Published in final edited form as:

Nat Chem Biol. 2013 January ; 9(1): 18–20. doi:10.1038/nchembio.1131.

RNA SHAPE analysis in living cells

Robert C. Spitale^{1,3}, Pete Crisalli^{2,3}, Ryan A. Flynn^{1,3}, Eduardo A. Torre¹, Eric T. Kool², and Howard Y. Chang¹

¹Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, CA 94305

²Department of Chemistry, Stanford University, Stanford, CA 94305

Abstract

RNA structure plays important roles in practically every facet of gene regulation, but the paucity of structural probes that function *in vivo* has limited current understanding. Here we design, synthesize, and demonstrate two novel chemical probes that enable selective 2'-hydroxyl acylation for accurate RNA structural analysis in living cells. RNA structure in embryonic stem cells and several other species is read out at single nucleotide resolution, revealing tertiary contacts and RNA-protein interactions.

Genomes are extensively transcribed to generate diverse coding and regulatory RNAs whose structure and function remain to be characterized.¹ Despite being composed of only four chemically similar nucleotides, RNA can base pair with itself and interact with other molecules to form complex secondary and tertiary structures. Predicting how RNAs fold and their corresponding functions are formidable challenges. Recently, *in vitro* RNA structure-probing experiments have improved the accuracy of secondary structure models, adding to our knowledge of RNA structural motifs;^{2,3} however, RNA structure *in vivo* is likely to be more complex, and in some cases fundamentally different than what is observed *in vitro*.⁴ RNA structure in cells is influenced by the rate of transcription, local solution conditions, the binding of small molecules, and interactions with numerous RNA-binding proteins.⁵ These observations hint that the physical state of RNA within the cell is crucial for its function, but current knowledge of RNA structures in cells is limited.

Several reagents have been developed to obtain structural maps of RNAs. Ribonucleases that are specific for single or double-stranded RNA can provide low-resolution measurements of RNA secondary structure.^{6,7} Alternatively, chemical mapping by direct RNA base methylation by dimethylsulfate (DMS), or cleavage of solvent exposed residues by hydroxyl radical or metal catalysis are currently the most rigorous methods for determining RNA secondary and tertiary structure.^{2,3} However, reagents such as DMS and kethoxal do not react with all four bases in RNA, and lead (II) cleavage has strong biases near sites of metal coordination that may confound the interpretation of secondary structure

Correspondence to: E.T.K. (kool@stanford.edu) and H.Y.C. (howchang@stanford.edu).

³These authors contributed equally.

Author Contributions:

R.C.S., E.T.K. and H.Y.C. conceived the project. R.C.S., P.C., R.A.F. and E.A.T. performed the experiments. R.C.S., P.C., R.A.F., E.T.K. and H.Y.C. analyzed the data. R.C.S. and H.Y.C. wrote the manuscript with input from all authors.

Competing Financial Interests:

The authors declare no competing financial interests.

Additional Information: Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to E.T.K or H.Y.C.

measurements. Such biases have limited the ability to obtain structural information for every nucleotide. While DMS has been used for RNA structure probing *in vivo*,⁸ other reagents cannot be used in intact mammalian cells. These observations point toward the need for a generalizable reagent that can be used in a wide range of organisms to accurately measure RNA structure inside living cells.

The 2'-hydroxyl group is a universal chemical feature in every RNA. Pioneering work by Weeks and colleagues has demonstrated the feasibility of a general method to measure the dynamics and solvent exposure of local RNA structure by probing the 2'-hydroxyl groups. Single-stranded or flexible RNA regions exhibit high 2'-hydroxyl reactivity, whereas RNA nucleotides engaged in base pairing or other interactions show lower reactivity. Selective 2'-hydroxyl acylation followed by primer extension, or SHAPE, has become the gold standard for monitoring the secondary structures of complex RNAs.^{9–11} An additional strength of SHAPE chemistry is that the reactivities of the 2'-hydroxy groups can be integrated to obtain pseudo-free energy terms^{3, 10}; these terms can then be applied to correct RNA secondary structure models, increasing their accuracy to ~95%¹¹. Thus, SHAPE reagents are powerful probes for measuring and predicting RNA secondary structure at single-nucleotide resolution.

SHAPE was recently used to interrogate the structural landscape of genomic RNA extracted from viral particles and inside intact virions.^{12,13} However, long incubation times (~50 min) and the comparably simple viral environment suggested to us that current SHAPE electrophiles may not be amenable to *in vivo* experiments in more complex cells. Indeed, we attempted to modify RNA in different species with NMIA, the canonical SHAPE electrophile. Despite numerous efforts, we could not detect any modification of 5S rRNA in these cells (see below). We surmised that this limited reactivity might be due to the high reactivity of NMIA (leading to a very short effective half-life in water), to its high degree of cross-reactivity with other nucleophiles in the cell, and to its limited solubility in aqueous solutions.

We sought to develop novel acylation electrophiles that are selectively reactive toward hydroxyl groups, soluble at high concentrations, and amenable to RNA modification inside living cells within a reasonable time frame. We screened several aromatic electrophiles (Supplementary Fig. 1), but none of these were able to fit all desired parameters. We designed and tested electrophiles with different predicted solubilities and leaving groups, and devised simple synthetic strategies that proceed in high yield (Supplementary Fig. 2), ultimately choosing 2-methylnicotinic acid imidazolide (NAI) and 2-methyl-3-furoic acid imidazolide (FAI) (Fig. 1a). These compounds are conveniently produced as 1:1 mixtures with imidazole in a DMSO stock solution by reaction of the carboxylic acids with carbonyldiimidazole. Both reagents display similar hydroxyl acylation specificities as NMIA (Fig. 1 c, d and Supplementary Fig. 3). The heteroatoms were included in the aromatic rings to increase solubility, and adjacent methyl groups tune reactivity by causing a twist to the carbonyl groups. The low-toxicity imidazole leaving groups were designed to modulate reactivity while retaining solubility. The hydrolysis rates of NAI ($t_{1/2}$ =33min) and FAI ($t_{1/2}$ =73min) are considerably lower than NMIA ($t_{1/2}$ =4min)⁹. RNA extraction or addition of betamercaptoethanol successfully halts the acylation (Supplementary Fig. 4) *in vitro* and *in vivo*; this step allows the experimenter to terminate the reaction at will.

First, we used NAI and FAI to probe the secondary structure of mouse embryonic stem cells 5S rRNA *in vitro*. We chose 5S rRNA because of its abundant nature, its highly characterized structure, and its ability to fold into a stable structure without the need for protein cofactors.¹⁴ We observed very similar quantitative patterns of 2'-hydroxyl acylation with NAI or FAI versus NMIA (R^2 =0.93) which all map to residues that are predicted to be

flexible (Fig. 1b, d and Supplementary Fig. 5, 10, 11). These data suggest that both NAI and FAI are suitable electrophiles for 2'-hydroxyl acylation on structured RNA molecules, yielding accurate structural information comparable to that obtained with existing probes.

Next we evaluated the ability of NAI and FAI to monitor RNA structure in live cells. Cultured mouse embryonic stem cells (ESC) were reacted with 13 mM NAI or FAI (the maximum solubility of NMIA) in aqueous buffer, but no 5S rRNA modification could be detected after one hour with the probes at this concentration. However, when the concentration of the probes was increased to above 20 mM, we observed positive signals for modification. Both NAI and FAI employed in the cellular experiments caused blocks in subsequent reverse transcription, which was suggestive of modification, while NMIA did not (Supplementary Fig. 6). In addition, NAI showed greater extent of modification than FAI, consistent with its higher reactivity toward ATP *in vitro*.

We focused on NAI for further structural probing due to its higher reactivity. Application of 100 mM NAI to murine ESCs resulted in 5S rRNA modification in as little as one minute with suitable signal-to-noise ratio. This signal begins to plateau by 15 minutes (Fig. 1e), indicating a reasonable timescale to study biological phenomena. Even after 30 minutes of NAI treatment, murine ESCs remained attached to tissue culture vessel, appeared morphologically normal and unstained by trypan blue (Supplementary Fig. 7). NAI is expected to acylate many cellular RNAs and other molecules; thus cytotoxicity after longer-term treatment needs to be evaluated with care. To test the ability of our reagent to modify lower abundant RNAs, we also determined the SHAPE pattern for two nuclear localized RNAs. We were able to detect significant RNA modification, which suggests that our reagent is able to enter the nucleus and react with lower abundant RNAs to give structural information (Supplementary Fig. 8). NAI also modifies 5S rRNA in cultured human cancer cells, *Drosophila* S2 cells, yeast, and *E. coli* (Supplementary Fig. 9), suggesting that it is a general cell-permeable probe of RNA structure.

To understand the pattern of 5S rRNA SHAPE in ESCs, we compared our data to the crystal structure of the 80S ribosome from yeast, which includes the 5S rRNA¹⁵. Yeast and mammalian 5S rRNA exhibit very high sequence similarity and functional domain architecture,^{16,17} as indicated by a CLUSTALW alignment score of 60.¹⁷ The crystal structure of the ribosome is validated by decades of molecular genetics and biochemical studies and likely represents a conformation that occurs *in vivo*. Overlaying our SHAPE data to the 5S crystal structure showed that practically all residues in flexible regions or not in canonical Watson-Crick base-pairs are modified, including singlestranded loops, unstable non-canonical base-pairs, and a single base flipped out of the helical duplex (Fig. 1f, g). These results indicate that NAI can probe RNA structure *in vivo* with high accuracy and single-nucleotide resolution.

Comparison of SHAPE profiles of 5S rRNA *in vivo* versus *in vitro* revealed key RNA-RNA and RNA-protein interactions that dock the 5S rRNA into the ribosome. Overall, the profiles looked similar, but a few key differences suggest differential interactions in the living system (Fig. 2a, b). Hereafter, residues in 5S are numbered per the mouse gene (*M. musculus*, *M.M.*); residues in other ribosomal subunits are numbered as in the yeast crystal structure (*S. cerevisiae*, *S.C.*). First, major differences between the *in vitro* and *in vivo* modification profiles were observed with residue *M.M.A49/M.M.A50*. Within the context of the crystal structure we noticed that the analogous residue *S.C.U50*, which sits near the nexus of Loop B and Helix III, is kinked to allow the docking of Loop C into the 28S rRNA (Resi. *S.C.C2684* and *S.C.U2683*). This conformation permits the residues *S.C.ARG218*, *S.C.LEU222*, *S.C.GLU221*, and *S.C.LYS224* of ribosomal protein L5 to be stacked against residues *S.C.A51* and *S.C.U50* (Fig. 2d). As a result, *S.C.U50* seems to be pushed out of the

helix, thus increasing its dynamic nature and exposing the 2'-OH for reactivity. Prior saturation mutagenesis showed that *S.C.A51* and *S.C.U50* make contact with ribosomal proteins *S.C.L11* and *S.C.L5* to form a critical structural link between the large and small ribosomal subunits and are essential for proper ribosome function and viability^{16,18}. Thus, NAI can read out alterations in the RNA tertiary structure as a result of sampling critical mature ribosome conformations.

Second, a three-nucleotide bridge that connects Helix II with Loop A showed slight differences (Fig. 2e, Supplementary Fig. 11). Within the crystal structure *S.C.C10*, *S.C.A11*, *S.C.U12*, and *S.C.G67* are engaged in a multi-nucleotide bridge.¹⁵ *S.C.G67* and ribosomal protein L5 interact with these residues and may stabilize them in lower-reactive conformations. *S.C.C10* also moves out of the helix to stack on *S.C.PHE20* from ribosomal protein L5. This conformation may stabilize and shield residues *S.C.C10* and *S.C.U12* from sampling the same reactive conformations as seen *in vitro*.

Third, the residues of *M.M.U84/M.M.A83*, which are in Loop D, were more reactive to NAI *in vitro*. Within the context of the 80S ribosome these residues are engaged in extensive hydrogen-bond contacts with residues *S.C.G1148* and *S.C.G1171* of 28S *S.C.rRNA*. *S.C.U86* is stacked upon *S.C.A1197* and is in an H-bonding contact with cobalt hexamine (Fig. 2f). Notably, mutations of these 5S residues in yeast result in gross defects in translational accuracy.¹⁶ These differences suggest that NAI is able to distinguish subtle dynamic differences that may be the result of protein interactions, yet can still identify residues that are unpaired and therefore more flexible in the context of the cell. These findings suggest *in vivo* versus *in vitro* SHAPE comparison as a powerful unbiased strategy to pinpoint key residues in ncRNA interaction and function.

Loop E of 5S RNA provided a prime example of the power of *in vivo* SHAPE analysis. In the context of the fully assembled ribosome, loop E adopts a unique bulge structure in yeast 5S RNA, but not in other species studied.^{19,20} The *S.C.* crystal structure shows C72 and C73 are pushed out of Loop E. These residues also have the highest bfactor, suggesting these residues are highly dynamic (Fig. 2g). We confirmed that loop E is accessible *in vivo* only in yeast, but not other species (Supplementary Fig. 9). Comparison of *in vitro* vs. *in vivo* modified yeast 5S rRNA revealed conserved residues with similar modification patterns as in the mouse 5S rRNA, including the key residue *S.C.U50* (*M.M.U49*). Further, residues with lower b-factors in the crystal structure were shown to be less reactive *in vivo*. Importantly, residues C72 and C73 displayed the largest differences in the cell, with a marked increase in reactivity (Fig. 2h, i). Moreover, residues with altered modification pattern *in vivo* are nearly always required for 5S function *in vivo* when mutated (Fig. 2i). Our analysis is the first comparison of 5S rRNA structure *in vitro* versus *in vivo*. Overall, our experiments establish that our acylation reagents are capable of modifying RNA *in vivo* and can sensitively read characteristics of RNA structure that are the result of unique conformations that RNA adopts in the cell, either due to changes in base-pairing characteristics or protein-RNA interactions.

Our results suggest an approach to directly probe RNA structure in living cells and assess dynamic changes in RNA structure in different cell states or in cells knocked out for any gene. Importantly, we demonstrate the utility of our chemical probe to read out specific and sensitive structural differences in RNA as a function of its tertiary structure and RNA-protein interactions. Currently, obtaining physiologically relevant RNA structural information requires much effort to reconstitute ribonucleoprotein complexes *in vitro*. As the catalog of non-coding RNA molecules²¹⁻²³ and functional motifs in coding transcripts^{24,25} continues to expand, the need to probe their structure in the cell will become increasingly important.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of the Chang lab, particularly Yue Wan and Grace Zheng, and R. Das for comments. Supported by NIH (R01 GM072705 to E.T.K. and R01-HG004361 to H.Y.C.), A.P. Giannini Foundation (R.C.S.), and Stanford MedScholars Program (R.A.F.). H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.

References

1. Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* 2001; 2:986–991. doi:10.1093/embo-reports/kve230 2/11/986 [pii]. [PubMed: 11713189]
2. Mathews DH, et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A.* 2004; 101:7287–7292. doi:10.1073/pnas.0401799101 0401799101 [pii]. [PubMed: 15123812]
3. Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A.* 2009; 106:97–102. doi:0806929106 [pii] 10.1073/pnas.0806929106. [PubMed: 19109441]
4. Mahen EM, Harger JW, Calderon EM, Fedor MJ. Kinetics and thermodynamics make different contributions to RNA folding in vitro and in yeast. *Mol Cell.* 2005; 19:27–37. doi:S1097-2765(05)01351-1 [pii] 10.1016/j.molcel.2005.05.025. [PubMed: 15989962]
5. Schroeder R, Grossberger R, Pichler A, Waldsich C. RNA folding in vivo. *Curr Opin Struct Biol.* 2002; 12:296–300. doi:S0959440X02003251 [pii]. [PubMed: 12127447]
6. Trang P, Hsu AW, Liu F. Nuclease footprint analyses of the interactions between RNase P ribozyme and a model mRNA substrate. *Nucleic Acids Res.* 1999; 27:4590–4597. doi:gkc674 [pii]. [PubMed: 10556315]
7. Tijerina P, Mohr S, Russell R. DMS footprinting of structured RNAs and RNA-protein complexes. *Nat Protoc.* 2007; 2:2608–2623. doi:nprot.2007.380 [pii] 10.1038/nprot.2007.380. [PubMed: 17948004]
8. Luo D, Condon C, Grunberg-Manago M, Putzer H. In vitro and in vivo secondary structure probing of the thrS leader in *Bacillus subtilis*. *Nucleic Acids Res.* 1998; 26:5379–5387. doi:gkb834 [pii]. [PubMed: 9826762]
9. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc.* 2005; 127:4223–4231. doi:10.1021/ja043822v. [PubMed: 15783204]
10. Gherghe CM, Shajani Z, Wilkinson KA, Varani G, Weeks KM. Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA. *J Am Chem Soc.* 2008; 130:12244–12245. doi:10.1021/ja804541s. [PubMed: 18710236]
11. Weeks KM, Mauger DM. Exploring RNA structural codes with SHAPE chemistry. *Acc Chem Res.* 2011; 44:1280–1291. doi:10.1021/ar200051h. [PubMed: 21615079]
12. Wilkinson KA, et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* 2008; 6:e96. doi:07-PLBI-RA-2987 [pii] 10.1371/journal.pbio.0060096. [PubMed: 18447581]
13. Watts JM, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature.* 2009; 460:711–716. doi:nature08237 [pii] 10.1038/nature08237. [PubMed: 19661910]
14. Kladwang W, VanLang CC, Cordero P, Das R. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat Chem.* 2011; 3:954–962. doi:nchem.1176 [pii] 10.1038/nchem.1176. [PubMed: 22109276]
15. Ben-Shem A, et al. The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science.* 2011; 334:1524–1529. doi:science.1212642 [pii] 10.1126/science.1212642. [PubMed: 22096102]

16. Smith MW, Meskauskas A, Wang P, Sergiev PV, Dinman JD. Saturation mutagenesis of 5S rRNA in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 2001; 21:8264–8275. doi:10.1128/MCB.21.24.8264-8275.2001. [PubMed: 11713264]
17. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994; 22:4673–4680. [PubMed: 7984417]
18. Yusupov MM, et al. Crystal structure of the ribosome at 5.5 Å resolution. *Science*. 2001; 292:883–896. doi:10.1126/science.1060089 1060089 [pii]. [PubMed: 11283358]
19. Brunel C, Romby P, Westhof E, Ehresmann C, Ehresmann B. Three-dimensional model of *Escherichia coli* ribosomal 5 S RNA as deduced from structure probing in solution and computer modeling. *J Mol Biol*. 1991; 221:293–308. doi:0022-2836(91)80220-O [pii]. [PubMed: 1717695]
20. Leontis NB, Westhof E. The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA*. 1998; 4:1134–1153. [PubMed: 9740131]
21. Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS. Non-coding RNAs: regulators of disease. *J Pathol*. 2010; 220:126–139. doi:10.1002/path.2638. [PubMed: 19882673]
22. Spitale RC, Tsai MC, Chang HY. RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics*. 2011; 6:539–543. doi:15221 [pii]. [PubMed: 21393997]
23. Breaker RR. Prospects for riboswitch discovery and analysis. *Mol Cell*. 2011; 43:867–879. doi:S1097-2765(11)00645-9 [pii] 10.1016/j.molcel.2011.08.024. [PubMed: 21925376]
24. Goodarzi H, et al. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature*. 2012 doi:nature11013 [pii] 10.1038/nature11013.
25. Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*. 2009; 324:255–258. doi:324/5924/255 [pii] 10.1126/science.1170160. [PubMed: 19359587]

Figure 1. Design and experimental evaluation of NAI and FAI

(a) Structural comparison between NMIA, FAI, and NAI. (b) Chemical mechanism of 2'-O-acylation of RNA. (c) Time course of ATP modification by NAI and FAI. (d) Correlation of 2'-RNA modification of 5S rRNA *in vitro*. (e) Denaturing gel electrophoresis of NAI modification of 5S rRNA in cells. (f) Secondary structure mapping of NAI modification of 5S rRNA in cells. (g) Three-dimensional model of *S.C.* 5S rRNA with modifications superimposed onto the structure (PDB 3U5H).

Figure 2. 5S rRNA has different modification patterns in cells

(a) Denaturing gel electrophoretic analysis of NAI modification of 5S rRNA in *M. musculus* Embryonic Stem cells and *in vitro*. (b) Normalized Differential profile of *M. musculus* Embryonic Stem cell 5S rRNA. (c) Three-dimensional model of *S.C.* 5SrRNA with differential modifications superimposed onto the structure (PDB 3U5H). (d) Closeup view of hyperreactivity of M.M. A49 and S.C. U50. U50 is near the kink of helix II and helix III, This conformation, which allows Loop C to interact with 28S rRNA promotes the ejection of S.C. U50 2'-OH thus rendering it more dynamic and reactive to NAI modification. (e) Zoomed-in view of a three-nucleotide bridge that joins loop A to helix II. This conformation

results in docking of A11, hiding its 2'-OH through a hydrogen bond with A13-OP2. In addition C10 is forced into a stacking interaction with PHE20, which exposes the 2'-OH, potentially increasing its reactivity with NAI. **(f)** Close up view of differential modification of Helix IV. The ribosome crystal structure shows many of the residues that are hypomodified in cells are engaged in extensive interactions with ribosomal RNA. Shown is the interaction of U86 with 28S rRNA. Such extensive bonding and stacking may be stabilizing the internucleotide linkages of these residues, limiting their acylation reactivity. **(g)** Three-dimensional model of *S. C.* 5SrRNA colored by b-factor (PDB 3U5H). **(h)** Denaturing gel electrophoretic analysis of NAI modification of 5S rRNA in *S. cerevisiae* cells and *in vitro*. **(i)** Normalized differential profile of *S. cerevisiae* 5S rRNA. Residues are labeled for their importance in 5S rRNA function as noted in reference 31.