

## SEISMICgraph: a web-based tool for RNA structure data visualization

### AUTHORS

Federico Fuchs Wightman<sup>1</sup>, Grant Yang<sup>1</sup>, Yves J. Martin des Taillades<sup>2</sup>, Casper L'Esperance-Kerckhoff<sup>1</sup>, Scott Grote<sup>1</sup>, Matthew F. Allan<sup>1</sup>, Daniel Herschlag<sup>2</sup>, Silvi Rouskin<sup>1</sup>, \* and Lauren D. Hagler<sup>2</sup>, \*.

<sup>1</sup> Department of Microbiology, Harvard School of Medicine, Boston, Massachusetts, 02115, United States.

<sup>2</sup> Department of Biochemistry, Stanford University, Stanford, California, 94305, United States.

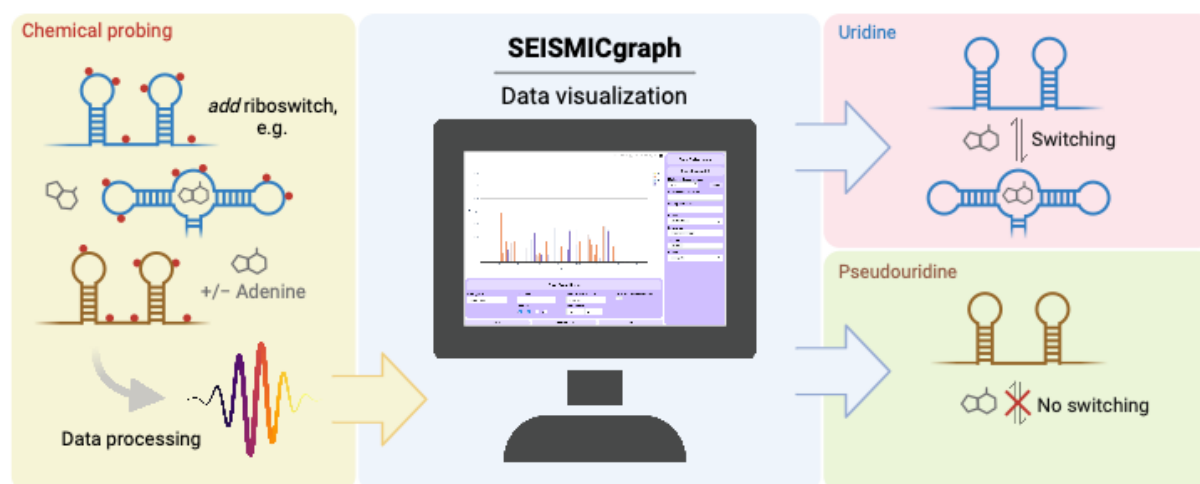
\* To whom correspondence should be addressed. Tel: +1 (617) 432-1935; Email: [silvi@hms.harvard.edu](mailto:silvi@hms.harvard.edu).

\* Correspondence may also be addressed to Tel: +1 (650) 723-6161; Email: [ldhagler@stanford.edu](mailto:ldhagler@stanford.edu).

†Author notes: The first three authors should be regarded as Joint First Authors.

Present Address: Silvi Rouskin, Department of Microbiology, Harvard School of Medicine, Boston, Massachusetts, 02115, United States. Lauren D. Hagler, Department of Biochemistry, Stanford University, Stanford, California, 94305, United States.

### GRAPHICAL ABSTRACT



## **ABSTRACT**

In recent years, RNA has been increasingly recognized for its essential roles in biology, functioning not only as a carrier of genetic information but also as a dynamic regulator of gene expression through its interactions with other RNAs, proteins, and itself. Advances in chemical probing techniques have significantly enhanced our ability to identify RNA secondary structures and understand their regulatory roles. These developments, alongside improvements in experimental design and data processing, have greatly increased the resolution and throughput of structural analyses. Here, we introduce SEISMICgraph, a web-based tool designed to support RNA structure research by offering data visualization and analysis capabilities for a variety of chemical probing modalities. SEISMICgraph enables simultaneous comparison of data across different sequences and experimental conditions through a user-friendly interface that requires no programming expertise. We demonstrate its utility by investigating known and putative riboswitches and exploring how RNA modifications influence their structure and binding. SEISMICgraph's ability to rapidly visualize adenine-dependent structural changes and assess the impact of pseudouridylation on these transitions provides novel insights and establishes a roadmap for numerous future applications.

## **INTRODUCTION**

Since its initial discovery as an intermediary of genetic information, RNA has been recognized for its diverse regulatory roles, including the control of transcription, translation, splice site selection, and RNA degradation (Allan et al., 2023; Wan et al., 2011). These functions are achieved through RNA's ability to bind proteins and fold into complex structures. The discovery of novel RNA interactions and their biological function has driven and been supported by the development of numerous techniques, each with specific advantages and limitations (Jones, 2016; Ma et al., 2022; Spitale and Incarnato, 2023; Turner and Mathews, 2016; Zhang et al., 2022).

Chemical probing methods provide detailed insights into RNA base-pairing, folding, and interactions with biomolecules such as proteins and ligands at single-nucleotide resolution. These techniques use small molecules such as dimethyl sulfate (DMS) and SHAPE reagents (e.g., 1-methyl-7-nitroisatoic anhydride (1M7)) to specifically and covalently modify RNA at chemically accessible sites (Ehresmann et al., 1987; Merino et al., 2005; Peattie and Gilbert,

1980; Soukup and Breaker, 1999). The differential chemical reactivity between nucleotides is used to distinguish between folded and unfolded regions of a given RNA structure (Douds et al., 2024; Siegfried et al., 2014; Zubradt et al., 2017). DMS methylates the Watson-Crick-Franklin face of solvent-exposed and unpaired A and C bases, while SHAPE reagents selectively acylate the 2'-hydroxyl group of bases in non-based paired regions that can attain the reactive configuration (Lucks et al., 2011; Rouskin et al., 2014; Watts et al., 2009). Both DMS and SHAPE techniques can be combined with mutational profiling (MaP), where chemically modified residues are converted into mutations during reverse transcription to enable the generation of high-resolution structural maps across RNA molecules (Homan et al., 2014; Siegfried et al., 2014; Zubradt et al., 2017). These methods are highly adaptable and can be employed to study how environmental conditions impact RNA interactions, allowing comparisons between *in vitro* and cellular contexts and providing new insights into the relationship between RNA sequence, structure, and function.

As chemical probing methods have become more widely adopted, several computational tools have been developed to analyze the resulting sequencing data. These include DREEM (Tomezsko et al., 2020), RNA Framework (Incarnato et al., 2018), and ShapeMapper 2 (Busan and Weeks, 2018), and the more recent SEISMIC-RNA (Allan et al., 2024), which builds upon DREEM. These tools facilitate the alignment of sequencing reads to reference sequences, the identification and quantification of mutations, and the generation of mutation profiles for each reference. The data generated through these analyses can be further used to test hypothesis-driven models and draw biophysical or biological conclusions, including the generation of secondary structure models and the assessment of the effects of mutations or environmental changes. Despite their utility, these tools typically require users to have at least basic programming skills to perform quality control assessments, visualize data, and compare samples, posing challenges for those less familiar with bioinformatics.

To overcome these limitations and allow the broader community to effectively use these approaches, we introduce SEISMICgraph, a web-based platform. SEISMICgraph is designed to streamline and standardize the visualization and analysis of chemical probing data. It was validated using datasets from DMS-MaPseq experiments involving well-characterized RNA structures. The application provides an accessible framework for generating plots and analyzing raw chemical probing data from RNA libraries. SEISMICgraph enables researchers to

produce standard visualizations of chemical probing data, making the interpretation of results more accessible and interpretable, while also facilitating broad use in the scientific community. SEISMICgraph is available at <https://seismicrna.org>.

## MATERIALS AND METHODS

### Library Design

The library used for the initial analysis contained 48 RNA structures and uses a common primer, flank, and barcoding strategy (Figure 2A). Library amplification, in vitro transcription, DMS-MaPseq, and data processing was performed as outlined below with minor changes. Specifically, DMS reactions varied in magnesium ( $\text{MgCl}_2$ ) concentration (0, 1, and 10 mM), while keeping the RNA input (10  $\mu\text{g}$ ), buffer (150 mM potassium cacodylate, pH 7.4), DMS concentration (5% v/v), reaction time (5 min), and temperature (37 °C) constant. One reference (*add* aptamer domain) and section (variable region, bases 40-112) were chosen from the library to help visualize the plot types in Figure 2.

A second library had the same design strategy (Figure 3A). We retrieved from the RiboswitchDB database (Mukherjee et al., 2019) 29 predicted  $\text{Mg}^{2+}$ -dependent riboswitches and 30 proposed to respond to purines. We included the *add* riboswitch from *Vibrio vulnificus* as a positive control. The total library encompasses 81 sequences, including above mentioned putative riboswitches and other well-known structures such as the HIV FSE (see Supplementary Table 1). To enable pooled chemical probing with DMS-MaPseq, we designed the library to include the following: (1) Common forward and reverse primer sequences (forward primer: 5'-TTAAACCGCCAACATACC -3'; reverse primer: 5'-TCGAAAGGAACGAGTAGCG -3'). (2) Variable length flanking sequences to make all the constructs of equal length. Flanking sequences were generated by randomly shuffling the order of Cs and Us. Sequences were filtered for consecutive stretches of Cs or Us longer than 3 nucleotides and were regenerated for these cases. (3) Unique barcodes were added to each sequence for later demultiplexing (Hamming distance > 3), which was required because of similarities between several library sequences and to identify parent sequences after DMS modification (Figure 2A and 3A). In combination, these features increase sequencing read depth and quality, by decreasing RT and/or PCR biases (thereby giving more uniform reads of all library members) and by unambiguously assigning very similar sequences in the library pool. The complete sequences of the 81-

member library are given in Supplementary Table 1. The DNA library was obtained as an oligo pool from Integrated DNA Technologies (IDT).

### **DMS-MaPseq**

The DNA library was first amplified via PCR, adding a T7 promoter sequence at the forward primer (5'- TAATACGACTCACTATAG -3'), using the Q5<sup>®</sup> High-Fidelity 2X Master Mix (NEB, Cat #M0492L) and the following program: 98 °C for 2 minutes; 6 cycles of 98 °C for 10 seconds, 63 °C for 30 seconds and 72 °C for 30 seconds; and a 2 minute final extension at 72 °C and final hold at 4 °C (and for long term storage, stored at -20 °C). The annealing temperature was calculated using the NEB's Tm calculator (<https://tmcaculator.neb.com/#!/main>). The reaction was carried in a final volume of 50 µl, adding 2.5 µl of each primer (10 µM) and 1 µl of the oligo pool (0.1 pmol/µl). The cycle number was optimized by screening increasing numbers of cycles until non-specific amplification was observed. To obtain the necessary yield of PCR product for the downstream applications, several reactions were run in parallel, pooled and purified using a Zymo DNA Clean & Concentrator kit (Cat. D4004).

*In vitro* transcription was performed using the T7 MEGAscript kit (Thermo Fisher, Cat. AM1354), starting with 200 ng of the amplified library. 20 µl reactions were incubated for 2-3 hours, followed by DNase treatment for 15 minutes. For pseudouridine experiments, UTP was replaced with 75 mM ΨTP or m1ΨTP. A yield of 30-50 µg was obtained per reaction. The RNAs were purified using the Zymo RNA Large Sample Clean-Up Kit (Cat. R1017). For chemical probing, 5-10 µg of the purified RNAs were diluted in 10 µl of nuclease-free water, denatured for 1 minute at 95 °C, and placed on ice for 2 minutes. The RNA samples were refolded in 300 mM sodium cacodylate, pH 7.2, 6 mM MgCl<sub>2</sub>, and optionally 5 mM adenine (or an equivalent volume of DMSO as a control). Refolding buffer (85 µl) was added to the RNAs, which were then incubated at 37 °C for 20 minutes in a thermomixer with a heated lid. Afterward, 5 µl of DMS (dimethyl sulfate) was added to the RNA solution and the reaction was allowed to proceed at 37 °C for 5 minutes with 800 rpm shaking in a thermomixer. The reaction was quenched by adding 60 µl of β-mercaptoethanol. The DMS-treated RNAs were purified using the Zymo RNA Clean & Concentrator-5 Kit (Cat. R1013). As recoveries can be as low as 10-20%, increasing the initial quantity of RNA per reaction is sometimes helpful.

The purified RNAs were reverse transcribed using the Induro Reverse Transcriptase (NEB #M0681), as described previously with minor modifications (Romero-Agosto et al., 2023). The reaction was performed as follows: 500-1000 ng of RNA, 1  $\mu$ l of the reverse primer at 10  $\mu$ M (sequence: 5' TCGAAAGGAACGAGTAGCG -3'), 1  $\mu$ l of 10 mM dNTPs, 0.2  $\mu$ l of RNase Inhibitor, Murine (40 U/ $\mu$ l), 4  $\mu$ l of the 5X Induro RT buffer, 1  $\mu$ l of Induro Reverse Transcriptase (200 U/ $\mu$ l), and nuclease-free water up to 20  $\mu$ l; the program was set for 30 minutes at 60  $^{\circ}$ C, followed by the addition of 1  $\mu$ l of 4 M NaOH for 3 minutes at 95  $^{\circ}$ C. The cDNAs were purified using the Zymo Oligo Clean & Concentrator kit (Cat# D4060) and gave a yield of approximately 250-500 ng. We then amplified the constructs via PCR using the Q5<sup>®</sup> High-Fidelity 2X Master Mix (NEB, Cat #M0492L) in a final volume of 50  $\mu$ l and the following program: 98  $^{\circ}$ C for 2 minutes, 15 cycles of 98  $^{\circ}$ C for 10 seconds, 63  $^{\circ}$ C for 30 seconds and 72  $^{\circ}$ C for 30 seconds, with a 2-minute final extension at 72  $^{\circ}$ C and a final hold at 4  $^{\circ}$ C.

Samples were indexed using NEBNext Ultra II DNA Library Prep Kit for Illumina (#E7645S/L), following the manufacturer's instructions, and sequenced on an Illumina NextSeq 1000 using a P1 (2x150 read) cartridge (Cat# 20050264).

### **SEISMIC-RNA Data Processing**

The first round of demultiplexing (separating the reads corresponding to each sample by their library index) was automatically performed by the Illumina platform BaseSpace (BCL-Convert 2.4.0) upon completion of the sequencing run. The FASTQ files were processed using SEISMIC-RNA (version 0.14.1) (Allan et al., 2024). Briefly, SEISMIC-RNA demultiplexes based on the barcode of a given reference, aligns reads to the known sequence, and quantifies the number of mutations on each read at each position. The function *wf* (for workflow) was executed including the option *-x* for 2 separate FASTQ files of paired-end reads; *--demult-on* for the second round of demultiplexing (a table with the barcode positions must be provided as well as a FASTA with the references, see Supplementary Table 1); *--sections-file* to annotate parts of the library sequence, including primers, flanking sequence, region of interest, and barcode; and *--export* for exporting a *json* file that can be provided to SEISMICgraph. Data were analyzed in SEISMICgraph. The filtered data for each plot were downloaded from SEISMICgraph and replotted in GraphPad Prism 10 to increase the resolution for publication.

## RESULTS

### SEISMICgraph: a web-based analysis tool

We developed SEISMICgraph (Figure 1 and Supplementary Figure 1) as a web-based application to address the need for rapid and simple visualization of chemical probing data, without requiring programming skills. SEISMICgraph consists of three main parts: (A) A front end user interface through which the user uploads data, makes selections, and requests plots; (B) A backend server that handles these requests and responds to them; (C) A python package (seismic-graph) that runs on the backend, processing uploads and generating plots.

SEISMICgraph uses data from chemical probing experiments to generate relevant plots. Users can upload data from multiple experiments at once, up to 32MB total. For datasets pre-processed with RNA Framework or SHAPMapper 2, SEISMICgraph automatically converts into SEISMIC-RNA format on upload. For details on how the data is structured, transformed, and organized, refer to Supplementary Information section "Input data structure".

*User Interface.* The user interface is separated into three main sections, (Figure 1B) Data Selections, Plot Selections, and the Plot Viewer.

In the Data Selections panel, users choose a subset of the uploaded data to plot. Certain plot types require only one row to be selected at a time, while others require two or more rows to be selected. Additional filtering options are available depending on the data selected, with more detailed information provided in the Supplementary Information section "Data Selection Options".

Under Plot Selections, the user can select from several plots provided by SEISMICgraph. The Supplementary Information lists the plots that can be generated, and Figure 2 highlights examples of each plot type. The full details for additional plotting options are given in the Supplementary Information section "Plot Selection Options".

Plot Viewer displays the rendered plot. The plot can be downloaded as an image or interactive HTML page. Additionally, the data can be downloaded as a CSV file for further analysis or to plot in ways not available in SEISMICgraph.

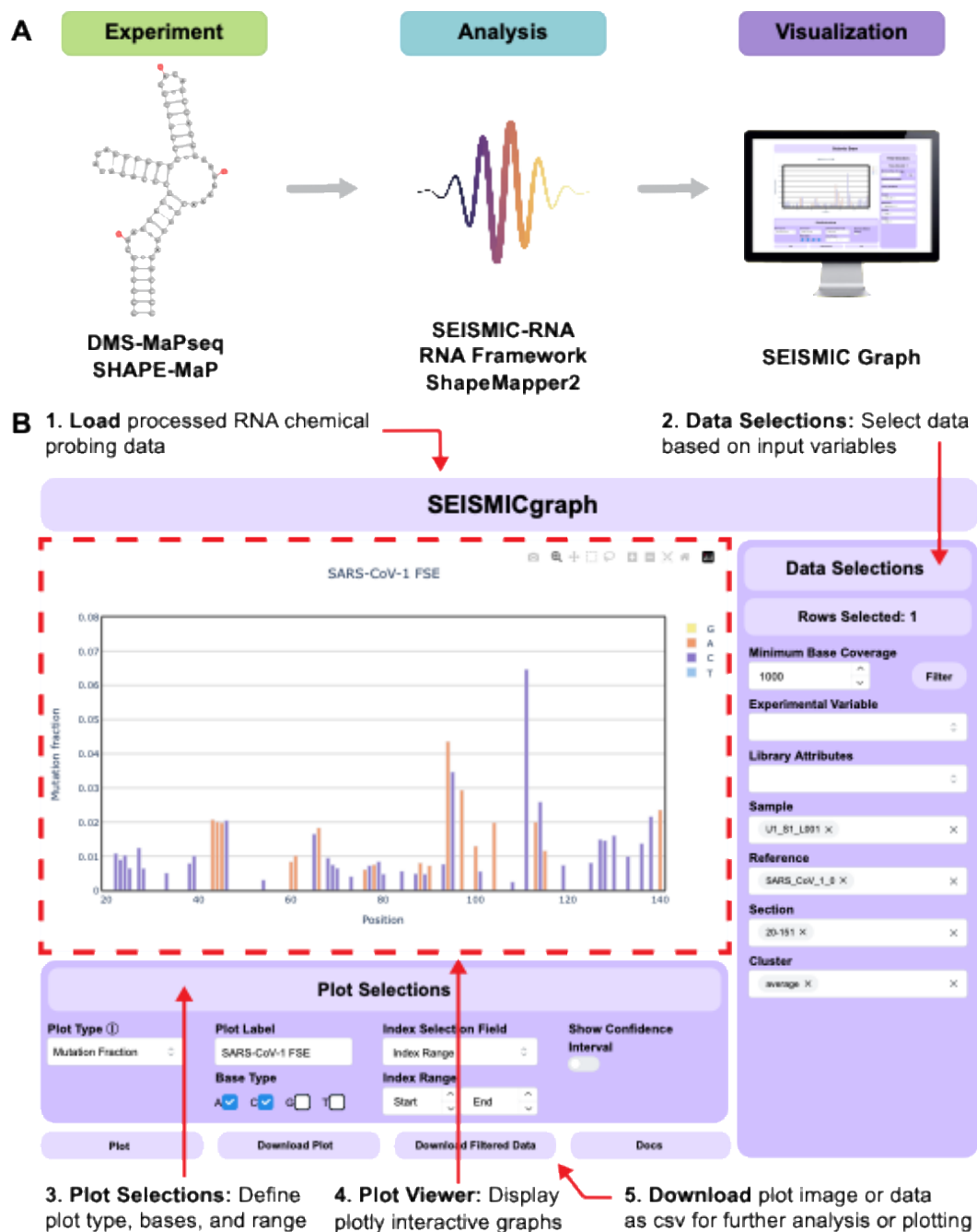


Figure 1. SEISMICgraph. (A) To study RNA secondary structures, DMS-MaPseq or SHAPE-MaP is performed by chemically modifying unstructured regions of RNA depicted here with red circles. Modifications are converted to mutations during RT-PCR and read out by next generation sequencing. The sequencing data from these experiments is then analyzed with SEISMIC-RNA, RNA Framework, or ShapeMapper 2 before being visualized with SEISMICgraph (see also Supplementary Figure 1). (B) Overview of the SEISMICgraph interface.



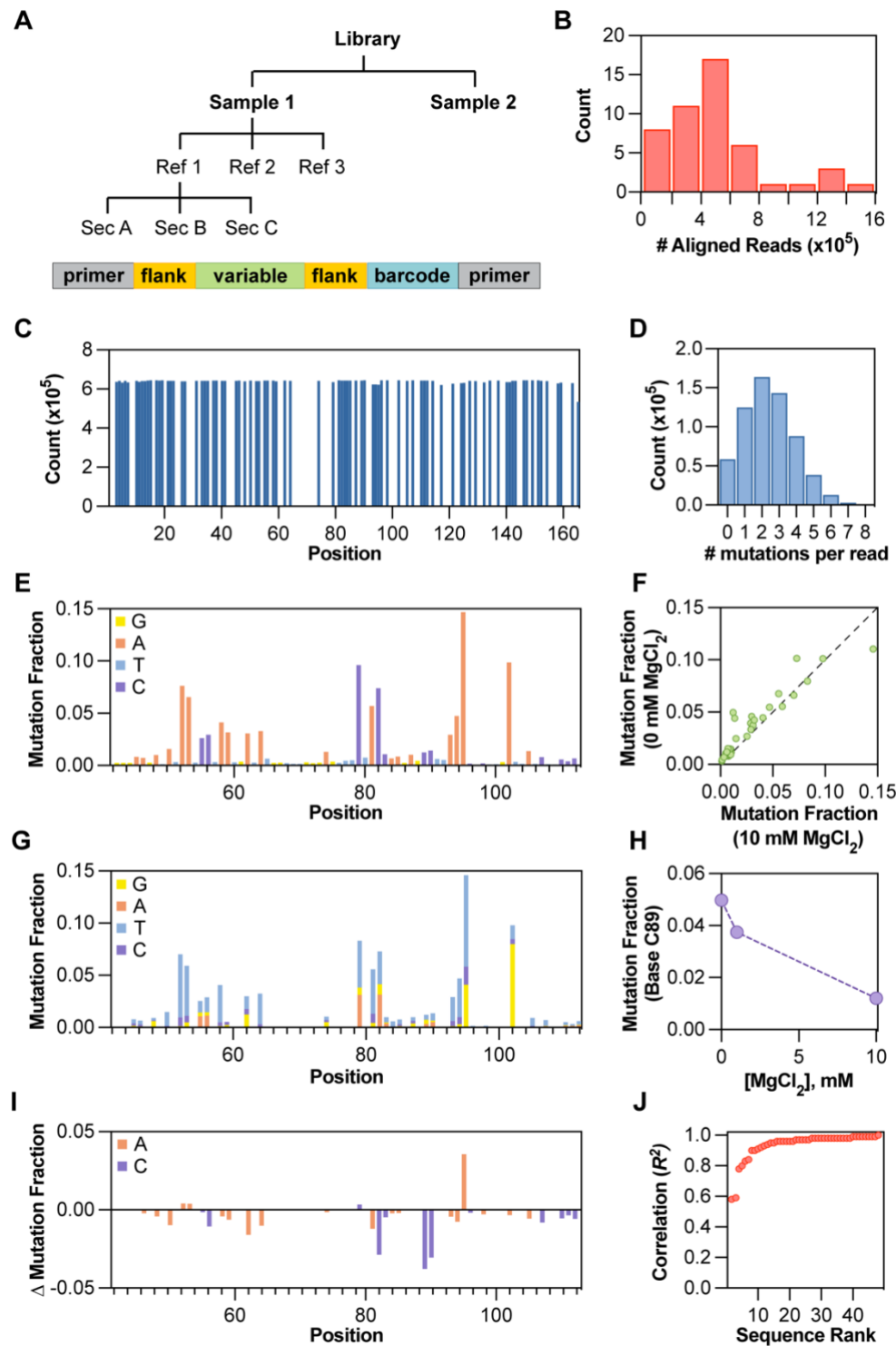


Figure 2. Examples of plot types available in SEISMICgraph. (A) Hierarchical data structure used in SEISMICgraph. For this library, each sample was separated into its reference sequences, which in turn were divided into sections. The references were structured to have common primers for library-specific amplification, a reference-specific barcode, variable-length flanking sequences to ensure all library members are of the same size, and a variable region that contains the RNA sequence of interest (see Methods and Supplementary Information). (B) #

**Aligned reads / Reference as Freq. Dist.** for one sample of a designed library of known RNA structures. (C-I) Example plots for one reference in the library where the variable region of interest is the *add* riboswitch (*Vibrio vulnificus*) aptamer domain. (C) **Base coverage** for entire amplicon (showing only As and Cs, section = full). (D) **Mutation per Read per Reference** shows the frequency distribution of mutations per read in bins. (E) **Mutation Fraction** for the variable section. Bases are colored by their identity in the reference sequence where A is orange, C is purple, T is light blue, and G is yellow. (F) **Compare Mutation Profiles** for two samples with no MgCl<sub>2</sub> or 10 mM MgCl<sub>2</sub>. Each green circle represents one reactive A or C base ( $N = 33$ ; Pearson  $R = 0.93$ , RMSE = 0.011). The black dashed line represents the line of identity predicted for the null model of no difference between the samples ( $x = y$ ). (G) **Mutation Fraction Identity** for the variable section. Each stacked bar represents the fraction mutated to that base, where orange represents bases mutated to A, purple represents bases mutated to C, light blue represents bases mutated to T, and yellow represents bases mutated to G. (H) **Experimental Variable Across Samples**. The mutation fraction from 3 samples (0, 1, and 10 mM MgCl<sub>2</sub>) are shown for one base (C at position 89). (I) **Mutation Fraction Delta** comparing the variable section of two samples (0 and 10 mM MgCl<sub>2</sub>) for only A and C bases. The change in mutation fraction here is calculated as sample 1 (10 mM MgCl<sub>2</sub>) – sample 2 (0 mM MgCl<sub>2</sub>). Bases are colored by identity in the reference where A is orange and C is purple. (J) **Correlation by Refs. Between Samples**. The Pearson correlations ( $R^2$ ) for all the members of the library were calculated by comparing the 0 vs. 10 mM MgCl<sub>2</sub> samples, where each red circle is an individual reference sequence.

## SEISMICgraph Identifies Riboswitches and Pseudouridylation's Impact on Function

We used SEISMICgraph to analyze DMS data for a designed library of RNA sequences to evaluate several predicted riboswitches, characterize several RNAs with well-established structures, and determine the effects of base modifications on the structure and function of these RNAs. Our library consisted of 81 RNAs, categorized as follows: 30 computationally-predicted purine riboswitches from various prokaryotic species (Mukherjee et al., 2019), 29 predicted  $Mg^{2+}$  riboswitches, and 22 other RNAs with well-characterized structures (Figure 3A). A complete list of sequences and their identities is provided in the Supplementary Table 1. Additionally, we provide detailed reports for all data and constructs in the Supplementary Dataset. We strongly encourage practitioners of chemical modification experiments to use and publish analogous summaries to make these studies more transparent and accessible to the scientific community.

*Identification of Purine-Responsive Riboswitches.* The designed library was transcribed *in vitro* and folded in the presence or absence of adenine. The samples were then subjected to DMS treatment followed by library-specific RT-PCR (Tomezsko et al., 2020) to detect conformational changes due to ligand binding (Figure 3A-C).

As a positive control, we first evaluated the adenosine deaminase (*add*) riboswitch from *Vibrio vulnificus*, which is known to alter its conformation upon adenine binding (Tian et al., 2018; Tomezsko et al., 2020). We demonstrated reproducibility in the DMS signal through two technical replicates, as shown in the **Mutation Fraction** plots of those replicates (Figure 3D and Supplementary Dataset "U1\_Vibrio\_vulnificus\_0 and U2\_Vibrio\_vulnificus\_0") and the **Compare Mutation Profiles** scatterplot (Figure 3E, Pearson  $R^2 = 0.98$  and Supplementary Figure 2B, Pearson  $R^2 = 0.99$ ). In contrast, a comparison of the *add* riboswitch with and without adenine revealed differences, indicated by a much lower Pearson correlation coefficient relative to the replicates (Pearson  $R^2 = 0.22$ ; Figure 3F). These differences were also apparent from inspection of the mutation fraction across the sequence (Figure 3D compared to Supplementary Figure 2A) and the difference plot between these conditions (Figure 3G).

The ability to download data from SEISMICgraph for further analysis allows users to closely examine single nucleotides changes under different conditions. For the *add* riboswitch, we identified individual bases within the aptamer binding domain that substantially decreased in

DMS reactivity when adenine was present (Figure 3H, I). The bases that were protected from DMS methylation upon ligand binding align with previous studies (Tian et al., 2018) and the hydrogen-bonding interactions enforced by adenine binding in the X-ray crystal structure of the ligand-bound state (Serganov et al., 2004). Such detailed analyses can enhance our understanding of the structural transformations induced by ligand binding. Future experiments can expand this approach to study the effect of other metabolites as well as studying these riboswitches in cells.

Next, we analyzed our entire RNA library using SEISMICgraph to compare the DMS reactivity in the presence and absence of adenine (Figure 3A). To provide an overall metric for adenine-dependent changes, we compared Pearson  $R^2$  values between replicates and between experimental conditions (Figure 3B, C). Three putative purine riboswitches, in addition to the established *add* riboswitch control, exhibited markedly lower  $R^2$  values ( $<0.6$ ) compared to all other sequences (Figure 3B, C). Thus, DMS-MaPseq chemical probing can be used to identify candidate riboswitches. In all three putative riboswitches, the adenine binding pocket is highly conserved (Supplementary Figure 4B-E), and a decrease in DMS reactivity is observed at these bases, akin to what is seen in the *add* riboswitch (Figure 3J-L). The putative riboswitch from *Shewanella piezotolerans* is found upstream of a gene encoding adenosine deaminase (Figure 3L), and the other two putative riboswitches (from *Thermotoga thermarum* and *Thermotoga lettingae*) are located upstream of operons involved in *de novo* IMP biosynthesis (Figure 3J, K). Thus, it is reasonable that the three identified riboswitches are authentic, although direct functional tests in cells are needed. Equally interesting are the RNAs with sequence homology (Supplementary Figure 3) that did not change conformation upon addition of adenine. These might adopt different, functional structures in cells or have evolved to bind different ligands or carry out different functions.

Surprisingly, one RNA lacking homology to adenine riboswitches also appeared to adopt an altered conformation in the presence of adenine. A short sequence from the *Arabidopsis thaliana* gene COOLAIR also appeared to be responsive to adenine (Figure 3M and Supplementary Figure 5). While the 5' domain of exon 1 from which this sequence was taken has several internal loops that might serve as a binding pocket for adenine (Yang et al., 2022), there is no observable pattern akin to that seen in the purine riboswitches. The shift in conformation may represent an accidental binding event (Peracchi et al., 1996) or an

underlying yet-to-be-identified riboswitch function in a eukaryotic RNA. Regardless, the ability of RNAs to associate with nucleobases and other small ligands has implications for evolution in an RNA world and for potential modern-day RNA functions (Benner et al., 1989; Peracchi et al., 1996).

The “other” RNA structures included in the library primarily serve as controls, allowing us to confirm that these structures can fold similarly to previous reports and to ensure that the addition of adenine does not affect the chemical probing experiment in the absence of binding. For example, the frameshifting stimulation element (FSE) from HIV has a mutational profile consistent with stem-loop formation in the absence of adenine (Supplementary Figure 6A, B). This pattern remains unchanged in presence of adenine (Supplementary Figure 6C, D).

Putative Mg-sensing riboswitch structures have a motif or junction that could act as the binding pocket for  $Mg^{2+}$ , similar to those found in known and putative purine riboswitches. We determined if sequences predicted to shift in response to  $Mg^{2+}$  also change conformations in the presence of adenine. Of the 15 sequences that passed quality control filters, none showed an observable conformational change (all Pearson  $R^2 > 0.6$ , Figure 3C) at the working conditions (300 mM Na-cacodylate, 6 mM  $MgCl_2$ ). For example, sequence NC\_0009792 appears to be folded in the absence of adenine (Supplementary Figure 7A, B), and its conformation does not change in the presence of adenine (Supplementary Figure 7C, D). Future experiments could measure the thermodynamics of folding by varying the  $Mg^{2+}$  concentration or temperature, using SEISMICgraph to analyze the results.

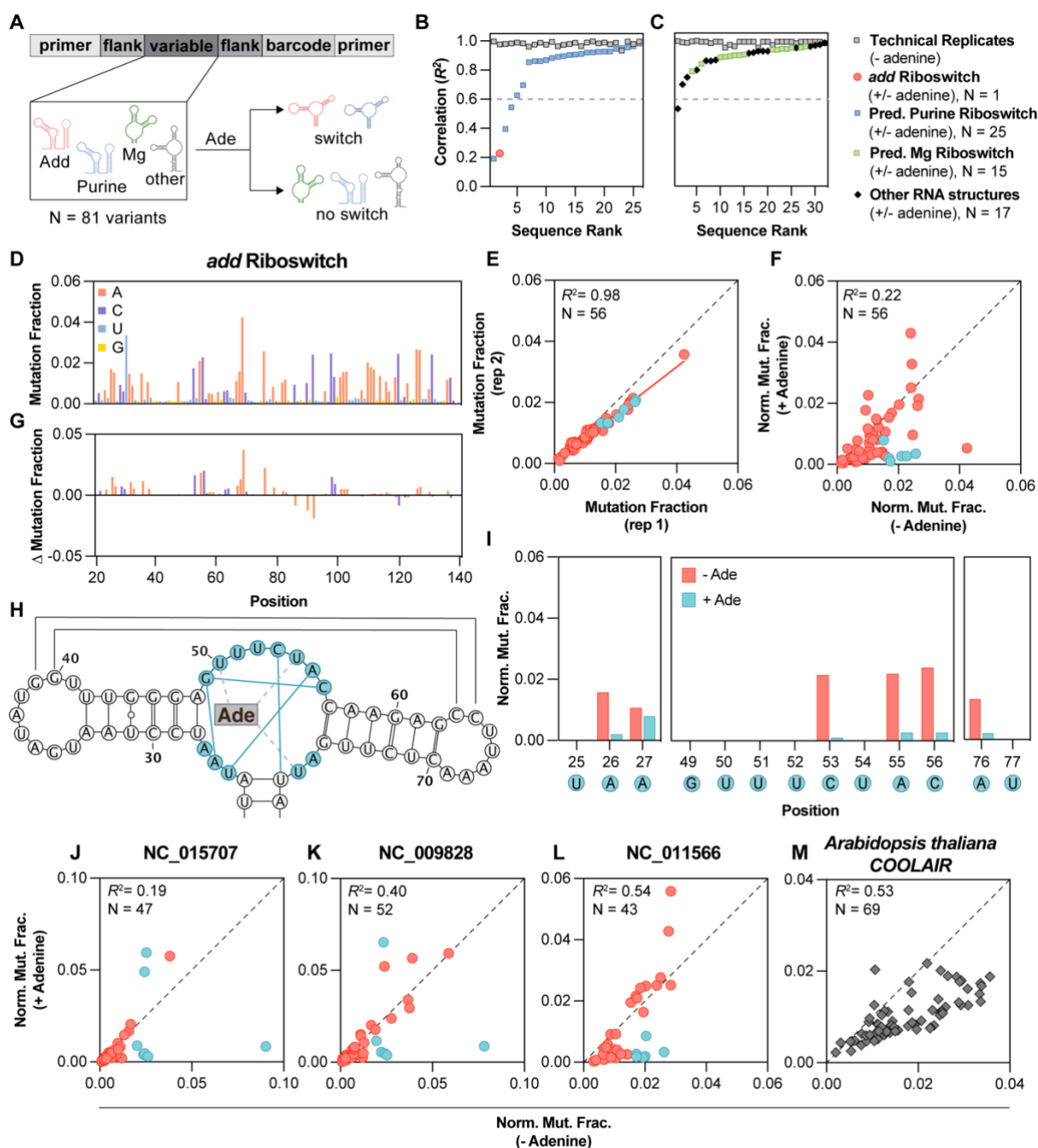


Figure 3. Experimental identification of adenine-responsive RNAs. (A) Library design for the riboswitch library. In this experiment the structure is probed with DMS in the presence or absence of adenine (Ade) to assess which structures undergo a conformation change. (B-C) Ranked correlation of references compared to each other in technical replicates (gray squares) or in samples with and without 5 mM adenine. (D) Mutation fraction at each position of the riboswitch section. (E) Comparison of mutation fraction at each position of two technical replicates where each red circle is an A or C base (N = 56). Turquoise circles depict A and C bases in the aptamer binding domain as in H. The Pearson correlation ( $R^2$ ) was calculated in

SEISMICgraph. The black dashed line represents the line of identity, and the red solid line represents the best-fit line from linear regression (slope = 0.77). (F) Comparison of the mutation fraction of samples in the presence and absence of 5 mM adenine. (G) Mutation fraction delta comparing samples in the presence and absence of 5 mM adenine, where the difference is calculated as sample 1 (with adenine) – sample 2 (without adenine). (H) Predicted secondary structure of the aptamer domain of the *add* riboswitch (Tian et al., 2018), where bases in the adenine binding site are depicted as turquoise circles. Hydrogen bond interactions enforced by binding are between bases in the binding site are shown with turquoise solid lines (Serganov et al., 2004). Hydrogen bonds between the binding site and adenine are shown as grey dashed lines (Serganov et al., 2004). (I) Comparison of bases in the binding site in the absence (red bars) and presence (turquoise bars) of 5 mM adenine. (J-L) Putative purine riboswitches that undergo conformational change in response to adenine. Bases are colored as in (F). (M) A sequence from *A. thaliana* that has a significant change in correlation upon addition of adenine. For (F, G, and I-M) samples were normalized as described in the Supplementary Information. See also Supplementary Figures 2, 4, and 5.

*Pseudouridylation Prevents Conformational Switching Upon the Addition of Adenine.* Pseudouridine ( $\Psi$ ) and N1-methylpseudouridine ( $m1\Psi$ ) are naturally occurring post-transcriptional modifications used widely in RNA therapeutics due to their ability to evade immune recognition and increase translation (Morais et al., 2021). Since pseudouridylation has previously been shown to exert both stabilizing and destabilizing effects on RNA structure (Davis, 1995; Meroueh et al., 2000; Vögele et al., 2023), we examined the functional consequences these modifications would have on the purine riboswitches analysed above. The procedure was repeated as before, but with uridine (U) replaced with either  $\Psi$  or  $m1\Psi$  during *in vitro* transcription of the library (Figure 4 and Supplementary Figures 8 and 9).

Incorporation of  $\Psi$  resulted in structural changes in more than half of the library sequences passing quality control filters (37 of 60; Figure 4A). These changes indicate broad structural transformations, with decreasing levels of change highlighted in Supplementary Figure 8. Similar DMS patterns, indicative of structural re-arrangements, were also observed in sequences modified with  $m1\Psi$ , detailed in Supplementary Figure 9B (17 of 59). These structural effects reflect the chemical properties of  $\Psi$  and  $m1\Psi$ , which both contain a C5-C1' bond that augments base-pairing, base-stacking, and duplex stability, likely reducing the energy cost of non-native structures (Davis, 1995; Parr et al., 2020). While the DMS reactivities in  $\Psi$  and  $m1\Psi$  modifications generally demonstrated some correlation, indicating similarities in their structural impacts, the extent of these similarities varied depending on the sequence (Supplementary Figure 9C-D). This variability highlights that  $\Psi$  and  $m1\Psi$  exhibit distinct behaviours in different sequences, likely due to the fact that  $\Psi$  has an extra hydrogen bond donor at N1, which is absent in  $m1\Psi$ , and is therefore capable of wobble base-pairing with G, U, or C in a duplex (Kierzek et al., 2014). The differences are further suggested by the even more drastic structural changes seen upon incorporation of  $\Psi$  compared to  $m1\Psi$ . Notably, the introduction of either modification prevented conformational changes in all four adenine-sensing riboswitches and the COOLAIR structure (Figures 4 C-L). Given that  $\Psi$  has been shown to exert context-dependent destabilizing effects in functional RNAs, as in the *E. coli* 23S rRNA and the neomycin-sensing riboswitch (Meroueh et al., 2000; Vögele et al., 2023), modifications in the U-rich aptamer domain of the *add* riboswitch would likely also disrupt tertiary contacts crucial for native folding, thereby eliminating adenine-sensing and functional switching of the pseudouridylated riboswitches.



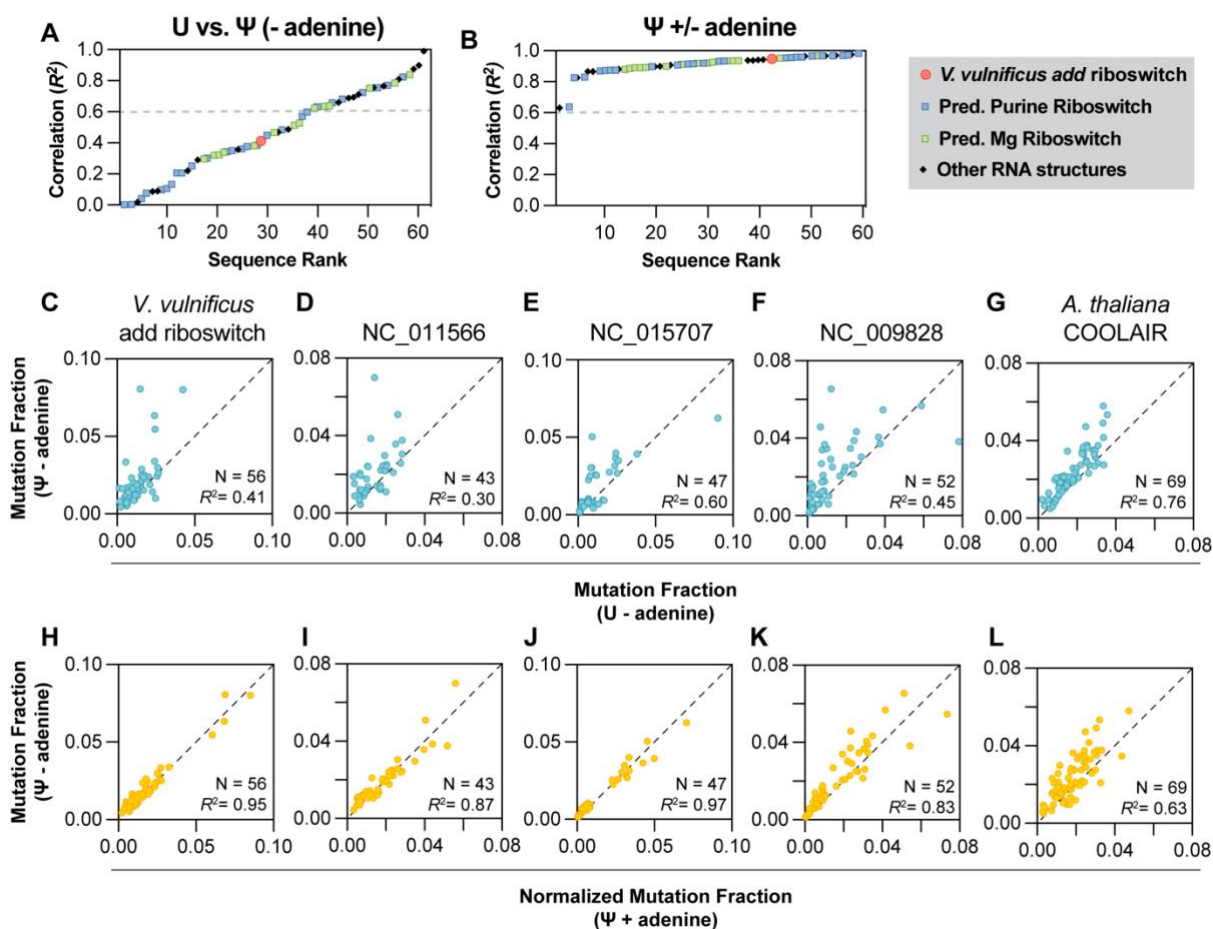


Figure 4. Effects of Pseudouridylation on RNA structure and riboswitching. (A) Ranked correlation of references compared to each other in samples containing all uridines or all pseudouridines. (B) Ranked correlation of references compared to each other in samples containing all pseudouridines in the presence and absence of 5 mM adenine. Each symbol represents one reference. (C-G) Comparison of the DMS signal of adenine responsive sequences from Figure 3 transcribed either with all uridines or all pseudouridines in the absence of adenine. Each turquoise circle represents one A or C base in the variable section. Neither sample was normalized. (H-L) Comparison of the DMS signal of adenine responsive sequences from Figure 4 transcribed with all pseudouridines in the presence or absence of adenine. Each yellow circle represents one A or C base in the variable section. The sample 2 ( $\Psi$  + Adenine) is normalized to sample 1 ( $\Psi$  - Adenine) as described in Supplementary Information. For (C-L), N represents the number of A and C bases. The Pearson  $R^2$  was calculated in SEISMICgraph.

## DISCUSSION

We developed SEISMICgraph to streamline and standardize the analysis of chemical probing data. This tool provides various plotting and filtering options that enable rapid assessment and quantitative evaluation of experimental samples. It facilitates a broad overview by allowing researchers to examine multiple gene references simultaneously while also supporting detailed analysis by focusing on specific sequences. Additionally, SEISMICgraph allows for fast and thorough comparisons across different samples, experimental conditions, and sequences, making it a versatile and efficient resource for RNA structure research.

We confirmed three key findings using SEISMICgraph. First, we validated the structural changes in a well-documented adenine riboswitch. Second, we characterized the structures of a library of 81 sequences including 30 of putative purine riboswitches and found that three exhibit changes in response to adenine. Third, we demonstrated that introducing base modifications results in defects in structural switching. Notably,  $\Psi$  and  $m^1\Psi$  modifications prevent the adenine-sensing function of the *add* riboswitch and the three other purine-responsive riboswitches.

These findings present a challenge for RNA therapeutics, as RNA modifications are often used to avoid host immune responses but may interfere with essential functions, as shown here. Chemical probing techniques such as DMS and SHAPE can help identify when and where these interferences occur. SEISMICgraph can be used to visualize and interpret these results, providing valuable insights for designing functional structures and developing RNA therapeutics.

## DATA AVAILABILITY

SEISMICgraph is available at <https://seismicrna.org>. Code is available at <https://rousikinlab.github.io/seismic-graph/>. All data collected in this study can be found at the Sequence Read Archive, accession PRJNA1165154.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

## AUTHOR CONTRIBUTIONS

Federico Fuchs Wightman: Conceptualization, Library design, Validation, Formal analysis, Methodology, Writing-original draft. Yves J. Martin des Taillades: SEISMICgraph library development, Webapp UX/UI design, Writing-review & editing. Grant Yang: Conceptualization, Validation, Formal analysis, Methodology, Writing-review & editing. Casper L'Esperance-Kerckhoff: SEISMICgraph library implementation, Webapp UX/UI design and development. Scott Grote: Conceptualization and Library design. Matthew F. Allan: SEISMIC development and adaptation to SEISMICgraph. Daniel Herschlag: Conceptualization and Writing-review & editing. Silvi Rouskin: Conceptualization, Formal analysis, Writing-original draft. Lauren D. Hagler: Conceptualization, Library design, Methodology, Formal analysis, Writing-original draft.

## **ACKNOWLEDGEMENTS**

We thank the members of the Rouskin and Herschlag labs for the general discussion of the manuscript and their support.

## **FUNDING**

This work was supported by the National Institutes of Health [DP2AI175475 to S.R., R01GM13289904 to D.H.]; and the Howard Hughes Medical Institute [Hanna H. Gray Postdoctoral Fellowship to L.D.H.]. Funding for open access charge: National Institutes of Health.

## **CONFLICT OF INTEREST**

DH is a member of the Scientific Advisory Board of Arrakis Therapeutics.

## **REFERENCES**

- Allan, M.F., Aruda, J., Plung, J.S., Grote, S.L., des Taillades, Y.J.M., de Lajarte, A.A., Bathe, M., Rouskin, S., 2024. Discovery and Quantification of Long-Range RNA Base Pairs in Coronavirus Genomes with SEARCH-MaP and SEISMIC-RNA. *Res. Sq.* rs.3.rs-4814547. <https://doi.org/10.21203/rs.3.rs-4814547/v1>
- Allan, M.F., Brivanlou, A., Rouskin, S., 2023. RNA levers and switches controlling viral gene expression. *Trends Biochem. Sci.* 48, 391–406. <https://doi.org/10.1016/j.tibs.2022.12.002>
- Benner, S.A., Ellington, A.D., Tauer, A., 1989. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci.* 86, 7054–7058. <https://doi.org/10.1073/pnas.86.18.7054>
- Busan, S., Weeks, K.M., 2018. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA N. Y. N* 24, 143–148. <https://doi.org/10.1261/rna.061945.117>
- Davis, D.R., 1995. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* 23, 5020–

5026. <https://doi.org/10.1093/nar/23.24.5020>

Douds, C.A., Babitzke, P., Bevilacqua, P.C., 2024. A new reagent for in vivo structure probing of RNA G and U residues that improves RNA structure prediction alone and combined with DMS. *RNA N. Y. N* 30, 901–919. <https://doi.org/10.1261/rna.079974.124>

Ehresmann, C., Baudin, F., Mougél, M., Romby, P., Ebel, J.P., Ehresmann, B., 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res.* 15, 9109–9128. <https://doi.org/10.1093/nar/15.22.9109>

Homan, P.J., Favorov, O.V., Lavender, C.A., Kursun, O., Ge, X., Busan, S., Dokholyan, N.V., Weeks, K.M., 2014. Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U. S. A.* 111, 13858–13863. <https://doi.org/10.1073/pnas.1407306111>

Incarinato, D., Morandi, E., Simon, L.M., Oliviero, S., 2018. RNA Framework: an all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications. *Nucleic Acids Res.* 46, e97. <https://doi.org/10.1093/nar/gky486>

Jones, S., 2016. Protein-RNA interactions: structural biology and computational modeling techniques. *Biophys. Rev.* 8, 359–367. <https://doi.org/10.1007/s12551-016-0223-9>

Kierzek, E., Malgowska, M., Lisowiec, J., Turner, D.H., Gdaniec, Z., Kierzek, R., 2014. The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.* 42, 3492–3501. <https://doi.org/10.1093/nar/gkt1330>

Lucks, J.B., Mortimer, S.A., Trapnell, C., Luo, S., Aviran, S., Schroth, G.P., Pachter, L., Doudna, J.A., Arkin, A.P., 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. <https://doi.org/10.1073/pnas.1106501108>

Ma, H., Jia, X., Zhang, K., Su, Z., 2022. Cryo-EM advances in RNA structure determination. *Signal Transduct. Target. Ther.* 7, 1–6. <https://doi.org/10.1038/s41392-022-00916-0>

Merino, E.J., Wilkinson, K.A., Coughlan, J.L., 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. <https://doi.org/10.1021/ja043822v>

Meroueh, M., Grohar, P.J., Qiu, J., SantaLucia Jr, J., Scaringe, S.A., Chow, C.S., 2000. Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of *Escherichia coli* 23S rRNA. *Nucleic Acids Res.* 28, 2075–2083. <https://doi.org/10.1093/nar/28.10.2075>

Morais, P., Adachi, H., Yu, Y.-T., 2021. The Critical Contribution of Pseudouridine to mRNA COVID-19 Vaccines. *Front. Cell Dev. Biol.* 9.

Mukherjee, S., Das Mandal, S., Gupta, N., Drory-Retwitzer, M., Barash, D., Sengupta, S., 2019. Ribod: a comprehensive database for prokaryotic riboswitches. *Bioinformatics* 35, 3541–3543. <https://doi.org/10.1093/bioinformatics/btz093>

Parr, C.J.C., Wada, S., Kotake, K., Kameda, S., Matsuura, S., Sakashita, S., Park, S., Sugiyama, H., Kuang, Y., Saito, H., 2020. N 1-Methylpseudouridine substitution enhances the performance of synthetic mRNA switches in cells. *Nucleic Acids Res.* 48, e35. <https://doi.org/10.1093/nar/gkaa070>

Peattie, D.A., Gilbert, W., 1980. Chemical probes for higher-order structure in RNA. *Proc. Natl. Acad. Sci. U. S. A.* 77, 4679–4682. <https://doi.org/10.1073/pnas.77.8.4679>

Peracchi, A., Beigelman, L., Usman, N., Herschlag, D., 1996. Rescue of abasic hammerhead ribozymes by exogenous addition of specific bases. *Proc. Natl. Acad. Sci.* 93, 11522–11527. <https://doi.org/10.1073/pnas.93.21.11522>

Romero-Agosto, G., Cox, E., Rouskin, S., 2023. DMS-MapSeq Analysis of Antisense Oligonucleotide Binding to lncRNA PANDA. <https://doi.org/10.1101/2023.10.22.563486>

- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., Weissman, J.S., 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505, 701–705. <https://doi.org/10.1038/nature12894>
- Serganov, A., Yuan, Y.-R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A.T., Hobartner, C., Micura, R., Breaker, R.R., Patel, D.J., 2004. Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs. *Chem. Biol.* 11, 1729–1741. <https://doi.org/10.1016/j.chembiol.2004.11.018>
- Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A.E., Weeks, K.M., 2014. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* 11, 959–965. <https://doi.org/10.1038/nmeth.3029>
- Soukup, G.A., Breaker, R.R., 1999. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA N. Y. N* 5, 1308–1325. <https://doi.org/10.1017/s1355838299990891>
- Spitale, R.C., Incarnato, D., 2023. Probing the dynamic RNA structurome and its functions. *Nat. Rev. Genet.* 24, 178–196. <https://doi.org/10.1038/s41576-022-00546-w>
- Tian, S., Kladwang, W., Das, R., 2018. Allosteric mechanism of the *V. vulnificus* adenine riboswitch resolved by four-dimensional chemical mapping. *eLife* 7, e29602. <https://doi.org/10.7554/eLife.29602>
- Tomezsko, P.J., Corbin, V.D.A., Gupta, P., Swaminathan, H., Glasgow, M., Persad, S., Edwards, M.D., Mcintosh, L., Papenfuss, A.T., Emery, A., Swanstrom, R., Zang, T., Lan, T.C.T., Bieniasz, P., Kuritzkes, D.R., Tsibris, A., Rouskin, S., 2020. Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* 582, 438–442. <https://doi.org/10.1038/s41586-020-2253-5>
- Turner, D.H., Mathews, D.H. (Eds.), 2016. *RNA Structure Determination: Methods and Protocols*, *Methods in Molecular Biology*. Springer, New York, NY. <https://doi.org/10.1007/978-1-4939-6433-8>
- Vögele, J., Duchardt-Ferner, E., Kruse, H., Zhang, Z., Sponer, J., Krepl, M., Wöhnert, J., 2023. Structural and dynamic effects of pseudouridine modifications on noncanonical interactions in RNA. *RNA* 29, 790–807. <https://doi.org/10.1261/rna.079506.122>
- Wan, Y., Kertesz, M., Spitale, R.C., Segal, E., Chang, H.Y., 2011. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* 12, 641–655. <https://doi.org/10.1038/nrg3049>
- Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Swanstrom, R., Burch, C.L., Weeks, K.M., 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460, 711–716. <https://doi.org/10.1038/nature08237>
- Yang, M., Zhu, P., Cheema, J., Bloomer, R., Mikulski, P., Liu, Q., Zhang, Y., Dean, C., Ding, Y., 2022. In vivo single-molecule analysis reveals COOLAIR RNA structural diversity. *Nature* 609, 394–399. <https://doi.org/10.1038/s41586-022-05135-9>
- Zhang, J., Fei, Y., Sun, L., Zhang, Q.C., 2022. Advances and opportunities in RNA structure experimental determination and computational modeling. *Nat. Methods* 19, 1193–1207. <https://doi.org/10.1038/s41592-022-01623-y>
- Zubradt, M., Gupta, P., Persad, S., Lambowitz, A.M., Weissman, J.S., Rouskin, S., 2017. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* 14, 75–82. <https://doi.org/10.1038/nmeth.4057>