A small cationic probe for accurate, punctate discovery of RNA tertiary structure

Jeffrey E. Ehrhardt[1,4], David Y. Qiu[1,4], Shouhong Jin[1], Mark A. Boerneke[1], Caroline J. Aufgebauer[2], Stacy M. Horner[2,3], and Kevin M. Weeks[1,*]

[1] Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290
[2] Department of Molecular Genetics and Microbiology, Duke University, Durham, NC 27710
[3] Department of Integrative Immunobiology, Duke University, Durham, NC 27710

[4] contributed equally
* correspondence, weeks@unc.edu

Abstract

RNA can fold to form complex three-dimensional tertiary structures, with large roles in RNA biology. However, accurate discovery of motifs that form true tertiary structures remains an unsolved challenge. Here we demonstrate that at rare but distinctive sites, RNA tertiary folding creates electronegative pockets that undergo selective chemistry with a small positively-charged chemical probe, trimethyloxonium (TMO). Sites preferentially reactive with the TMO cation, relative to uncharged dimethyl sulfate (DMS), termed T-sites, reflect preferential interaction of TMO at punctate electronegative pockets created by juxtaposition of a reactive nucleobase with non-bridging oxygen atoms in the RNA backbone. T-sites map specifically to centers of higher-order structural interactions and functional cores in diverse RNAs, including in specific states in conformational ensembles. Probing the 10,723-nt dengue virus RNA genome revealed exactly three T-sites, each of which is embedded within a complex RNA structure with direct roles in viral replication. T-site chemistry creates broad opportunities for high-confidence discovery and analysis of functional RNA tertiary structures across long and complex RNAs, including transcriptome-wide.

Introduction

RNA molecules encode information through both primary genetic sequence and via higher-order structural elements. RNA elements with complex structures mediate biological functions ranging from catalysis, to ligand responses, to translation regulation[1–4]. Chemical probing has been used extensively to characterize and define RNA structures, and recent developments in massively

parallel sequencing have enabled high-throughput characterization of RNA secondary structure. In particular, methods using SHAPE and DMS probing and read out based on mutational profiling (MaP) and single-molecule correlated chemical probing (smCCP) strategies are revealing that coding and non-coding RNAs contain extensive, often directly functional, internal secondary structure elements that are expressed both as stable structures and as ensembles[5–7]. Experimentally-informed secondary structure modeling, including for large and complex RNAs, is now well established and typically highly accurate[6,8–10]. Chemical probing is now an essential tool for understanding RNA structure-function interrelationships at the level of RNA base pairing, extending from understanding the regulatory logic of RNA viruses, to defining how RNA structure functions as a ubiquitous switch or rheostat in bacterial and eukaryotic systems, to optimizing RNA medicines[5,6,11–13].

By contrast, the extent to which true *tertiary* structures form in RNAs remains poorly explored. RNA tertiary structures are highly diverse and include non-canonical pairing, kissing loop contacts, pseudoknots, and indirect backbone interactions[14,15]. The structural diversity of tertiary interactions and the fact that even complex tertiary structure motifs contain extensive simple base pairing has made devising broadly applicable chemistry for definitive RNA tertiary structure discovery challenging. Conventional chemical probing technologies are certainly sensitive to formation of tertiary contacts, as these methods generally measure conformational flexibility or steric accessibility. However, these approaches do not distinguish whether a given reactivity measurement or difference is reflective of base pairing versus higher-order tertiary interactions. Existing probing strategies that show partial selectivity for RNA tertiary structures have done so via measurements of solvent accessibility, crosslinking (UV, psoralen or SHAPE based), reaction with soft metal ions (for example, Pb^{2+} and Tb^{3+}), or single-molecule correlated chemical probing[7,16–18]. Drawbacks of extant approaches include low signal above background, high read depth requirements when read out by massively parallel sequencing, and convolution of secondary and tertiary structure features. Current strategies are thus limited due to challenging implementation, difficulty assigning measurements specifically to RNA tertiary structure, and lack of suitability for *de novo* RNA tertiary structure discovery.

Here we devise a chemistry-based probing strategy highly selective for RNA tertiary structure by taking advantage of structure and charge features unique to RNA motifs with complex folds. RNAs with complex structures create surfaces with distinctive electrostatic profiles, which result from confining the negative charge in the phosphoribose backbone in tightly folded tertiary

motifs[15,19,20]. We developed a comparative chemical strategy that measures reactivity differences between positively charged and neutral reagents. In probing an ~11,000-nt dengue virus (DENV) RNA genome, we identified exactly three tertiary structure elements: two show the precise geometry expected of an electronegative pocket based on comparison with known structures, and the third corresponds to a proven regulator of viral replication. This charge-based chemical probing strategy thus identifies distinctive electronegative pockets in a precise and punctate way, enabling detection of RNA tertiary structures with high accuracy and low false-positive discovery.

Results

Differential reactivity of TMO and DMS for RNAs with complex structures

We reasoned that the key to developing a chemical probing strategy highly selective for true RNA *tertiary* structure would be to focus on RNA electrostatics. Individual RNAs feature diverse backbone geometries, extending from the smooth helix characteristic of simple helices to tight kinks and turns found in riboswitches and RNAs with complex structures[15]. Specific unusual conformations of an RNA backbone might create an electrostatic environment favoring or disfavoring reaction by charged molecules.

We recently introduced trimethyloxonium (TMO) as a reagent for RNA structure probing[21]. TMO is similar in size to and alkylates RNA via the same electrophilic mechanism as dimethyl sulfate (DMS)[21] (Figure 1). However, TMO has a much shorter half-life than DMS and is positively charged. The short half-life of TMO can be advantageously leveraged to study time-resolved RNA folding and to capture through-space interaction networks during folding[21]. Here we will leverage the positive charge on TMO to create a chemistry-based strategy for RNA tertiary structure discovery. TMO and DMS reactivities are efficiently quantified using massively parallel sequencing via mutational profiling (MaP) reverse transcription[21,22], including at low read depths (Figure S1).

We probed three RNAs of known three-dimensional structure: the catalytic core of a bacterial RNase P (termed RNase P), the thiamine pyrophosphate (TPP) aptamer domain from the TPP riboswitch, and a group II intron. All three RNAs contain complex tertiary structures critical to their biological function[23–25]. RNase P cleaves tRNAs during their biogenesis, the TPP riboswitch binds TPP to regulate gene expression in prokaryotes, and group II introns are self-splicing ribozymes. Across all three RNAs, the reactivity patterns of TMO and DMS reactivity

were broadly similar. Both probes reacted with nucleotides in unstructured loop and bulge regions, consistent with their expected reaction mechanism with RNA (Figures 1B, 2A).

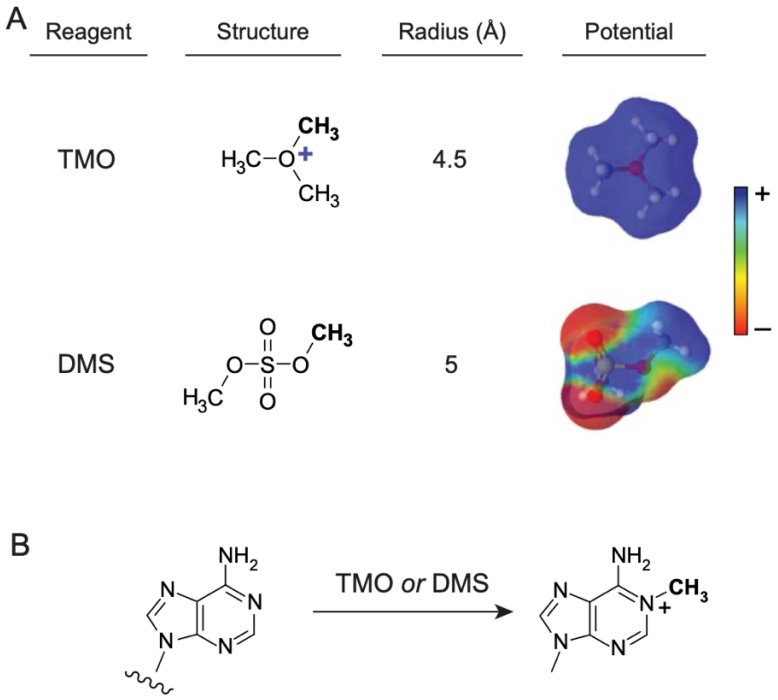

Figure 1: Characteristics and RNA chemistry for TMO and DMS. (**A**) Structures and surface charge. (**B**) Reaction with adenine.

Strikingly, we observed distinctive, large-magnitude, TMO-specific reactivity at a single site in each model RNA. In each case, the differentially reactive nucleotide was an adenosine. In RNase P, the largest difference in reactivity was at adenosine 45, where TMO was ~20-fold more reactive than DMS (Figure 2C). Although nucleotide 45 is unpaired in secondary structure diagrams, it is structurally constrained in the catalytic core of the RNase P RNA by base stacking and non-canonical interactions (Figure 2A)[24]. In the TPP aptamer domain at nucleotide A44 and in the group II intron at nucleotide A376, we observed 10- and 8-fold increases in reactivity of TMO over DMS, respectively (Figure 2B, 2D). As with RNase P, these sites are in the complexly folded tertiary core of each RNA[23,25]. We refer to these TMO-reactive sites, indicative of RNA tertiary structure, as T-sites.

We also observed the converse: At a subset of adenosine residues in the RNase P and group II intron RNAs, DMS was more reactive than TMO (Figure 2C, 2D). We will show below that these DMS-reactive sites correspond to adenosines involved in A•G base pairs. These differential

TMO and DMS reactivity signals are distinctive in that one reagent generally shows low reactivity, while the other is highly reactive (Figure 2E, 2F). These patterns imply that there is a structure-based mechanism that creates specific probe-binding pockets that govern each class of reactivity.

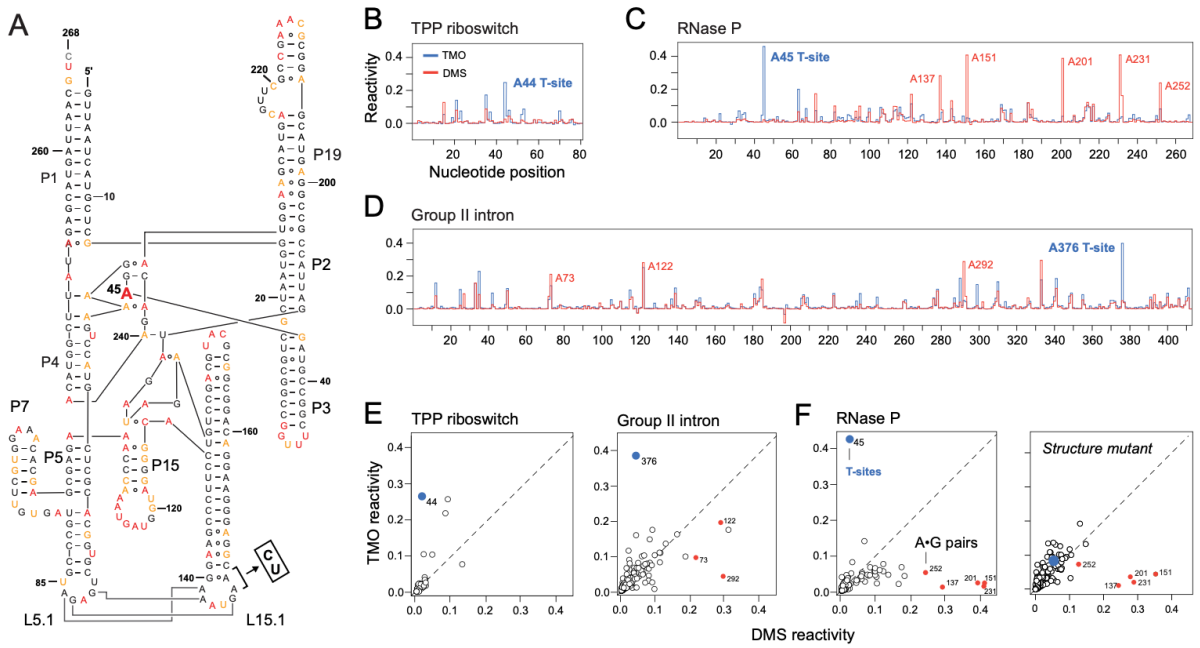

Figure 2: TMO and DMS reactivity profiles for three RNAs. (**A**) Secondary structure of RNase P, annotated with mutation that disrupts the L5.1-L15.1 interaction. T-site (at position 45) is shown as large bold A; loop-loop interaction was disrupted by AA to CT mutation in L15.1 (box). Nucleotides are colored by TMO reactivity; red, orange and black correspond to high, medium and low reactivities, respectively. (**B-D**) TMO (blue) and DMS (red) reactivity profiles for (**B**) TPP aptamer domain, (**C**) RNase P, and (**D**) group II intron. Sites with significant differential reactivities are labeled; T-sites are in bold blue text; sites with DMS reactivity greater than TMO reactivity (A•G pairs) are labeled in red. (**E**) Correlation plots of reactivity rates for RNA modified with DMS (x-axis) and TMO (y-axis). T-sites (blue) and A•G pairs (red) with distinctive reactivities are labeled. (**F**) For RNase P, both the native sequence and a tertiary structure-disrupting mutant are shown.

T-site reactivities are strongly and specifically dependent on RNA tertiary structure

There is a single T-site in each of the model RNAs, which span a wide range in size (80 to 412 nts) and have completely different global architectures. The reactivity signal was separated from background for each RNA by large factors, among or exceeding the largest we have ever

observed in chemical probing experiments (Figure 2E, 2F). These features suggest that T-sites are strong predictors of specific RNA tertiary structure motifs.

We examined the hypothesis that T-sites are strictly dependent on formation of higher order RNA structure in two ways. First, we mutated two nucleotides in the L15.1 loop in RNase P to weaken a loop-loop interaction that is critical to higher-order folding of this RNA[21,24] (Figure 2A). As expected, nucleotide reactivities increased precisely at the (L5.1-L15.1) loop-loop interaction, consistent with its disruption (Figure S2A). Probing with TMO and DMS showed that the overall reactivity pattern for each reagent was otherwise largely unperturbed, consistent with maintenance of the native base-paired RNA structure. By contrast, differential TMO-vs-DMS reactivity was lost at the T-site, reflective of selective loss of tertiary structure in this mutant (Figure 2F). In sum, T-site reactivity, but not general per-nucleotide reactivity, is specifically and strongly dependent on formation of native tertiary structure in RNase P.

Second, we examined TMO and DMS reactivities for the TPP aptamer domain, RNase P and group II intron RNAs as a function of Mg^{2+}concentration, which stabilizes closely packed negatively charged helices in RNA tertiary structures. All three examined RNAs adopted native-like RNA secondary structure folds in the absence of Mg^{2+}, as supported by the correlation between per-nucleotide reactivities of the two reagents. In contrast, the T-site signal is substantially reduced when the three RNAs were folded and probed in the absence of Mg^{2+} (Figure S2B). These observations are consistent with the model that a specific, constrained higher-order tertiary structure enables an experimentally observable T-site.

T-site and A•G base pair structure determinants

Each of the T-sites we observed occurs in a region of closely packed RNA strands (Figure 3A). The T-sites in the model RNAs are structurally distinct, but each site shares specific atomistic features (Figure 3B). First, the reactive T-site in each of the three RNAs is an unpaired adenosine. Second, local folding of the RNA backbone places the structurally constrained adenosine residue within 4-7 Å of two nonbridging phosphate oxygen atoms. Third, the reactive adenosine residue of each T-site stacks with nearby nucleobases. Despite these atomistic similarities, the specific arrangements of the RNA strands in the three RNAs are completely different (Figure 3B). Indeed, in the three RNAs studied here, the sequence distance between the T-site adenosine and a non-bridging oxygen varies from 30 nucleotides in the TPP aptamer domain to 371 nucleotides in the group II intron. There are thus a wide variety of local RNA

geometries capable of creating a T-site, but such sites occur only in regions of complex tertiary interactions. These features support a chemical model in which TMO preferentially binds in an electronegative pocket adjacent to the reactive adenosine N1 atom (Figure 3C).

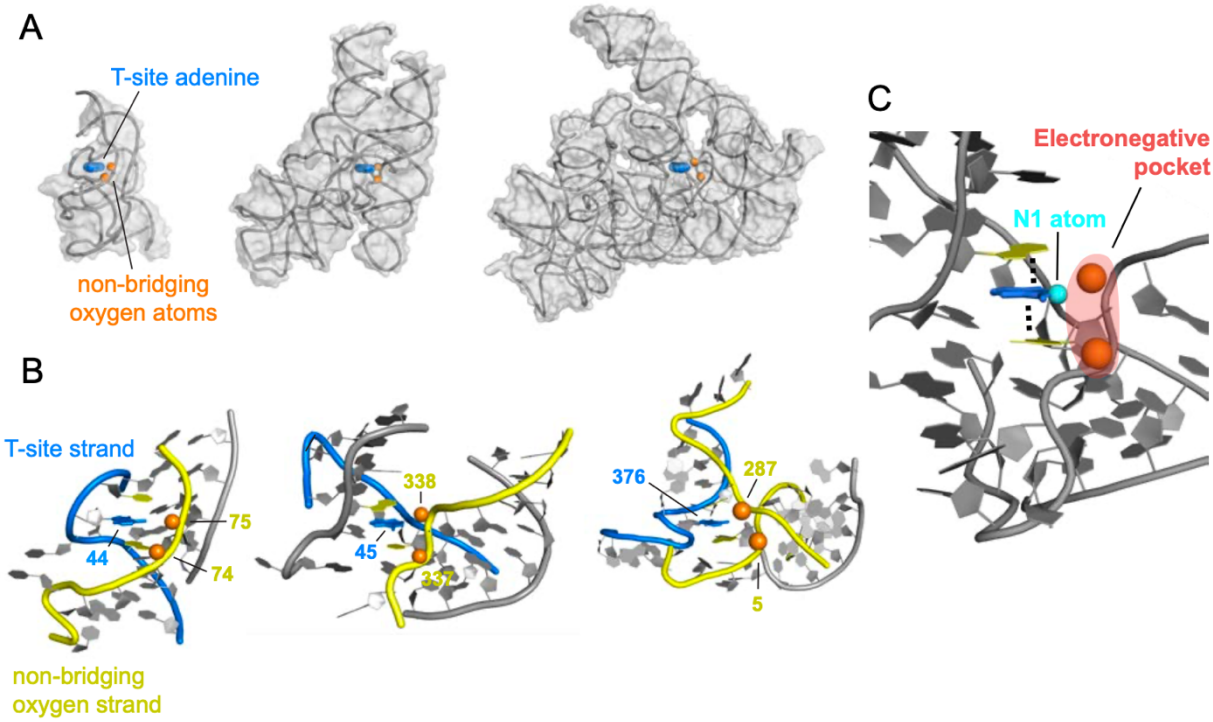

Figure 3: Visualization of T-sites in RNAs with complex tertiary folds. (**A**) RNA structures (grey) showing T-site residue (blue) situated proximal to nonbridging phosphate oxygen atoms (orange spheres). (*left* to *right*) TPP riboswitch aptamer domain, RNase P and group II intron. RNAs are positioned such that T-site adenosine residues have the same orientation. (**B**) Detailed views showing geometry of the differentially reactive adenine nucleobase (blue), proximal nonbridging oxygen atoms (orange), and nucleotides that stack with the reactive position (yellow). Sequence positions of T-site adenosine and proximal non-bridging oxygen atoms are numbered. RNA backbones for the strands on which the T-site adenosine and non-bridging phosphate atoms are located are emphasized in blue and yellow, respectively. (**C**) Summary of features found at T-sites. T-site adenosine is blue; reactive N1 atom is cyan; the two closest non-bridging oxygens are orange; and nucleotides that engage in base stacking with the T-site adenine nucleobase are yellow. Region of electronegative potential is emphasized in red.

In addition to the single adenosine residues in each RNA that were preferentially reactive with TMO, we also identified multiple sites in the RNase P and group II intron RNAs that showed the converse behavior and reacted more strongly with DMS than with TMO (Figure 2). These sites all occur at A•G base pairs[24,25] (Figure S3), one of the most common non-canonical pairs[26,27]. All pairs that are preferentially reactive toward DMS are *trans* A•G pairs that interact via the Hoogsteen face of A and the sugar edge of G. This pairing geometry creates an adenosine nucleotide with a distinctive conformation in which the adenosine N1 atom is accessible for reaction and is juxtaposed near the 2'-OH of the pairing G residue (Figure S3). We hypothesize that this 2'-OH group facilitates optimal positioning of the DMS reagent, which is predictive for A•G non-canonical base pairing and allows their identification in complex RNAs.

Mapping higher-order RNA structure around T-sites

To understand the relationship between T-sites and higher-order structure, we assessed the effect of disrupting, by mutation, RNA tertiary structure specifically at each T-site. For each model RNA, we mutated the T-site adenosine to uridine or introduced 1- or 3-nucleotide deletions centered at the T-site. The overall reactivity profiles for each of the mutants were similar to those of the native sequence when the RNAs were probed with TMO, emphasizing that the mutants adopt native-like RNA secondary structure. However, specific nucleotides within each RNA mutant displayed punctate increases in reactivity (Figure 4A and S4). The nucleotides with increased reactivities were not involved in canonical base pairing interactions but all lie close in three-dimensional space to the T-site and are often located distant in the primary sequence (Figure 4B). In sum, analysis of RNA T-site disruption mutants reveals the close-in-space environment surrounding the T-site and further supports the model that T-sites report critical, core interactions in an RNA tertiary structure.

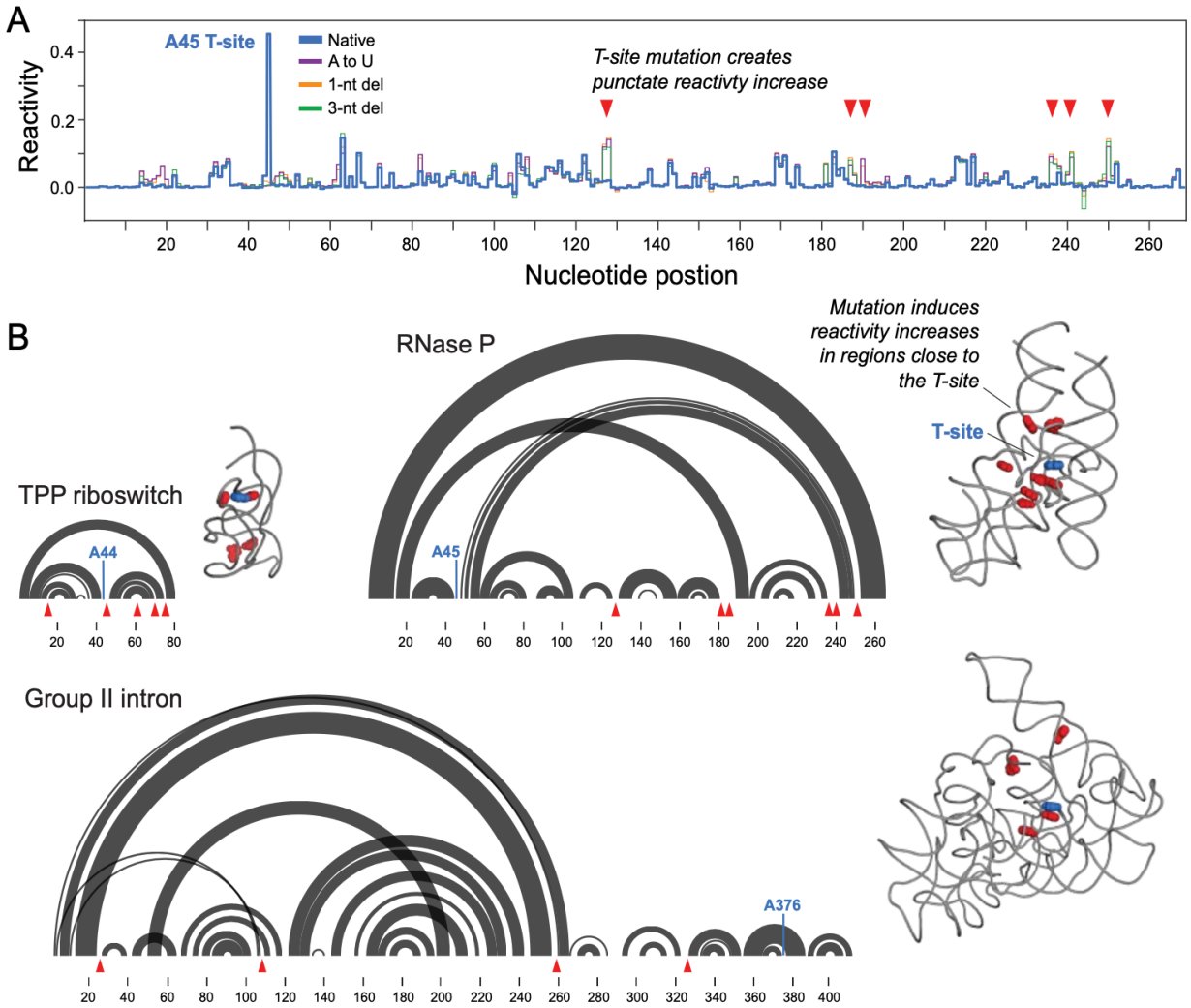

Figure 4: Mapping the structural neighborhood surrounding a T-site by mutation. (**A**) Comparison of TMO reactivity profiles for native and mutant sequences for the RNase P RNA. Nucleotides that show significant (≥0.05) increases in reactivity rate in the 3-nt deletion mutation are denoted with red triangles. (**B**) Secondary and tertiary structures for the TPP aptamer domain, RNase P, and group II intron RNAs, the locations of T-sites, and perturbations induced by mutating the T-site. Secondary structures (base pairing) are shown as arcs (grey). Three dimensional structures shown as gray backbone. T-sites are labeled in blue. Nucleotides that show significant reactivity increases in 3-nt deletion mutants are emphasized in red. Reactivity profiles for the TPP riboswitch and group II intron are shown in Figure S2.

T-sites identify the tertiary structure-containing state in a conformational ensemble

Many RNAs populate conformational ensembles, as exemplified by riboswitches[28]. Riboswitches typically contain a ligand-sensitive aptamer domain and an adjoining expression

platform (Figure 5A). The ThiM riboswitch populates translational OFF and ON states, where the OFF state is stabilized upon binding of the TPP ligand to the aptamer domain, which has the well-defined tertiary structure characterized above (Figure 3A). The OFF state also sequesters the Shine-Dalgarno (SD) site, essential for translation initiation, in a stable helix[29]. As demonstrated here, this state contains a T-site at nucleotide 44 (Figure 5A). The ON state adopts a fold containing simple stem-loop motifs and that presents the start codon and Shine-Dalgarno sequence in a single stranded region, accessible for interaction with the ribosome and initiator tRNA (Figure 5B). The OFF state is populated in both the presence and absence of TPP, at ~80 and 40%, respectively, for our structure-switching riboswitch construct (Figure S5).

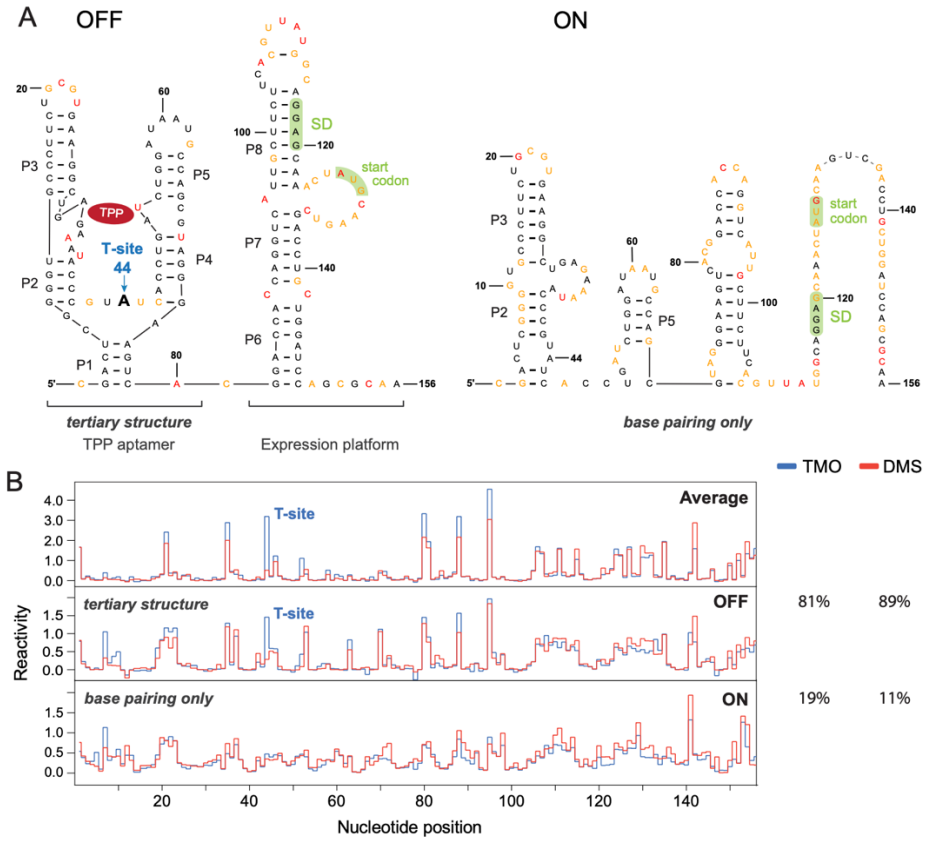

Figure 5: Detecting tertiary structure specific to individual state in a conformational ensemble for a structure-switching TPP riboswitch. Riboswitch was probed in the presence of saturating concentrations of the TPP ligand. (**A**) Secondary structures of the ThiM TPP riboswitch in the OFF and ON translation states, modeled by DANCE-MaP based on DMS-MaP data. Nucleotides are colored by DMS reactivity. Red, orange and black correspond to high, medium, and low reactivities, respectively. T-site (at position 44) is shown as large bold A; the Shine–Dalgarno (SD) and translational start codon are highlighted with green. (**B**) TMO (blue) and DMS (red) reactivity profiles for the TPP riboswitch. The upper panel shows the averaged,

bulk reactivities, prior to deconvolution. Middle and lower lanes show reactivities deconvoluted by DANCE-MaP[31]. Population percentages for the OFF and ON states, determined from deconvolution of TMO and DMS data, respectively, are listed at right. Comparative data, obtained in the presence and absence of TPP, are provided in Figure S3.

We probed the structure-switching riboswitch RNA in the presence and absence of TPP with both TMO and DMS. We used a single-molecule correlated chemical probing (smCCP) strategy, called DANCE-MaP, to deconvolute multiple conformations in the riboswitch ensemble[30,31]. As expected, bulk RNA reactivities showed a T-site at position A44 in the presence of TPP (Figure 5B, *top*). By contrast, ensemble deconvolution, based on either reagent, yielded similar populations of OFF and ON states (Figure 5B, *right*). Based on the secondary structures modeled for each state by DANCE, we assigned the SD-sequestered configuration as the OFF state. We could then recreate the reactivity patterns for each state, for each reagent. We observed the T-site only in the OFF state, which therefore correctly identifies the state with the TPP-binding *tertiary* structure (Figure 5B). Strikingly, in the absence of TPP, where the OFF state comprises only ~40% of the population, the T-site is not visible in the bulk per-nucleotide reactivity profiles but becomes identifiable in the deconvoluted OFF state (Figure S5). In sum, T-site probing specifically identifies the state in a conformational ensemble that contains higher-order tertiary structure and can reveal an, otherwise hidden (low population), tertiary structure in a complex ensemble.

Tertiary structure discovery across a viral RNA genome

Viral RNA genomes contain a diverse array of complex structures, and indeed encompass a large fraction of all known RNA tertiary structure motifs relevant to eukaryotic organisms[4]. We probed DENV serotype 2 (DENV2, New Guinea C strain) RNA, extracted from virions, with TMO and DMS. We obtain high-quality data across the entire 10,723-nt genome and identified three strong T-sites at positions A8293, A10417, and A10473 (Figure 6A; and Figure S6 for the entire genome). Strikingly, all three sites play pivotal roles in DENV2 replication. Two sites lie in the 3′-UTR, a region densely populated with replication control elements and RNA tertiary structures[4,32]. These two sites, A10417 and A10473, reside within an exonuclease resistant RNA (xrRNA) motif and a Dumbbell motif, respectively.

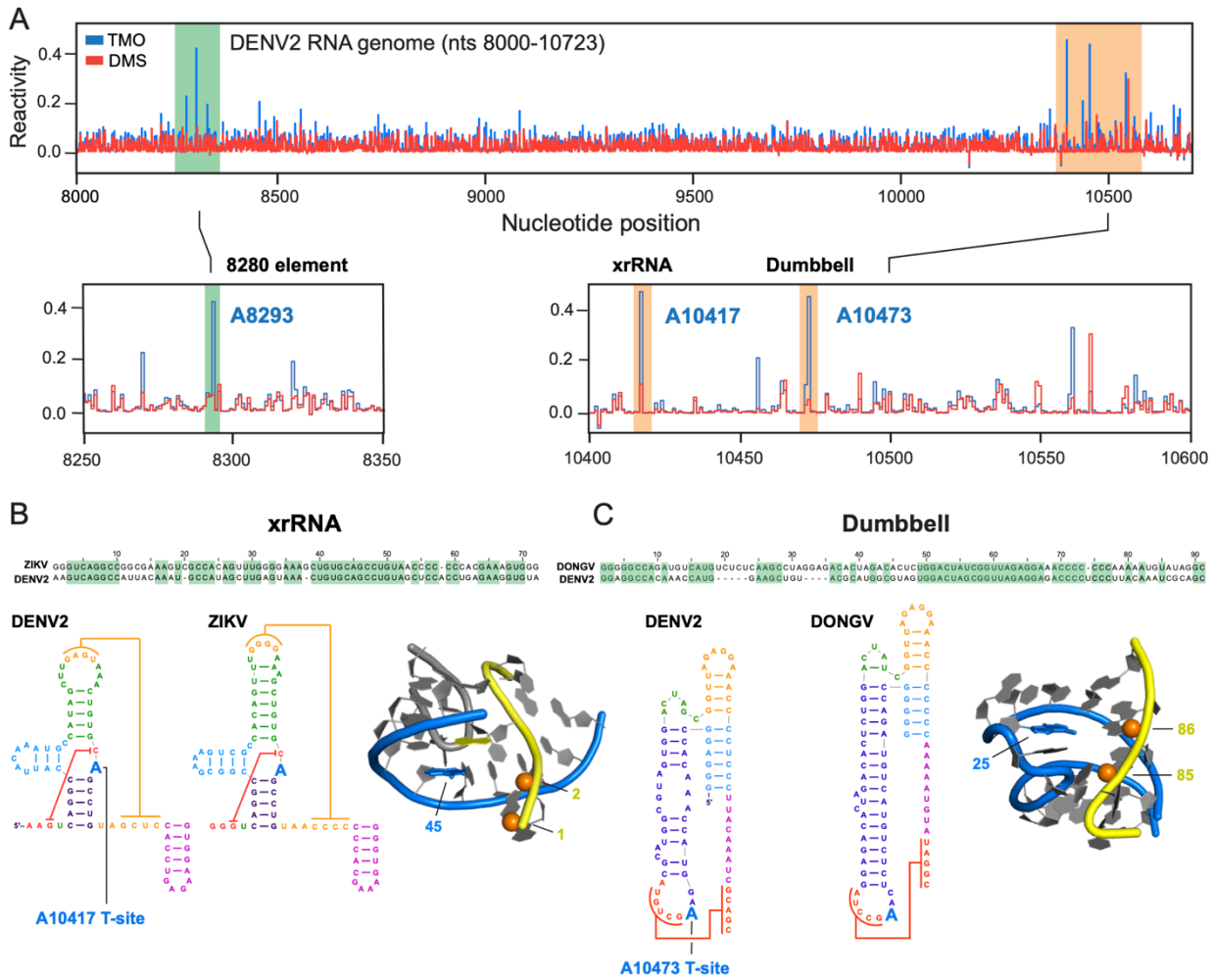

Figure 6: T-site detection across the DENV2 RNA genome. (**A**) Reactivity profiles for TMO (blue) and DMS (red) over a representative region of the DENV2 genome. Regions containing T-sites are shaded, and T-sites are labeled. Sequence alignments, secondary structure models, and three-dimensional models for (**B**) xrRNA and (**C**) Dumbbell motifs. DENV2 structures are compared with their ZIKV and DONGV analogs, for xrRNA and Dumbbell motifs, respectively. T-site nucleotides are emphasized with large blue nucleotide lettering. For the three-dimensional models, the T-site nucleotide and non-bridging phosphate oxygen atoms are shown in blue and as orange spheres, respectively; RNA backbones for the strands on which the T-site adenosine and non-bridging phosphate atoms are located are emphasized in blue and yellow, respectively. Nucleotides are numbered based on the crystal structures[33,34]. Full TMO and DMS probing data across the DENV2 genome are shown in Figure S6.

xrRNA and Dumbbell motifs are found across multiple viral superfamilies and representative structures of these elements have been determined[33,34]. xrRNA motifs form a highly stable,

pseudoknot-containing structure, which prevents degradation by cellular 5'→3' exonucleases[35]. The DENV2 xrRNA sequence is highly conserved with that of the Zika virus (ZIKV), whose structure is known (Figure 6B). Examination of the adenosine nucleotide in ZIKV corresponding to the DENV A10417 T-site shows that this nucleotide, indeed, adopts the precise T-site conformation[33]. The adenosine N1 atom points toward non-bridging phosphate oxygen atoms from a nearby backbone (Figure 6B). Dumbbell motifs regulate the transition from viral replication to translation[36]. Analysis of a homologous Dumbbell structure from a related insect-specific flavivirus, Donggang virus (DONGV), corresponding to the DENV A10473 T-site, again, reveals this nucleotide forms a T-site geometry (Figure 6C)[34]. For the third T-site, at position A8293, no 3D structure is known. However, the motif, termed the 8280 element, is conserved among all four DENV serotypes and mutations near A8293 significantly decrease viral replication and infectivity, highlighting the biological significance of this T-site motif[37]. These results emphasize that T-site probing enables *de novo* detection of RNA tertiary structure, across long RNAs, in a highly specific and punctate manner.

Both the xrRNA and Dumbbell motifs occur twice in the DENV2 3′-UTR, and we detect only one instance for each element. The first xrRNA motif (we detect the second) has a uridine instead of an adenosine nucleotide at the T-site position, and T-site chemistry (Figure 1B) currently only detects tertiary structure at adenosine residues. The potential T-site (position A10561) in the second Dumbbell motif (we detect the first) does display moderate TMO over DMS reactivity (Figure 6A), but not large enough in magnitude to meet defined T-site thresholds derived from model 3D RNA structures. We posit that the tertiary structure in the second Dumbbell is less stable than in the first, resulting in the lower differential reactivity. The presence of a T-site is thus highly predictive for RNA tertiary structure, but the absence of a T-site does not imply absence of higher-order structure.

Discussion

The key to precisely mapping RNA *tertiary* structure using chemical tools lies in finding a measurement that reflects a through-space interaction strongly correlated with higher-order structure. Absent such an insight, efficient *de novo* tertiary structure discovery has remained an important, widely sought, but unrealized goal. Here we show that differences in per-nucleotide reactivities for TMO and DMS, similar-in-size reagents with different charges, identify electrostatic pockets present in specific RNA tertiary structures. Our chemical probing strategy is unique in its ability to measure long-range, through-space interactions – between an

adenosine N1 atom and distantly positioned non-bridging oxygen groups – via an individual-nucleotide reactivity measurement (Figure 3C).

We identified a single T-site in each of three highly structured RNAs with known tertiary structures. We also observed eight sites with the reverse behavior – enhanced DMS reactivity – predictive of sheared A•G Hoogsteen-ribose face base pairs, one of the most common non-canonical pairing interactions[26,27]. The dual probing experiment chemistry is simple and direct, and read-out employs the efficient, well-validated MaP strategy[21,22,38]. Data interpretation is straightforward; the raw difference in reactivity rates between TMO and DMS yields the T-site signal. This differential signal is among the largest we have ever observed in a chemical probing experiment and is readily detectable at low sequencing read depths, opening the door for applications on transcriptome-wide scales. By introducing a mutation at a T-site (we favor the 3-nt deletion), the through-space structural neighborhood surrounding the T-site is defined as punctate increases in reactivity due to loss of tertiary interactions. The strong signal-to-background and rarity of T-site signals suggest this strategy can be broadly applied for novel RNA tertiary structure discovery.

T-sites are created by motifs that form unusual, but stable, electronegative pockets immediately adjacent to an accessible adenosine N1 atom (Figure 3C). This precise geometry is formed from highly diverse tertiary structure folds. Indeed, in the five RNAs visualized here (Figures 3 and 6), there is no similarity in backbone geometry or RNA strand orientation, beyond the proximal positions of the reactive adenine nucleobase and non-bridging oxygen atoms. In RNase P and the group II intron, the reactive adenosine and two critical non-bridging phosphate oxygen atoms are separated by hundreds of nucleotides. The structural diversity surrounding these T-sites emphasizes that a wide range of tertiary structures will be detectable. T-site motifs consistently map to the center of important tertiary structure interactions or catalytic cores in our model RNAs.

Our whole-genome DENV T-site studies emphasize important features of T-site chemistry. First, T-site probing is highly selective. Only three T-sites were detected across the DENV RNA genome. Two T-sites display the specific TMO-reactive conformation in known 3D RNA structures[33,34] and the third T-site occurs in a conserved structured region known to be functionally important[37]. Second, T-site chemistry identifies conserved tertiary structures despite sequence variation, as evidenced by our identification of T-sites in structurally conserved xrRNA

and Dumbbell motifs, whose sequences varied between RNA virus classes. Third, the presence of a T-site reveals a tertiary structure with high specificity, but the absence of a T-site does not rule out a local tertiary structure. Fourth, RNA tertiary structures, specifically discoverable as T-sites, are harbingers of function as evidenced by the observation that all three DENV T-sites are functionally critical for viral replication (in addition to the model 3D RNA structures with known functions). T-sites occur selectively at complex sites in RNA and such sites also comprise favored targets for small molecule ligands[39,40]. In sum, T-site chemistry is a facile, robust strategy that is poised to create broad opportunities to discover, analyze, and ligand RNA tertiary structures.

Methods

Electrostatics calculations

Solvated surface area and electrostatic potential for TMO and DMS were estimated using the MolCalc web server (molcalc.org)[41].

Synthesis of RNA constructs

DNA templates (IDT) for native and mutant sequences of the *Escherichia coli* TPP riboswitch (both aptamer domain and structure-switching constructs), the *Bacillus stearothermophilus* RNase P catalytic domain, and the *Oceanobacillus iheyensis* group II intron were embedded within 5' and 3' structure cassette flanking sequences and preceded with a 5' T7 promoter sequence[42]. DNA templates were amplified by PCR [1 mL; 0.2 mM each dNTP, 500 nM each forward and reverse primer (IDT), 40 nM template, 0.02 U/μL DNA Polymerase (Q5 Hot Start High-Fidelity, NEB) using the temperature program: [95 °C for 30 s, 30 × (95 °C for 15 s, 65 °C for 30 s, 72 °C for 30 s), 72 °C for 2 min]. DNA template was recovered (Omega Mag-Bind beads, 1× ratio beads to PCR volume) and eluted in H_2O. RNAs were transcribed in vitro [1 mL; 2.5 mM each NTP, 25 mM $MgCl_2$, 40 mM Tris, pH 8.0, 2.5 mM spermidine, 0.01% (wt/vol) Triton X-100, 10 mM DTT, 0.025 units pyrophosphatase, 2.5 μg T7 polymerase, 1.125 μg DNA template, in 50 μL H_2O; 37 °C; 6 h]. RNA was recovered (Omega Mag-Bind beads, 1.8× ratio beads to RNA volume) and resuspended in 50 μL of 0.1× TE (1 mM Tris pH 8.0, 0.1 mM EDTA).

TMO handling

TMO was used as the tetrafluoroborate salt (MilliporeSigma). TMO stocks were handled in a fume hood, stored in a desiccator at -20 °C, and were used only once after opening. Stock solutions of TMO were prepared in 1:2 (vol/vol) nitromethane:sulfolane (from ThermoFisher and

Sigma-Aldrich, respectively), as follows: TMO was first fully dissolved in nitromethane; subsequently, the TMO solution was diluted with two volumes sulfolane[21]. Final concentrations of TMO stock solutions used in this work (before addition of RNA) were 1.0 M. Sulfolane is a solid at room temperature and was warmed to 37 °C immediately prior to use. Due to the rapid kinetics of TMO hydrolysis (half-life ~ 7 sec[21]), reactions were initiated by adding the larger volume of RNA solution to the smaller TMO volume, followed by mixing with immediate rapid pipetting.

RNA folding and structure probing

Under the optimized conditions used here, both TMO and DMS react with all four ribonucleotides, with a preference for reaction with unpaired A and C residues[43]. The TPP riboswitch RNAs, RNase P, and group II intron RNA (5 pmol in 5 µL 0.1× TE for TPP constructs or in 6 µL 0.1× TE for RNase P and group II intron) were denatured at 95 °C for 2 min, cooled on ice, and incubated with 3 µL of 3× folding buffer (for TPP and RNase P, 33 mM $MgCl_2$, 333 mM NaCl, 1 M bicine, pH 8; for the group II intron, 375 mM $MgCl_2$, 333 mM NaCl, 1 M bicine, pH 8) at 37 °C for 20 min. If present, thiamine pyrophosphate ligand (1 µL of 50 µM thiamine pyrophosphate in H_2O) was added to the TPP aptamer or structure-switching riboswitch samples, and all constructs were incubated at 37 °C for 10 additional minutes. After folding, the RNA solutions (9 µL) were added into 1 µL of reagent (100 mM TMO or 170 mM DMS final; both reagents in 1:2 v/v nitromethane-sulfolane). The TMO reaction is self-quenching; DMS was quenched with 1 µL of 2-mercaptoethanol after 6 min incubation at 37 °C[21]. The no-reagent control treatment contained 1 µL 1:2 v/v nitromethane-sulfolane. Modified RNA was purified (Illustra MicroSpin G-50 columns, GE Healthcare). No-Mg^{2+} structure probing was carried out exactly as described above but with $MgCl_2$ omitted from the 3× folding buffer.

MaP-RT

Modified RNAs (resuspended in TE) were subjected to reverse transcription essentially as described[38,43]. A 10 µL solution containing RNA, 200 nM gene specific primer, and 2 mM premixed dNTPs was incubated at 65 °C for 5 min followed by incubation at 4 °C for 2 min. To this solution was added 9 µL of 2.22× MaP buffer (1× MaP buffer contains 1 M betaine, 50 mM Tris, pH 8.0, 75 mM KCl, 10 mM DTT, 6 mM $MnCl_2$), and the mixture was incubated at room temperature for 2 min. Reverse transcriptase (1 µL SuperScript II, Invitrogen) was added, and reverse transcription was performed using the following temperature program: 25 °C for 10 min,

42 °C for 90 min, 10 × [50 °C for 2 min, 42 °C for 2 min], 72 °C for 10 min. cDNA was then purified (Omega Mag Bind beads) and resuspended in H_2O .

Library preparation and sequencing

Sequencing libraries were prepared from cDNA products using a two-step PCR approach. In the first PCR step, 5-µL aliquots of purified cDNA were amplified for Illumina sequencing with the following temperature program: 98 °C for 30 s, 15 × [98 °C for 5 s, 68 °C for 20 s, 72 °C for 20 s], 72 °C for 2 min. The PCR products were recovered (Omega Mag-Bind beads) and eluted in water. In the second PCR step, treatment-specific barcodes were added to the ends of amplicons, with the following temperature program: 98 °C for 30 s, 10 × [98 °C for 5 s, 68 °C for 20 s, 72 °C for 20 s], 72 °C for 2 min. PCR products were recovered (Omega Mag-Bind beads), pooled, and sequenced. All libraries were sequenced on an Illumina MiSeq instrument using 300, 500, and 600 cycle kits for the TPP riboswitch, RNase P, and group II intron, respectively.

Sequence alignment, mutation parsing, and deconvolution of states

ShapeMapper (v2.1.5) was used to align reads to the reference sequences and to identify mutations created during MaP[44]. Default ShapeMapper settings were used to obtain profile.txt files, which are used downstream to analyze per-nucleotide reactivity data and to compare TMO and DMS reactivities. The --per-read-histograms option was used to generate high-level TMO and DMS reactivity distributions. The --output-parsed-mutations option was employed to produce mutation data for future ensemble deconvolution, if needed.

DANCE deconvolution of TPP structure-switching RNA

Ensembles for the structure-switching construct of the TPP riboswitch were deconvoluted with DanceMapper (v1.0)[31]. DANCE clustering was limited to a maximum of four components (—maxc 4); in practice, DanceMapper identified two-state models as the best fit for experiments performed in the presence and absence of TPP. Structure cassette sequences were ignored and the first seven positions of the native riboswitch sequence were also ignored, to maintain sequence alignment with the simplified TPP riboswitch aptamer sequence[29].

Structure modeling

ShapeKnots (based on software package RNAstructure, v6.1) was used to model the secondary structures of the TPP aptamer, RNase P and group II intron RNAs[8,45]. Required parameters were the RNA primary sequences and the output.ct file name. Pseudo-free energy restraints

were generated with the flag -dmsnt to use TMO- or DMS-normalized reactivities and the flag -x to incorporate PairMapper .bp files, thus introducing two TMO- or DMS-based bonuses (per-nucleotide and pairwise correlation) into the secondary structure calculation.[43] The -m 1 option was used to select only the minimum free energy structure, outputting a .ct file containing the secondary structure model. foldClusters.py (part of DanceMapper) was used to model MFE structures (default options) for the structure-switching TPP riboswitch RNA.

T-site and mutant reactivity analysis

Differential reactivity was assessed using reactivity profiles output by the ShapeMapper program. Per-nucleotide reactivity differences were computed by subtracting TMO reactivities from DMS reactivities. Ratios were computed similarly with TMO reactivities divided by DMS reactivities. Per-nucleotide reactivity profiles were visualized using RNavigate[46]. Three-dimensional structures of each RNA construct were visualized using PyMol (pymol.org). Individual motifs detected by differential probing were superimposed using the PyMol pair fitting function to identify conserved features. T-sites were defined as adenosine positions having ≥0.2 absolute reactivity difference between TMO and DMS, and a reactivity ratio of 2.5-fold for TMO versus DMS treatment, based on raw mutation rates. To identify the effect of T-site mutations, significant changes were defined as nucleotide positions having ≥0.05 absolute TMO reactivity difference between the mutant and native sequence RNA.

DENV2 RNA structure probing

DENV2 (Dengue virus 2 New Guinea C; full name: Dengue virus 2 Thailand/NGS-C/1944) viral particles were purified from tissue culture supernatants from infected C6/36 cells as described[47]. Briefly, infected supernatant was layered over a 20% sucrose solution followed by ultracentrifugation. Virion pellets were then resuspended in 100 mM HEPES, pH 7.9 and 50 mM NaCl, followed by centrifugation. The DENV2 virion pellet was resuspended in 100 mM HEPES, pH 7.9 and RNA was extracted from purified virions (TRIzol LS reagent; ThermoFisher Scientific). The DENV2 RNA was then refolded in 10 mM $MgCl_2$, 100 mM NaCl, 0.33 M Bicine, pH 8 and subjected to T-site chemistry and MaP-RT as described above. cDNA from MaP-RT was subjected to second strand synthesis (non-directional second strand synthesis; NEBNext), followed by library preparation (Ultra II library prep kit; NEBNext). Two replicates, performed on separate days, were obtained and the same T-sites were observed in each experiment.

Data Availability

Sequencing data are available from the Gene Expression Omnibus (GEO) at GSE286293.

For review, data are immediately available at

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE286293

using reviewer token: svibokoajjefnob

Supporting Information

One table of RNA sequences and six figures showing additional features of T-site chemistry.

Disclosure

K.M.W. is a founder at ForagR Medicines, Ribometrix and A-Form Solutions.

Acknowledgements

We are indebted to Amy Aponte Santana for insightful discussions. This work was supported by the US National Science Foundation (MCB-2027701 to K.M.W.), National Institutes of Health grants R01AI125416 (S.M.H) and T32GM142605 (C.J.A), and a National Science Foundation Graduate Research Fellowship (C.J.A).

References

(1) Cech, T. R. Crawling Out of the RNA World. *Cell* **2009**, *136* (4), 599–602. https://doi.org/10.1016/j.cell.2009.02.002.

(2) Sharp, P. A. The Centrality of RNA. *Cell* **2009**, *136* (4), 577–580. https://doi.org/10.1016/j.cell.2009.02.007.

(3) Mustoe, A. M.; Corley, M.; Laederach, A.; Weeks, K. M. Messenger RNA Structure Regulates Translation Initiation: A Mechanism Exploited from Bacteria to Humans. *Biochemistry* **2018**, *57* (26), 3537–3539. https://doi.org/10.1021/acs.biochem.8b00395.

(4) Jaafar, Z. A.; Kieft, J. S. Viral RNA Structure-Based Strategies to Manipulate Translation. *Nat. Rev. Microbiol.* **2019**, *17* (2), 110–123. https://doi.org/10.1038/s41579-018-0117-x.

(5) Strobel, E. J.; Yu, A. M.; Lucks, J. B. High-Throughput Determination of RNA Structures. *Nat. Rev. Genet.* **2018**, *19* (10), 615–634. https://doi.org/10.1038/s41576-018-0034-x.

(6) Weeks, K. M. SHAPE Directed Discovery of New Functions in Large RNAs. *Acc. Chem. Res.* **2021**, *54* (10), 2502–2517. https://doi.org/10.1021/acs.accounts.1c00118.

(7) Arney, J. W.; Laederach, A.; Weeks, K. M. Visualizing RNA Structure Ensembles by Single-Molecule Correlated Chemical Probing. *Curr. Opin. Struct. Biol.* **2024**, *88*, 102877. https://doi.org/10.1016/j.sbi.2024.102877.

(8) Hajdin, C. E.; Bellaousov, S.; Huggins, W.; Leonard, C. W.; Mathews, D. H.; Weeks, K. M. Accurate SHAPE-Directed RNA Secondary Structure Modeling, Including Pseudoknots. *Proc. Natl. Acad. Sci.* **2013**, *110* (14), 5498–5503. https://doi.org/10.1073/pnas.1219988110.

(9) Bevilacqua, P. C.; Ritchey, L. E.; Su, Z.; Assmann, S. M. Genome-Wide Analysis of RNA Secondary Structure. *Annu. Rev. Genet.* **2016**, *50*, 235–266. https://doi.org/10.1146/annurev-genet-120215-035034.

(10) Wayment-Steele, H. K.; Kladwang, W.; Strom, A. I.; Lee, J.; Treuille, A.; Becka, A.; Eterna Participants; Das, R. RNA Secondary Structure Packages Evaluated and Improved by High-Throughput Experiments. *Nat. Methods* **2022**, *19* (10), 1234–1242. https://doi.org/10.1038/s41592-022-01605-0.

(11) Rausch, J. W.; Sztuba-Solinska, J.; Le Grice, S. F. J. Probing the Structures of Viral RNA Regulatory Elements with SHAPE and Related Methodologies. *Front. Microbiol.* **2018**, *8*, 2634. https://doi.org/10.3389/fmicb.2017.02634.

(12) Spitale, R. C.; Incarnato, D. Probing the Dynamic RNA Structurome and Its Functions. *Nat. Rev. Genet.* **2023**, *24* (3), 178–196. https://doi.org/10.1038/s41576-022-00546-w.

(13) Metkar, M.; Pepin, C. S.; Moore, M. J. Tailor Made: The Art of Therapeutic mRNA Design. *Nat. Rev. Drug Discov.* **2024**, *23* (1), 67–83. https://doi.org/10.1038/s41573-023-00827-x.

(14) Brion, P.; Westhof, E. Hierarchy and Dynamics of Rna Folding. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26* (1), 113–137. https://doi.org/10.1146/annurev.biophys.26.1.113.

(15) Vicens, Q.; Kieft, J. S. Thoughts on How to Think (and Talk) about RNA Structure. *Proc. Natl. Acad. Sci.* **2022**, *119* (17), e2112677119. https://doi.org/10.1073/pnas.2112677119.

(16) Solayman, M.; Litfin, T.; Singh, J.; Paliwal, K.; Zhou, Y.; Zhan, J. Probing RNA Structures and Functions by Solvent Accessibility: An Overview from Experimental and Computational Perspectives. *Brief. Bioinform.* **2022**, *23* (3), bbac112. https://doi.org/10.1093/bib/bbac112.

(17) Graveley, B. R. RNA Matchmaking: Finding Cellular Pairing Partners. *Mol. Cell* **2016**, *63* (2), 186–189. https://doi.org/10.1016/j.molcel.2016.07.001.

(18) Pyle, A. M.; Barton, J. K. Probing Nucleic Acids with Transition Metal Complexes. In *Progress in Inorganic Chemistry*; John Wiley & Sons, Ltd, 1990; pp 413–475. https://doi.org/10.1002/9780470166390.ch7.

(19) Chin, K.; Sharp, K. A.; Honig, B.; Pyle, A. M. Calculating the Electrostatic Properties of RNA Provides New Insights into Molecular Interactions and Function. *Nat. Struct. Biol.* **1999**, *6* (11), 1055–1061. https://doi.org/10.1038/14940.

(20) Draper, D. E. A Guide to Ions and RNA Structure. *RNA N. Y. N* **2004**, *10* (3), 335–343. https://doi.org/10.1261/rna.5205404.

(21) Ehrhardt, J. E.; Weeks, K. M. Time-Resolved, Single-Molecule, Correlated Chemical Probing of RNA. *J. Am. Chem. Soc.* **2020**, *142* (44), 18735–18740. https://doi.org/10.1021/jacs.0c06221.

(22) Siegfried, N. A.; Busan, S.; Rice, G. M.; Nelson, J. A. E.; Weeks, K. M. RNA Motif Discovery by SHAPE and Mutational Profiling (SHAPE-MaP). *Nat. Methods* **2014**, *11* (9), 959–965. https://doi.org/10.1038/NMETH.3029.

(23) Serganov, A.; Polonskaia, A.; Phan, A. T.; Breaker, R. R.; Patel, D. J. Structural Basis for Gene Regulation by a Thiamine Pyrophosphate-Sensing Riboswitch. *Nature* **2006**, *441* (7097), 1167–1171. https://doi.org/10.1038/nature04740.

(24) Kazantsev, A. V; Krivenko, A. A.; Pace, N. R. Mapping Metal-Binding Sites in the Catalytic Domain of Bacterial RNase P RNA. *RNA N. Y. N* **2009**, *15* (2), 266–276. https://doi.org/10.1261/rna.1331809.

(25) Marcia, M.; Pyle, A. M. Visualizing Group II Intron Catalysis through the Stages of Splicing. *Cell* **2012**, *151* (3), 497–507. https://doi.org/10.1016/j.cell.2012.09.033.

(26) Westhof, E.; Fritsch, V. RNA Folding: Beyond Watson–Crick Pairs. *Structure* **2000**, *8* (3), R55–R65. https://doi.org/10.1016/S0969-2126(00)00112-X.

(27) Olson, W. K.; Li, S.; Kaukonen, T.; Colasanti, A. V; Xin, Y.; Lu, X.-J. Effects of Noncanonical Base Pairing on RNA Folding: Structural Context and Spatial Arrangements of G·A Pairs HHS Public Access. *Biochemistry* **2019**, *58* (20), 2474–2487. https://doi.org/10.1021/acs.biochem.9b00122.

(28) Roth, A.; Breaker, R. R. The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Annu. Rev. Biochem.* **2009**, *78*, 305–334. https://doi.org/10.1146/annurev.biochem.78.070507.135656.

(29) Winkler, W.; Nahvi, A.; Breaker, R. R. Thiamine Derivatives Bind Messenger RNAs Directly to Regulate Bacterial Gene Expression. *Nature* **2002**, *419* (6910), 952–956. https://doi.org/10.1038/nature01145.

(30) Mustoe, A. M.; Weidmann, C. A.; Weeks, K. M. Single-Molecule Correlated Chemical Probing: A Revolution in RNA Structure Analysis. *Acc. Chem. Res.* **2023**, *56* (7), 763–775. https://doi.org/10.1021/acs.accounts.2c00782.

(31) Olson, S. W.; Turner, A.-M. W.; Arney, J. W.; Saleem, I.; Weidmann, C. A.; Margolis, D. M.; Weeks, K. M.; Mustoe, A. M. Discovery of a Large-Scale, Cell-State-Responsive Allosteric Switch in the 7SK RNA Using DANCE-MaP. *Mol. Cell* **2022**, *82* (9), 1708-1723.e10. https://doi.org/10.1016/j.molcel.2022.02.009.

(32) Gebhard, L. G.; Filomatori, C. V.; Gamarnik, A. V. Functional RNA Elements in the Dengue Virus Genome. *Viruses* **2011**, *3* (9), 1739–1756. https://doi.org/10.3390/v3091739.

(33) Akiyama, B. M.; Laurence, H. M.; Massey, A. R.; Costantino, D. A.; Xie, X.; Yang, Y.; Shi, P.-Y.; Nix, J. C.; Beckham, J. D.; Kieft, J. S. Zika Virus Produces Noncoding RNAs Using a Multi-Pseudoknot Structure That Confounds a Cellular Exonuclease. *Science* **2016**, *354* (6316), 1148–1152. https://doi.org/10.1126/science.aah3963.

(34) Akiyama, B. M.; Graham, M. E.; O′Donoghue, Z.; Beckham, J. D.; Kieft, J. S. Three-Dimensional Structure of a Flavivirus Dumbbell RNA Reveals Molecular Details of an RNA Regulator of Replication. *Nucleic Acids Res.* **2021**, *49* (12), 7122–7138. https://doi.org/10.1093/nar/gkab462.

(35) Pijlman, G. P.; Funk, A.; Kondratieva, N.; Leung, J.; Torres, S.; Van Der Aa, L.; Liu, W. J.; Palmenberg, A. C.; Shi, P.-Y.; Hall, R. A.; Khromykh, A. A. A Highly Structured, Nuclease-Resistant, Noncoding RNA Produced by Flaviviruses Is Required for Pathogenicity. *Cell Host Microbe* **2008**, *4* (6), 579–591. https://doi.org/10.1016/j.chom.2008.10.007.

(36) Alvarez, D. E.; De Lella Ezcurra, A. L.; Fucito, S.; Gamarnik, A. V. Role of RNA Structures Present at the 3′UTR of Dengue Virus on Translation, RNA Synthesis, and Viral Replication. *Virology* **2005**, *339* (2), 200–212. https://doi.org/10.1016/j.virol.2005.06.009.

(37) Boerneke, M. A.; Gokhale, N. S.; Horner, S. M.; Weeks, K. M. Structure-First Identification of RNA Elements That Regulate Dengue Virus Genome Architecture and Replication. *Proc. Natl. Acad. Sci.* **2023**, *120* (15). https://doi.org/10.1073/pnas.2217053120.

(38) Homan, P. J.; Favorov, O. V.; Lavender, C. A.; Kursun, O.; Ge, X.; Busan, S.; Dokholyan, N. V.; Weeks, K. M. Single-Molecule Correlated Chemical Probing of RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (38), 13858–13863. https://doi.org/10.1073/PNAS.1407306111.

(39) Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principles for Targeting RNA with Drug-like Small Molecules. *Nat. Rev. Drug Discov.* **2018**, *17* (8), 547–558. https://doi.org/10.1038/nrd.2018.93.

(40) Veenbaas, S. D.; Koehn, J. T.; Irving, P. S.; Lama, N. N.; Weeks, K. M. Ligand-Binding Pockets in RNA and Where to Find Them. *Proc. Natl. Acad. Sci.* **2025**, *122* (17), e2422346122. https://doi.org/10.1073/pnas.2422346122.

(41) Jensen, J. H.; Kromann, J. C. The Molecule Calculator: A Web Application for Fast Quantum Mechanics-Based Estimation of Molecular Properties. *J. Chem. Educ.* **2013**, *90* (8), 1093–1095. https://doi.org/10.1021/ed400164n.

(42) Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE): Quantitative RNA Structure Analysis at Single Nucleotide Resolution. *Nat. Protoc.* **2006**, *1* (3), 1610–1616. https://doi.org/10.1038/nprot.2006.249.

(43) Mustoe, A. M.; Lama, N. N.; Irving, P. S.; Olson, S. W.; Weeks, K. M. RNA Base-Pairing Complexity in Living Cells Visualized by Correlated Chemical Probing. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (49), 24574–24582. https://doi.org/10.1073/pnas.1905491116.

(44) Busan, S.; Weeks, K. M. Accurate Detection of Chemical Modifications in RNA by Mutational Profiling (MaP) with ShapeMapper 2. *RNA* **2018**, *24* (2), 143–148. https://doi.org/10.1261/rna.061945.117.

(45) Reuter, J. S.; Mathews, D. H. RNAstructure: Software for RNA Secondary Structure Prediction and Analysis. *BMC Bioinformatics* **2010**, *11* (1), 129. https://doi.org/10.1186/1471-2105-11-129.

(46) Irving, P. S.; Weeks, K. M. RNAvigate: Efficient Exploration of RNA Chemical Probing Datasets. *Nucleic Acids Res.* **2024**, *52* (5), 2231–2241. https://doi.org/10.1093/nar/gkae089.

(47) Jensen, S. M.; Nguyen, C. T.; Jewett, J. C. A Gradient-Free Method for the Purification of Infective Dengue Virus for Protein-Level Investigations. *J. Virol. Methods* **2016**, *235*, 125–130. https://doi.org/10.1016/j.jviromet.2016.05.017.