutamine, 1000 units/mL ESGRO leukemia inhibitory factor (Millipore Sigma), 100 units/mL penicillin, 100 μg/mL streptomycin, and 0.1 mM 2-mercaptoethanol. C2C12 myoblast cells were cultured on tissue culture plates in DMEM supplemented with 10% FBS, 100 units/mL penicillin, and 100 μg/mL streptomycin, and experiments were performed before cells were completely confluent. HSV40 fibroblasts were cultured on tissue culture plates in DMEM supplemented with 15% FBS, 100 units/mL penicillin, and 100 μg/mL streptomycin. LNCaP cells were cultured on tissue culture plates in RPMI-1640 supplemented with 10% FBS, 100 units/mL penicillin, and 100 μg/mL streptomycin. Jurkat T cells were grown in suspension culture flasks using the same medium that was used for LNCaP cells.

Mammalian Cell Lysis and Protein Digestion. C2C12 myoblasts, H4SV fibroblasts, LNCaps, or mouse embryonic stem cells were grown in two 10 cm dishes to ~80% confluency. Both plates were washed once in PBS before being scraped and before lysis in 2.5 mL of ice-cold cytoplasmic lysis buffer [40 mM Tris (pH 8), 175 mM NaCl, 6 mM MgCl$_2$, 1 mM CaCl$_2$, 256 mM sucrose, 0.5% Triton X-100, 0.5 unit/μL RNasin (Promega), and 0.45 unit/μL DNase I (Roche)]. Jurkat cells (grown in a suspension culture) were pelleted at 3000g for 5 min, washed once with PBS, and then resuspended in lysis buffer. Cells were lysed for 5 min on ice with intermittent mixing. Cell nuclei were pelleted at 3000g for 5 min at 4 °C (nuclear fraction), and the resulting supernatant (cytoplasmic fraction) was transferred to a new tube. A volume of 2.5 mL of proteinase K buffer [40 mM Tris (pH 8), 200 mM NaCl, 1.5% sodium dodecyl sulfate, and 0.5 mg/mL proteinase K] was added to the nuclear pellets. The supernatant (cytoplasmic fraction) volume was increased to 5 mL with appropriate solutions to yield final concentrations of 200 mM NaCl, 1.5% sodium dodecyl sulfate, and 0.5 mg/mL proteinase K (Thermo Fisher). Proteins were digested for 45 min at 23 °C with intermittent mixing.

***E. coli* Cell Lysis and Protein Digestion.** A 25 mL aliquot of E. coli cells at an OD$_{600}$ of 0.5 was pelleted at 8000g and 4 °C for 10 min. Cells were lysed in 16.5 mL of E. coli lysis buffer [15 mM Tris (pH 8), 450 mM sucrose, 8 mM EDTA,

and 0.4 mg/mL lysozyme] for 5 min at 23 °C, and then for 10 min at 0 °C. Protoplasts were collected at 5000g and then resuspended in 2 mL of proteinase K buffer [50 mM HEPES (pH 8), 200 mM NaCl, 5 mM MgCl$_2$, 1.5% sodium dodecyl sulfate, and 0.2 mg/mL proteinase K], vortexed for 10 s, and then incubated at 23 °C for 5 min and 0 °C for 10 min.

RNA Extraction and Treatment with 1M7, 5NIA, or NAI. Nucleic acids were extracted twice with 1 volume of a phenol/chloroform/isoamyl alcohol solvent (25:24:1) that was pre-equilibrated with either 1.1× mammalian RNA folding buffer [110 mM HEPES (pH 8), 110 mM NaCl, and 5.5 mM MgCl$_2$] or 1.1× E. coli RNA folding buffer [110 mM HEPES (pH 8), 110 mM NaCl, and 11 mM MgCl$_2$]. Excess phenol was removed through two subsequent extractions with 1 volume of chloroform. The final aqueous layer was buffer exchanged into 1.1× RNA folding buffer using PD-10 desalting columns (GE Healthcare Life Sciences). The resulting RNA solution (7 mL for mammalian cells and 3.5 mL for E. coli) was incubated at 37 °C for 20 min before being split into two equal volumes. For mammalian samples, 350 μL of 10× SHAPE reagent (250 mM 5NIA, 100 mM 1M7, or 1 M NAI) in dimethyl sulfoxide (DMSO) was added to one sample, and 350 μL of neat DMSO was added to the other. For E. coli samples, the volume added was 175 μL. Samples were incubated at 37 °C for 10 min with 5NIA (CAS Registry No. 4693-02-1; AstaTech catalog no. 69445) and NAI and for 5 min with 1M7 before the reagent was quenched with 1 M dithiothreitol (DTT) (550 μL for mammalian cell samples and 225 μL for E. coli samples). RNA was precipitated with 1/10 volume of 2 M NH$_4$OAc and 1 volume of isopropanol. After one wash with 75% ethanol, the resulting pellet was dried and resuspended in 88 μL of water and 10 μL of 10× TURBO DNase buffer and 4 units of DNase (TURBO DNase, Thermo Fisher) were added. The mixture was incubated at 37 °C for 1 h. RNA was purified using a 1.8× ratio of RNA-binding magnetic beads (Agencourt RNAClean XP, Beckman Coulter) and eluted into 20 μL of nuclease-free water.

In-Cell SHAPE Treatment and RNA Extraction. Adherent cells were grown in individual wells (35 mm diameter) of a six-well culture plate as described above. At ~80% confluency, cells were washed with PBS and replenished with 900 μL of fresh medium. To the cells was added 100 μL of SHAPE reagent in DMSO; the same volume of DMSO was added to the control cells. Jurkat cells were grown to ~80% confluency and pelleted at 3000g for 5 min, followed by one wash with PBS and resuspension in fresh growth medium. A 900 μL aliquot of this mixture was transferred to a well of a six-well culture plate and incubated at 37 °C. After 10 min, 100 μL of SHAPE reagent (1M7, 1M6, and NMIA at 100 mM; 5NIA at 250 mM; and NAI at 1 M) in DMSO was added to the well; controls were treated with DMSO. Cells were incubated at 37 °C for 15 min. Cells were either quenched at 15 min using 125 mM DTT or left unquenched. Cells were pelleted and washed once with PBS, followed by RNA extraction (TRIzol, Invitrogen). RNA pellets were dried and resuspended in 88 μL of nuclease-free water. Treatment with DNase (TURBO DNase, Thermo Fisher) and affinity purification (Agencourt RNAClean XP magnetic beads) were performed as described above for cell-free RNA extraction.

MaP of Total RNA. From each sample of RNA (defined above, E. coli or murine total RNA, extracted and SHAPE-modified or untreated), 1−3 μg was subjected to MaP reverse transcription (requiring Superscript II and addition of Mn^{2+} to

the RT buffer[25,39]) using random nonamer primers. The cDNA generated was buffer exchanged (Illustra microspin G-50 columns, GE Healthcare), and the volume increased to 68 μL. For second-strand cDNA synthesis, 8 μL of 10× buffer (Second Strand Synthesis Reaction Buffer, NEB) and 4 μL of enzyme (Second Strand Synthesis Enzyme mix, NEB) were added to the cDNA product and incubated for 2.5 h at 16 °C. Double-stranded cDNA was fragmented and amplified with sequencing indexes (Nextera XT library prep kit, Illumina). Nextera polymerase chain reaction (PCR) products were affinity purified (using a 0.8× ratio of Agencourt AMPure XP beads, Beckman Coulter) and eluted in 20 μL of nuclease-free water.

U1 snRNA MaP. From each sample of eukaryotic RNA, 1−3 μg was subjected to MaP reverse transcription (requiring Superscript II and addition of Mn^{2+} to RT buffer[25,39]) using primers specific for either mouse or human U1 RNA. The generated cDNA was buffer exchanged (over Illustra microspin G-50 columns, GE Healthcare). Output cDNA (5 μL) was used as a template for 25 μL PCR reactions (Q5 Hot-start polymerase, NEB) with primers made to amplify U1 and add adapter sequences (1× Q5 reaction buffer, each primer at 500 nM, 200 μM dNTPs, and 0.02 unit/μL Q5 Hot-start polymerase). PCR proceeded in a touchdown format: 98 °C for 2 min, 20 cycles of 98 °C for 10 s, 72 °C (decreasing by 1 °C each cycle until 64 °C) for 30 s, and 72 °C for 20 s, and finally 72 °C for 2 min. Step 1 PCR products were affinity purified (using a 1× ratio of Agencourt AMPure XP beads, Beckman Coulter) and eluted in 20 μL of nuclease-free water. Purified PCR products (2 ng) were used as a template in 50 μL of the PCR mixture to add the multiplex indices and remaining sequences necessary for Illumina sequencing (1× Q5 reaction buffer, each index primer at 500 nM, 200 μM dNTPs, and 0.02 unit/μL Q5 Hot-start polymerase). Step 2 PCR proceeded as follows: 98 °C for 2 min, 10 cycles of 98 °C for 10 s, 66 °C for 30 s, and 72 °C for 20 s, and finally 72 °C for 2 min. Step 2 PCR products were affinity purified (using a 0.8× ratio of Agencourt AMPure XP beads, Beckman Coulter) and eluted in 20 μL of nuclease-free water.

Illumina Sequencing. Libraries were pooled and sequenced on an Illumina Miseq instrument, outputting 2 × 300 paired-end data sets. For U1 libraries, 2 × 150 paired-end data sets were used.

MaP Analysis of Strongly Expressed Mouse Transcripts. Initial transcript target sequences were retrieved from Gencode release M15 (GRCm38.p5);[40] there were 131 100 transcripts. Transcripts corresponding to a single parent gene were limited to the first listed transcript (in most cases the best-supported spliced variant), giving 52 553 transcripts. The transcript for *Gapdh* (ENSMUST00000118875.7, Gapdh-203) was manually chosen on the basis of read counts mapped to each exon. Missing sequences for the 5.8S, 18S, and 28S rRNAs (accession numbers NR_003280.2, NR_003278.3, and NR_003279.1, respectively) were appended. Experimental reads were pseudomapped against these transcripts with *Kallisto*.[41] Transcripts were selected with at least 1000 pseudomapping reads (including reads mapping to multiple loci), yielding 77 transcripts. Reads were filtered to only those pseudomapping to any of these 77 targets. The *ShapeMapper* pipeline (version 2.1.2) was run on these filtered reads against each of the 77 target transcripts separately (this includes a mapping step using *Bowtie2*[42]). Transcripts with a coverage of more than 1000 reads over at least half of the sequence in both

plus reagent and minus reagent samples were retained, yielding 13 transcripts. Two exactly duplicated transcript sequences were removed. Pairwise sequence alignment was performed for the 11 remaining targets using *Needle*[43] with default parameters. Hierarchical clustering was performed in *python* using the *scipy.cluster.hierarchy* package based on 1 − (alignment score/alignment length), with the resulting distribution of scores scaled from 0 to 1. Linkage was calculated with *hierarchy.linkage* with *method* = '*average*'. Clusters were assigned with *hierarchy.fcluster* with *criterion* = '*distance*' and a threshold of 1. Each cluster of highly similar sequences was reduced to the single best-annotated transcript, first by eliminating sequences whose GENCODE information indicated no human curation and then by eliminating all but the longest sequence in a given cluster. This process eliminated two apparent 7SK sequences, one apparent 18S rRNA sequence, and a shortened sequence highly similar to CT010467.1, yielding seven final transcript sequences. Finally, a single *ShapeMapper* analysis run was performed with the pseudomapping reads as input and these seven transcripts as targets.

Analysis of RT-Truncation Data for Strongly Expressed Mouse Transcripts. Sequence reads for RT-truncation data sets were retrieved from the NCBI Sequence Read Archive accessions SRR1534952, SRR1534953, SRR1534954, and SRR1534955.[24] Reads were pseudomapped against seven transcripts of interest using *Kallisto*,[41] and any reads that mapped were used as inputs for a published RT-truncation analysis pipeline (*icSHAPE_pipeline.pl*).[44] Calculated reactivity values below 0 and above the 95th percentile were not clipped. Analysis of one transcript did not yield any usable data, possibly due to its short length (57 nucleotides); this transcript is therefore not included in Table S1.

Read Depth. Across all MaP data sets reported here, for a given position to be included in analysis, the read depth was required to be >5000 in both untreated and SHAPE-modified samples.

Mutation Detection and Adduct Inference. A significant part of the information contained in a MaP experiment lies in insertions, deletions, and sequence changes more complex than simple mutations relative to the reference sequence, and it is important to account for this signal accurately. *ShapeMapper2*[35] locally realigns ambiguous sequence deletions and inserts such that a deletion or insert is placed at the 5′-most position out of the set of compatible alignment locations. Mutations separated by five or fewer unchanged reference positions are grouped and treated as arising from a single chemical adduct. For each group of mutations, the adduct location is inferred to be the reference position 5′ of the unchanged reference position 3′ of the mutation group. See Figure 1D and Figure S2 in ref 35 for examples; implementation details are provided in the software documentation.

Reference Structures. Accepted secondary structure models for an *E. coli* TPP riboswitch and tRNA-Phe were based on crystallographic structures, as previously reported.[45] Structures for *E. coli* rRNAs (both large and small subunits, including pseudoknots, corresponding to GenBank sequence accession number J01695) and for *Mus musculus* 18S rRNA and 12S mitochondrial rRNA were retrieved from the Comparative RNA Web site.[46] A structure for *M. musculus* 7SK snRNA was retrieved from Ensembl for transcript

ENSMUST00000083103.[47] The secondary structure for *Homo sapiens* U1 was based on a crystal structure.[48]

RESULTS

MaP Accuracy. A critical criterion for any structure probing strategy is that chemical adducts created on the RNA be reported accurately and quantitatively. There is close agreement between previous direct, one-step electrophoresis-based primer extension readouts and the massively parallel sequencing-based MaP strategy (Figure 1). Prior work includes
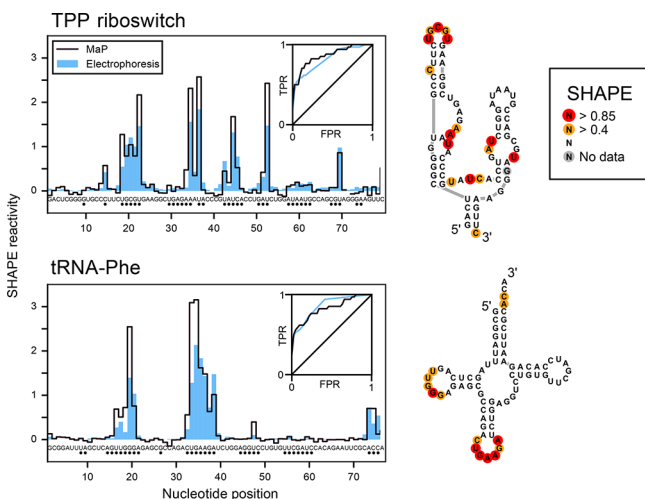

Figure 1. SHAPE data read out by MaP or one-step electrophoresis-detected reverse transcription. Histograms show normalized SHAPE reactivities, using 1M7, as a function of detection approach. Dots below the nucleotide sequence indicate unpaired nucleotides according to accepted structure models. Insets show receiver operator characteristic (ROC) curves for reactivity profiles compared to base pairing status. Sweeping a threshold over all ranked reactivity values, the true positive rate (TPR) is calculated as the number of unpaired nucleotides with reactivities above a given threshold divided by the total number of unpaired nucleotides and the false positive rate (FPR) is calculated as the number of base-paired nucleotides with reactivities above a given threshold divided by the total number of base-paired nucleotides. Areas under the ROC curve for the TPP riboswitch are 0.86 and 0.82 for MaP and electrophoresis, respectively; for tRNA-Phe, areas are 0.82 and 0.86 for MaP and electrophoresis, respectively. Reactivities read out by MaP are shown on the accepted secondary structure models. It is expected that some nucleotides in regions shown as single-stranded in the secondary structure are unreactive, as these regions are constrained by noncanonical and tertiary interactions. MaP[25] and electrophoresis data[45] were reported previously. Both RNA elements are based on *E. coli* sequences and were transcribed *in vitro* and refolded prior to probing.

extensive comparisons between electrophoresis readouts and MaP, based on analyses of long and complex transcripts such as bacterial rRNAs and diverse smaller structured RNAs and on benchmarking analyses that show that MaP reads out SHAPE experiments with the same or greater accuracy as simpler (but low-throughput) methods.[25,35]

The detection accuracy of MaP is closely linked both to performing the experiment under the right conditions and to the design and quality of the algorithms used to call adduct-induced sequence changes. MaP conditions have been carefully optimized for detection of SHAPE adducts.[25,39] Some prior reports that have attempted to evaluate MaP have used ad hoc enzymes and reverse transcription conditions unlikely to

produce the optimal signal over background.[49,50] Both point mutations (which are relatively easy to detect) and more complex deletions, insertions, and other sequence changes contribute to the MaP signal. Indeed, approximately 70% of the MaP signal is due to these more complex features (see Figure 1C in ref 35). Some software packages for reading out MaP experiments have ignored these components of the signal,[49,51] potentially raising sequencing coverage requirements for accurate readout or leading to systematic errors in SHAPE adduct detection and quantification. Currently, *ShapeMapper2* (https://github.com/Weeks-UNC/shapemapper2) is one of the few software packages that carefully accounts for deletions and complex multinucleotide mutations (see Methods) and integrates this signal into SHAPE adduct detection.[35]

A variety of SHAPE reagents are used in RNA structure probing experiments. All five of the examined reagents, 1M6, 1M7, NMIA, NAI, and 5NIA, report on base pairing with comparable accuracies when read out by MaP, as shown by similar areas under the receiver operator characteristic (ROC) curve calculated with respect to unpaired versus paired nucleotide classification in accepted structures of *E. coli* rRNA and human U1 snRNA (Figure 2). An area under the curve of 1 indicates perfect agreement between nucleotide reactivity and pairing status, and a value of 0.5 indicates that nucleotide reactivity reports no base pairing information. SHAPE reactivities are primarily measurements of nucleotide
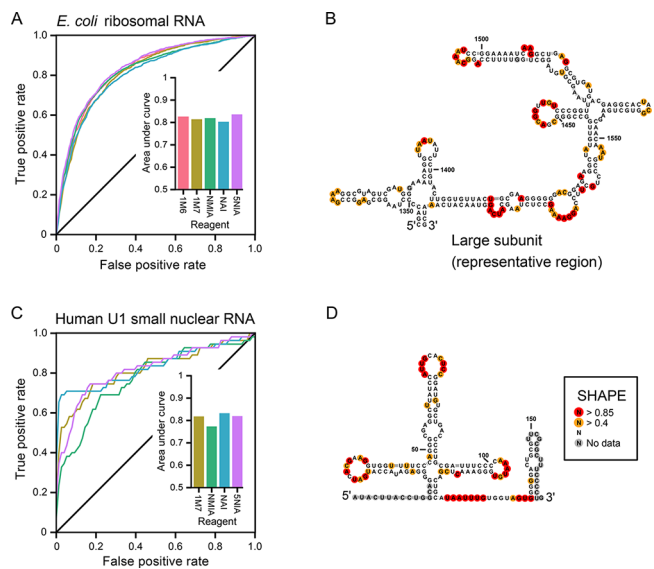

Figure 2. MaP readout for SHAPE reagents compared to base pairing status. (A) ROC curves for MaP data obtained from *E. coli* 16S and 23S rRNAs probed with five different SHAPE reagents under protein-free conditions. True positive rate and false positive rate calculated as described in the legend of Figure 1, with respect to base pairs in accepted CRW structures.[46] The inset shows the area under the ROC curve with respect to base pairing status. This analysis directly compares SHAPE−MaP reactivity profiles and accepted structures and does not involve SHAPE-directed structure modeling. (B) 1M7 reactivities shown on a representative region of the accepted 23S rRNA structure. (C) ROC curves for MaP data from the U1 snRNA isolated from Jurkat cells and probed with four SHAPE reagents under protein-free conditions. (D) 1M7 reactivities shown on the accepted U1 secondary structure. Data for 1M6-, 1M7-, and NMIA-probed rRNAs were reported previously.[34] See Methods for details on other data sets.

conformational flexibility. Base pairing is a major, but not the only, contributor to nucleotide conformation and flexibility; therefore, SHAPE reactivities show close but not perfect agreement with base pairing. When coupled with best-practices thermodynamic pseudo-free energy change minimization strategies,[45,52] data obtained from probing cell-free *E. coli* rRNA with each of the five reagents, read out with MaP and analyzed with *ShapeMapper2*, resulted in accurate RNA secondary structure models (Figure 3).
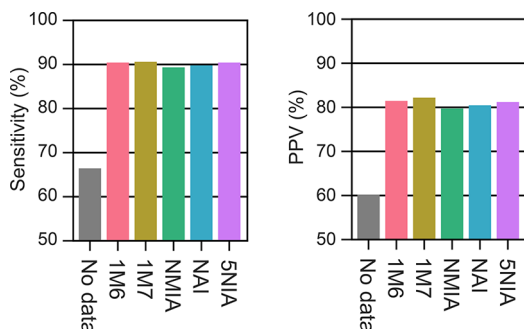

Figure 3. SHAPE-guided structure modeling accuracies for cell-free *E. coli* rRNA. Plots of sensitivity and positive predictive value (PPV) calculated by comparison of secondary structures modeled using data from the indicated SHAPE reagent to the accepted structure.[46] Structures were modeled using the *Fold* module of *RNAstructure*[52] as described previously.[25,45,60] For scoring, noncanonical base pairs and reference pairs >600 nucleotides apart in primary sequence were excluded. Sensitivity is the number of pairs shared between the accepted and modeled structures (allowing offsets of ±1 nucleotide) divided by the number of pairs in the accepted structure. PPV is the number of pairs in the modeled structure also present in the accepted structure (allowing offsets of ±1 nucleotide) divided by the number of pairs in the modeled structure. Regions in which SHAPE reactivities do not support the formation of specific helices[60,45,25] were excluded.

Reagent Biases. 1M7 was identified as an especially useful general purpose SHAPE reagent because of its short, but manageable, reaction half-life and its ability to report on the flexibility of all four ribonucleotides with similar reactivity.[53] We examined the per-nucleotide reactivity of 1M7 relative to background using a >4000-nucleotide data set. Per-nucleotide 1M7 reactivities (detected as MaP mutation rates) for cell-free *E. coli* rRNA confirmed the low nucleotide bias of this reagent (Figure 4, top). There is clear separation between the distributions of unpaired (red) and paired (blue) mutation rates, indicating that 1M7 reacts preferentially with unpaired nucleotides with very little bias for a particular nucleotide. We next evaluated whether the other four reagents had nucleotide biases (Figure 4 and Figure S1). NAI showed a pattern of MaP mutation rates markedly different from that of 1M7. NAI effectively distinguished unpaired and paired adenosine residues but less effectively distinguished unpaired from paired guanosine and cytosine (Figure 4, middle, and Figure S2). We suggest that the bias observed with NAI primarily reflects differences in nucleotide reactivity, but adduct detection rates and adduct reversal[54] may contribute. 5NIA showed excellent discrimination for all four ribonucleotides but reacted with adenosine at high relative rates compared to the other three nucleotides (Figure 4, bottom row, and Figure S2). Because adenosine residues are over-reactive with 5NIA, we recommend rescaling 5NIA SHAPE reactivity profiles on a per-nucleotide basis using factors estimated from the median
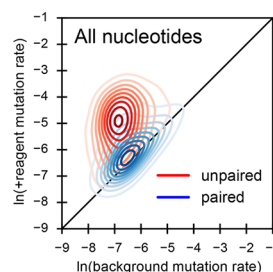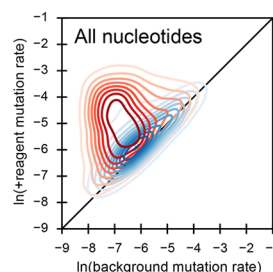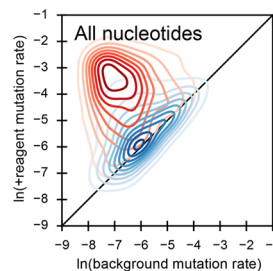

Figure 4. Nucleotide-specific base pair discrimination for various SHAPE reagents. Plots of mutation rates from SHAPE-treated versus background samples as calculated by *ShapeMapper2*.[35] For a given RNA position, the mutation rate is calculated as the mutation count divided by the local read depth. Data shown correspond to protein-free *E. coli* rRNA (both large and small subunits, same data as in Figure 2A). Contour lines show smoothed density, similar to a histogram. Data for the 1M6 and NMIA reagents are reported in Figure S1; distributions of per-nucleotide reactivity for all reagents are shown in Figure S2.

background-subtracted mutation rates of unpaired nucleotides (Figure S2). Rescaling facilitates visual interpretation of reactivity profiles and yields structure models of equivalent accuracy for the rRNAs examined here. Consistent with prior work,[55,34] the choice of reagent can have a strong effect on the reactivity profile.

MaP versus Click-Selective RT Truncation. We performed SHAPE on total RNA extracted from mouse B-lymphocytes, probing with the NAI reagent and reading out adducts with MaP. We compared this profile with a previously published NAI-N$_3$ click-selective RT-truncation data set of poly-A enriched extracted mouse RNA[24] (see Methods). SHAPE data read out by MaP and by click-selective RT truncation generally showed poor to modest correlations [Spearman r values spanned the range of 0−0.5 (Table S1)]. The observed low correlation is consistent with a study comparing truncated and internally mutated cDNA products generated by RT of DMS-modified RNA under conditions not optimal for MaP.[50] Despite an only modest correlation, for

transcripts with well-determined secondary structures, both readout methods showed comparable agreement with accepted base pairing models, based on AUC analyses for the few RNAs for which overlapping data sets are available. In general, very few RNAs with known structures have been examined with good read depth in their entirety using RT-truncation methods, limiting this analysis.

SHAPE Modification in Cells. Mutation rates obtained by SHAPE−MaP probing in cells are generally lower than those from probing cell-free RNAs, in agreement with previous reports.[38,37,56] This difference varies by reagent and cell line (Figure 5). The modification efficiency is likely affected by
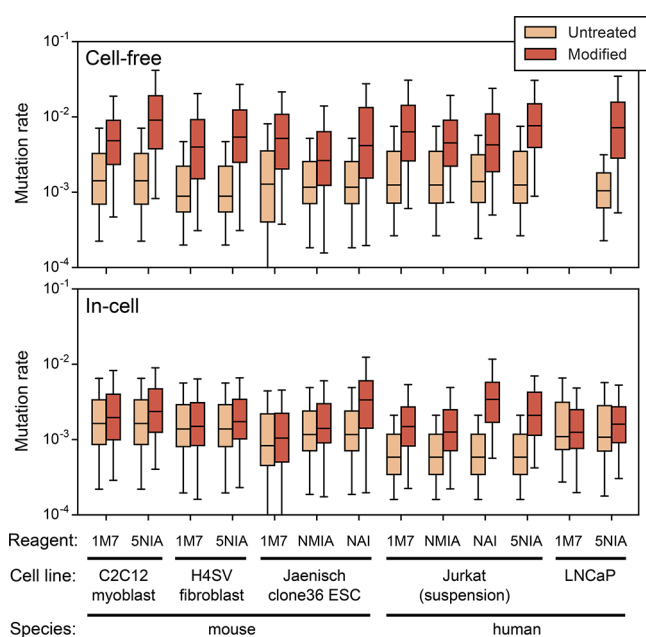

Figure 5. SHAPE reagent performance in live cells. SHAPE−MaP reactivities for U1 snRNA probed in total RNA isolated under nondenaturing conditions (cell-free) and in cells. U1 snRNA reactivities were determined by MaP using gene-specific primers[39] (see Methods). Box plots span the central 50% of the data, the interquartile range [IQR, from quartile 1 (Q1) to quartile 3 (Q3)]. The center line indicates the median. Whiskers indicate the most extreme points from Q1−1.5 × IQR to Q2 + 1.5 × IQR.

reagent half-life, reagent diffusion rate, and cell permeability. A few general trends are clear. All reagents evaluated here modify RNA in cells, based on a single reagent treatment, in the context of conscientiously performed experiments. Reactivity above background was clearly detected using a MaP readout for all reagents. 5NIA and NAI are more effective than 1M7 at modifying RNA in cells for all examined cell lines. Adherent cell lines (C2C12, H4SV, ESC, and LNCaP in Figure 5) typically had lower modification efficiencies with all reagents. An extracellular matrix or a reduced available cell surface area may reduce the in-cell SHAPE modification level. Conversely, cells that grow in suspension (Jurkat) were readily modified by all reagents (Figure 5).

■ DISCUSSION

Experimental Considerations. The design of an RNA structure probing experiment requires weighing the following criteria. The reagent should react readily under structurally informative or physiologically relevant conditions. Results from
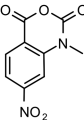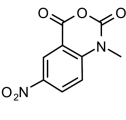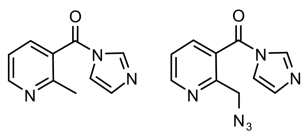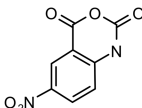
a probing experiment should be repeatable and tolerant of slight changes in conditions, and the procedure should be concise and contain no unnecessary steps. The experiment should accurately report on base pairing or other structural information and should ideally report on all four nucleotides with similar reactivity and high signal above background regardless of local sequence context. Importantly, data for evaluating these criteria should be obtained from carefully executed experiments with equivalent sequencing coverage across experiments and attentive adherence to published methods.

For RNA probing and structure analysis and modeling investigations focusing on *in vitro* transcribed and cell-free extracted RNAs, SHAPE experiments using 1M7 modification and MaP readout broadly satisfy these criteria. 1M7 remains notable for its relatively fast but manageable[53] and unbiased reactivity toward all four nucleotides (Figure 4). SHAPE probing in living cells requires a nuanced balancing of the strengths and weaknesses of each reagent.[56] Two broad trends are apparent from this and prior work. First, bacterial cells[57,58,30] and eukaryotic cells grown in a suspension culture [for example, Jurkat cells (Figure 5)] show higher levels of reactivity for all reagents. Second, longer-lived reagents like NAI and 5NIA exhibit better in-cell reactivity than faster-reacting reagents; however, these reagents have nucleotide biases, and for NAI, a quench step is necessary. Individual investigators and laboratory teams will balance these factors in different ways, but we offer the following recommendations (Table 1):

- 1M7 or 1M6 for general use, especially for experiments performed under cell-free conditions.
- NMIA for general use if 1M7 or 1M6 is not readily available; however, reaction times will be extended, increasing the risk of RNA degradation.
- 5NIA for in-cell probing, with rescaling by per-nucleotide relative reactivity factors (Figure S2).
- NAI for in-cell experiments where reactivities with alternative reagents are very low. The long half-life of NAI facilitates diffusion into cells, but the requisite long reaction time means that RNAs likely turn over many times during the experiment. This reagent has strong per-nucleotide reactivity variation (Figure 4) and must be quenched. TRIzol treatment does not inactivate NAI (Figure S3).

Adduct Detection Strategy. The experimental approaches and technical requirements for MaP versus RT-truncation strategies (either with or without click chemistry enrichment) differ substantially in the number of experimental steps, amount of experimental and hands-on time, and diversity of techniques and supplies required. Both methods share some procedures, including in-cell chemical probing, RNA purification, library amplification, and sequencing. An entire MaP procedure, starting with live cells and including RT and all steps shared with click-selective RT truncation, can be performed in 1−2 days. Click-selective RT truncation requires additional biochemical steps, including a copper-free click reaction, RNA fragmentation and end repair, adapter ligation, gel-based size selection of RNAs, streptavidin pull-down of RNA−cDNA hybrids, and further gel-based size selection of eluted cDNAs. These steps require an additional 2−3 days and may introduce readout biases. Likely due in part to these extensive differences in the experimental steps, there is a low to

Table 1. SHAPE Reagent Properties[a]

Short name	Name and structure	Solubility	Half-life at 37 °C	Treatment concentration (mM)	Notes
1M7	1-methyl-7-nitroisatoic anhydride	modest	17 s	10	General purpose reagent for cell-free studies. Very even per-nucleotide reactivity.
1M6	1-methyl-6-nitroisatoic anhydride	modest	31 s	10	
NMIA	N-methylisatoic anhydride	modest	260 s	10	Readily commercially available. Not recommended for in-cell probing.
NAI, NAI-N₃	2-methylnicotinic acid imidazolide[-azide]	high	~30 min	100-200	High modification rate in cells. Bias against guanosine and cytidine. Requires specific quench step.
5NIA	5-nitroisatoic anhydride	modest	~100 s	25	Fast-acting. Self-quenching. Highly reactive with adenosine. Useful for in-cell probing.

[a]Half-life refers to the reagent half-life in aqueous solution at pH 8.

modest quantitative correlation between experiments read out by MaP and those read out by RT truncation (Table S1 and ref 36). We did find that, despite the low per-nucleotide correlation, AUC analysis revealed similar discrimination for unpaired nucleotides in the small subunit rRNA (Table S1).

Some studies have proposed that MaP and RT-truncation readouts provide complementary information.[49,50] However, prior comparative analyses employed RT conditions and enzymes not optimized for MaP, very low read depth cutoffs, and (in specific cases) software that ignored deletions[49] and presentation formats that compressed data points.[50] These factors introduce noise and sampling bias and impact interpretability. Moreover, the quantitative accuracy of MaP would benefit only if RT truncations were to show patterns substantially different from those of mutations after accounting for convolution by sequencing library preparation steps and if their inclusion improved correspondence with a validated external metric.

We recommend MaP over RT truncation as a readout method because of its extensive validation based on long RNAs with complex structures,[25,35,39] its much simpler and more concise experimental implementation, and the ability to directly examine rare cellular transcripts by target-specific RT-PCR without requiring additional steps.[27,56] The MaP readout works well for all of the reagents evaluated in this work. MaP also provides an intuitive within-sample divisor

(read depth) for computing relative adduct frequencies, enabling direct comparison between reactivities of different transcripts probed in the same experiment,[30] without requiring use of proxy metrics such as the Gini index.[59] For example, with a MaP readout, coding RNAs can be directly compared and shown to have SHAPE chemical modification rates higher than those of noncoding RNAs in *E. coli*, suggesting that they are less structured overall.[30]

Areas for Future Development. SHAPE–MaP is now a mature technology, but there is room for improvement in several areas. If SHAPE reagent solubility in water could be enhanced without substantially lengthening the reaction half-life, such reagents could enable higher signal above background. SHAPE reagents such as NAI enable signal enrichment in cell types in which reactivities of 1M7 and 1M6 are low but show nucleotide biases that might be reduced through further chemical exploration. 1M7 and 5NIA show excellent overall discrimination of most unpaired and paired nucleotides; however, discrimination at cytidines is not at the theoretical limit (Figure 4 and Figure S2) and could potentially be improved by increasing adduct formation or detection rates at these nucleotides. MaP reverse transcription enzyme and reaction conditions have been optimized for adduct read-through and signal above background,[17] but further reverse transcriptase design or screening could identify enzymes or

conditions with improved efficiencies or detection rate signalto- noise.

In summary, MaP is now a mature and well-validated strategy for RNA secondary structure probing with diverse SHAPE reagents both *in vitro* and in cells. The accuracy of MaP-based approaches for reading out the results of the chemical modification step matches that of direct one-step electrophoresis approaches. Compared to RT-truncation strategies, the MaP strategy is much simpler, is less time-consuming, and requires less initial material. The reagents examined here in cell-free experiments, 1M6, 1M7, NMIA, NAI, and 5NIA, report on RNA flexibility and enable accurate modeling of RNA secondary structure. For live cell probing, suspension cultures are readily amenable to treatment with many SHAPE reagents, whereas some adherent cell cultures may require longer half-life reagents to enhance signal above background. However, trade-offs displayed by longer-lived reagents, such as nucleotide biases and the potential to miss shorter time-scale information, should be weighed when designing an experiment. We think the potential of comprehensive, nucleotide-resolution probing of RNA structure has just begun to be realized. We expect that there will be many creative applications that will result in deeper understandings of the wide-ranging roles of RNA structure in governing information transfer in diverse biological systems.

ASSOCIATED CONTENT

Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.8b01218.

Three figures and three tables (PDF)

SHAPE reactivity profiles for all MaP experiments (Table S2) and reprocessed icSHAPE profiles (ZIP)

AUTHOR INFORMATION

Corresponding Author
*E-mail: weeks@unc.edu.

ORCID
Kevin M. Weeks: 0000-0002-6748-9985

Author Contributions
†S.B. and C.A.W. contributed equally to this work.

Notes
The authors declare the following competing financial interest(s): K.M.W. is an advisor to and holds equity in Ribometrix, to which mutational profiling technologies have been licensed.

REFERENCES
(1) Scott, W. G. (1998) RNA catalysis. *Curr. Opin. Struct. Biol. 8*, 720−726.
(2) Jones, S., Daley, D. T., Luscombe, N. M., Berman, H. M., and Thornton, J. M. (2001) Protein−RNA interactions: a structural analysis. *Nucleic Acids Res. 29*, 943−954.
(3) Buratti, E., and Baralle, F. E. (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol. 24*, 10505−10514.
(4) Zarudnaya, M. I., Kolomiets, I. M., Potyahaylo, A. L., and Hovorun, D. M. (2003) Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res. 31*, 1375−1386.
(5) Nielsen, S., Yuzenkova, Y., and Zenkin, N. (2013) Mechanism of eukaryotic RNA polymerase III transcription termination. *Science 340*, 1577−1580.
(6) Deutscher, M. P. (2006) Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res. 34*, 659−666.
(7) Pop, C., Rouskin, S., Ingolia, N. T., Han, L., Phizicky, E. M., Weissman, J. S., and Koller, D. (2014) Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol. Syst. Biol. 10*, 770.
(8) Merino, E. J., Wilkinson, K. A., Coughlan, J. L., and Weeks, K. M. (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc. 127*, 4223−4231.
(9) Stoddard, C. D., Gilbert, S. D., and Batey, R. T. (2008) Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA 14*, 675−684.
(10) Weeks, K. M., and Mauger, D. M. (2011) Exploring RNA Structural Codes with SHAPE Chemistry. *Acc. Chem. Res. 44*, 1280−1291.
(11) Rice, G. M., Busan, S., Karabiber, F., Favorov, O. V., and Weeks, K. M. (2014) RNA SHAPE Analysis of Small RNAs and Riboswitches. *Methods Enzymol. 549*, 165.
(12) Salim, N. N., and Feig, A. L. (2010) An Upstream Hfq Binding Site in the fhlA mRNA Leader Region Facilitates the OxyS-fhlA Interaction. *PLoS One 5*, e13028.
(13) Todd, G. C., and Walter, N. G. (2013) Secondary structure of bacteriophage T4 gene 60 mRNA: implications for translational bypassing. *RNA 19*, 685−700.
(14) Xue, S., Tian, S., Fujii, K., Kladwang, W., Das, R., and Barna, M. (2015) RNA regulons in Hox 5′ UTRs confer ribosome specificity to gene regulation. *Nature 517*, 33−38.
(15) Lisowiec, J., Magner, D., Kierzek, E., Lenartowicz, E., and Kierzek, R. (2015) Structural determinants for alternative splicing regulation of the MAPT pre-mRNA. *RNA Biol. 12*, 330−342.
(16) Choi, E. K., Ulanowicz, K. A., Nguyen, Y. A. H., Frandsen, J. K., and Mitton-Fry, R. M. (2017) SHAPE analysis of the htrA RNA thermometer from Salmonella enterica. *RNA 23*, 1569−1581.
(17) Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr, Swanstrom, R., Burch, C. L., and Weeks, K. M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature 460*, 711.
(18) Burrill, C. P., Westesson, O., Schulte, M. B., Strings, V. R., Segal, M., and Andino, R. (2013) Global RNA Structure Analysis of Poliovirus Identifies a Conserved RNA Structure Involved in Viral Replication and Infectivity. *J. Virol. 87*, 11670−11683.
(19) Pollom, E., Dang, K. K., Potter, E. L., Gorelick, R. J., Burch, C. L., Weeks, K. M., and Swanstrom, R. (2013) Comparison of SIV and HIV-1 Genomic RNA Structures Reveals Impact of Sequence Evolution on Conserved and Non-Conserved Structural Motifs. *PLoS Pathog. 9*, e1003294.
(20) Pirakitikulr, N., Kohlway, A., Lindenbach, B. D., and Pyle, A. M. (2016) The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures. *Mol. Cell 62*, 111−120.
(21) Loughrey, D., Watters, K. E., Settle, A. H., and Lucks, J. B. (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−e165.

(22) Seetin, M. G., Kladwang, W., Bida, J. P., and Das, R. (2014) Massively parallel RNA chemical mapping with a reduced bias MAP-seq protocol. *Methods Mol. Biol. 1086*, 95−117.

(23) Poulsen, L. D., Kielpinski, L. J., Salama, S. R., Krogh, A., and Vinther, J. (2015) SHAPE Selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data. *RNA 21*, 1042−1052.

(24) Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H. Y., Batista, P. J., Torre, E. A., Kool, E. T., and Chang, H. Y. (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature 519*, 486−490.

(25) Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E., and Weeks, K. M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods 11*, 959−965.

(26) Lavender, C. A., Gorelick, R. J., and Weeks, K. M. (2015) Structure-Based Alignment and Consensus Secondary Structures for Three HIV-Related RNA Genomes. *PLoS Comput. Biol. 11*, e1004230.

(27) Smola, M. J., Christy, T. W., Inoue, K., Nicholson, C. O., Friedersdorf, M., Keene, J. D., Lee, D. M., Calabrese, J. M., and Weeks, K. M. (2016) 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. 113*, 10322−10327.

(28) Larman, B. C., Dethoff, E. A., and Weeks, K. M. (2017) Packaged and Free Satellite Tobacco Mosaic Virus (STMV) RNA Genomes Adopt Distinct Conformational States. *Biochemistry 56*, 2175−2183.

(29) Lavender, C. A., Lorenz, R., Zhang, G., Tamayo, R., Hofacker, I. L., and Weeks, K. M. (2015) Model-free RNA sequence and structure alignment informed by SHAPE probing reveals a conserved alternate secondary structure for 16S rRNA. *PLoS Comput. Biol. 11*, e1004126.

(30) Mustoe, A. M., Busan, S., Rice, G. M., Hajdin, C. E., Peterson, B. K., Ruda, V. M., Kubica, N., Nutiu, R., Baryza, J. L., and Weeks, K. M. (2018) Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell 173*, 181−195.e18.

(31) Corley, M., Solem, A., Phillips, G., Lackey, L., Ziehr, B., Vincent, H. A., Mustoe, A. M., Ramos, S. B., Weeks, K. M., Moorman, N. J., and Laederach, A. (2017) An RNA structure-mediated, posttranscriptional model of human α-1-antitrypsin expression. *Proc. Natl. Acad. Sci. U. S. A. 114*, E10244−E10253.

(32) Schulmeyer, K. H., Diaz, M. R., Bair, T. B., Sanders, W., Gode, C. J., Laederach, A., Wolfgang, M. C., and Yahr, T. L. (2016) Primary and secondary sequence structure requirements for recognition and discrimination of target RNAs by Pseudomonas aeruginosa RsmA and RsmF. *J. Bacteriol. 198*, 2458−2469.

(33) Ball, C. B., Solem, A. C., Meganck, R. M., Laederach, A., and Ramos, S. B. V. (2017) Impact of RNA structure on ZFP36L2 interaction with luteinizing hormone receptor mRNA. *RNA 23*, 1209−1223.

(34) Rice, G. M., Leonard, C. W., and Weeks, K. M. (2014) RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA 20*, 846−854.

(35) Busan, S., and Weeks, K. M. (2018) Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA 24*, 143−148.

(36) Smola, M. J., Calabrese, J. M., and Weeks, K. M. (2015) Detection of RNA−Protein Interactions in Living Cells with SHAPE. *Biochemistry 54*, 6867−6875.

(37) Lee, B., Flynn, R. A., Kadina, A., Guo, J. K., Kool, E. T., and Chang, H. Y. (2017) Comparison of SHAPE reagents for mapping RNA structures inside living cells. *RNA 23*, 169−174.

(38) Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T., and Chang, H. Y. (2013) RNA SHAPE analysis in living cells. *Nat. Chem. Biol. 9*, 18−20.

(39) Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A., and Weeks, K. M. (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc. 10*, 1643−1669.

(40) Harrow, J., Denoeud, F., Frankish, A., Reymond, A., Chen, C.-K., Chrast, J., Lagarde, J., Gilbert, J. G., Storey, R., Swarbreck, D., Rossier, C., Ucla, C., Hubbard, T., Antonarakis, S. E., and Guigo, R. (2006) GENCODE: producing a reference annotation for ENCODE. *Genome Biol. 7*, S4.

(41) Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol. 34*, 525−527.

(42) Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods 9*, 357−359.

(43) Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y. M., Buso, N., and Lopez, R. (2015) The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res. 43*, W580−W584.

(44) Flynn, R. A., Zhang, Q. C., Spitale, R. C., Lee, B., Mumbach, M. R., and Chang, H. Y. (2016) Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nat. Protoc. 11*, 273−290.

(45) Hajdin, C. E., Bellaousov, S., Huggins, W., Leonard, C. W., Mathews, D. H., and Weeks, K. M. (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U. S. A. 110*, 5498−5503.

(46) Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Müller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinf. 3*, 2.

(47) Aken, B. L., Achuthan, P., Akanni, W., Amode, M. R., Bernsdorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., et al. (2017) Ensembl 2017. *Nucleic Acids Res. 45*, D635−D642.

(48) Kondo, Y., Oubridge, C., van Roon, A.-M. M., and Nagai, K. (2015) Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *eLife 4*, e04986.

(49) Novoa, E. M., Beaudoin, J.-D., Giraldez, A. J., Mattick, J. S., and Kellis, M. (2017) Best practices for genome-wide RNA structure analysis: combination of mutational profiles and drop-off information. *bioRxiv*, DOI: 10.1101/176883.

(50) Sexton, A. N., Wang, P. Y., Rutenberg-Schoenberg, M., and Simon, M. D. (2017) Interpreting Reverse Transcriptase Termination and Mutation Events for Greater Insight into the Chemical Probing of RNA. *Biochemistry 56*, 4713−4721.

(51) Zubradt, M., Gupta, P., Persad, S., Lambowitz, A. M., Weissman, J. S., and Rouskin, S. (2017) DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods 14*, 75−82.

(52) Reuter, J. S., and Mathews, D. H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinf. 11*, 129.

(53) Mortimer, S. A., and Weeks, K. M. (2007) A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry. *J. Am. Chem. Soc. 129*, 4144−4145.

(54) Chan, D., Feng, C., Zhen, Y., Flynn, R. A., and Spitale, R. C. (2017) Comparative Analysis Reveals Furoyl in Vivo Selective Hydroxyl Acylation Analyzed by Primer Extension Reagents Form Stable Ribosyl Ester Adducts. *Biochemistry 56*, 1811−1814.

(55) Steen, K.-A., Rice, G. M., and Weeks, K. M. (2012) Fingerprinting Noncanonical and Tertiary RNA Structures by Differential SHAPE Reactivity. *J. Am. Chem. Soc. 134*, 13160−13163.

(56) Smola, M. J., and Weeks, K. M. (2018) In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc. 13*, 1181−1195.

(57) Tyrrell, J., McGinnis, J. L., Weeks, K. M., and Pielak, G. J. (2013) The Cellular Environment Stabilizes Adenine Riboswitch RNA Structure. *Biochemistry 52*, 8777−8785.

(58) McGinnis, J. L., Liu, Q., Lavender, C. A., Devaraj, A., McClory, S. P., Fredrick, K., and Weeks, K. M. (2015) In-cell SHAPE reveals that free 30S ribosome subunits are in the inactive state. *Proc. Natl. Acad. Sci. U. S. A. 112*, 2425−2430.

(59) Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J. S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature 505*, 701−705.

(60) Deigan, K. E., Li, T. W., Mathews, D. H., and Weeks, K. M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U. S. A. 106*, 97−102.