Fingerprinting tertiary structure in complex RNAs using single-molecule correlated chemical probing

Ana C. Tan[1], Patrick S. Irving[1], Jordan T. Koehn, Shouhong Jin, David Y. Qiu and Kevin M. Weeks*
Department of Chemistry, University of North Carolina, Chapel Hill NC 28599-3290

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

Abstract

Single-molecule correlated chemical probing (smCCP) is an experimentally concise strategy for characterizing higher-order structural interactions in RNA. smCCP data yield rich, but complex, information on base pairing, conformational ensembles, and tertiary interactions. To date, through-space communication specifically measuring RNA tertiary structure has been difficult to isolate from structural communication reflective of other interactions. Here we introduce mutual information as a filtering metric to isolate tertiary structure communication contained within smCCP data and use this strategy to characterize the structural ensemble of the SAM-III riboswitch. We identified a smCCP fingerprint that is selective for states containing tertiary structure that forms concurrently with cognate ligand binding. We then successfully applied mutual information filters to independent RNAs and isolated through-space tertiary interactions in riboswitches and large RNAs with complex structures. smCCP, coupled with mutual information criteria, can now be used as a tertiary structure discovery tool, including to identify specific states in an ensemble that have higher-order structure. These studies pave the way for use of the straightforward smCCP experiment for discovery and characterization of tertiary structure motifs in complex RNAs.

Introduction

As soon as RNA molecules are synthesized, most begin folding back on themselves to form internal structures, involving both simple base paired elements and complex motifs potentially involving through-space tertiary interactions. Specific regions can also populate multiple states, forming ensembles. These internal structures play broad roles in governing RNA biology. Single-molecule correlated chemical probing (smCCP) is emerging as a powerful strategy for directly measuring RNA structure and conformational ensembles in complex RNAs (1, 2).

The smCCP strategy uses mutational profiling (MaP) to encode, as mutations in cDNA sequencing libraries, the multiple chemical adducts that occur in each individual RNA strand in a chemical probing experiment (Scheme 1A) (3). Standard MaP experiments measure population-average per-nucleotide modification rates and identify local nucleotide flexibility and averaged base-pairing status. In contrast, smCCP measures *per-molecule* chemical modification patterns and relationships to identify through-space structural communication within an RNA (Scheme 1B). smCCP data are dense with structural information, but complex to interpret.

smCCP data are optimally interpreted within frameworks that select for and identify a specific class of RNA structural communication. For example, individual base pairs (called PAIRs) in an RNA molecule can be directly measured using PAIR-MaP (pairs ascertained from interacting RNA strands measured by mutational profiling) which provides direct measurements of base pairing interactions, and facilitates highly accurate secondary structure modeling (4). In principle, non-canonical and tertiary interactions (RNA interaction groups, or RINGs) can be mapped using RING-MaP (3). In current implementations, RING-MaP measurements are dense, complex, and have proven difficult to confidently assign to tertiary structure in the context of convoluting signals from other classes of structural communication. DANCE-MaP (deconvolution and annotation of nucleic acid ensembles), and independently developed frameworks, identify the dominant states of an RNA structural ensemble (5–7). DANCE-MaP can assign individual sequencing reads to one of the states of the ensemble, allowing further smCCP analysis of PAIRs and RINGs in each RNA state independently.

In principle, the RING-MaP experiment enables discovery and analysis of novel RNA tertiary structures, including in cells. However, analysis of a RING-MaP experiment is complex. Reflecting these complexities, prior investigations of RNA structure using RING-MaP have used

a variety of heuristic filters, based on underlying secondary structure and the significance of correlations between chemical modification events, to identify interactions likely to be linked to true tertiary structure formation (3, 5, 8, 9). Moreover, measures of correlation significance employed to date are dependent on the number of sequencing reads collected in an experiment. This read depth dependence makes it difficult to compare RINGs between experiments, or between states identified in an ensemble-deconvolution experiment, which often contain different numbers of sequencing reads. Currently, read depth and significance thresholds cannot be determined without prior knowledge involving comparison to a known structure and no single set of criteria has been validated that applies universally to RNAs of unknown structure and complexity. Here, we describe an improved RING filtering strategy and its use in characterizing the structural ensemble of the SAM-III riboswitch. These new filters, based on mutual information, facilitate broad application of the smCCP experiment for discovery and analysis of RNA tertiary structure (Scheme 1C).

We initially used the SAM-III riboswitch (10–12) to develop a filtering strategy for assigning RING correlations to tertiary structure domains in complex RNAs. Bacterial riboswitches are structural domains of mRNAs that enable self-regulation of gene expression through sensing of cellular metabolites. Coexisting riboswitch RNA conformations function in a rheostat-like way to either favor or disfavor gene expression, and binding by a cognate ligand influences the relative abundances of these conformations (13). S-adenosylmethionine (SAM) binds to the SAM-III riboswitch to regulate translation of the bacterial SAM synthetase (10, 14). The riboswitch samples two secondary structure conformations that exist in equilibrium: a translation ON state that presents the Shine-Dalgarno sequence in a ribosome-available single-stranded region and a translation OFF state that sequesters the Shine-Dalgarno sequence in a base-paired helix (Figure 1A). SAM binding to the SAM-III riboswitch shifts this equilibrium to more strongly favor the OFF state. Here, we apply chemical probing and smCCP frameworks to model the conformational ensemble of the SAM-III riboswitch. We use DANCE-MaP and RING-MaP data in combination to develop read-depth-agnostic mutual information-based filters to isolate correlations reflective of true tertiary structure (Scheme 1C). We then show that this framework identifies tertiary interactions broadly, in the adenine and TPP riboswitches, and in a large bacterial RNase P RNA.

Results

DANCE-MaP distinguishes two conformations of the SAM-III riboswitch

The SAM-III riboswitch forms two conformations in equilibrium. Upon binding by SAM, the SAM-III ensemble equilibrium shifts towards the translation OFF state (Figure 1A) (14). We obtained structure probing data for an *in vitro* transcribed construct of the SAM-III riboswitch, spanning the SAM-binding and expression domains, in the presence and absence of SAM, using dimethyl sulfate (DMS). We modeled the (average) secondary structures for each state using base-pairing bonuses derived from per-nucleotide DMS-MaP reactivities and internucleotide PAIRs (4). In the absence of SAM ligand, the modeled structure is identical to the accepted (11) ON state (Figure 1B, *left*). With SAM present, the modeled structure of the SAM-III riboswitch resembles the OFF state but, as compared to the accepted structure, an incorrect extra helix is present (Figure 1B, *right*). Major features differentiating each state are specifically supported by direct PAIR measurements. This conventional analysis, relying on population-average measurements, thus yields plausible secondary structure models but, nonetheless, gives no indication that multiple conformations might be present in each experiment, nor that any state is comprised of both base pairing and tertiary interactions.

We next used DANCE-MaP analysis to deconvolute the same DMS structure probing data. DANCE analysis reports whether the per-molecule mutation patterns identified by MaP likely arose from a single underlying structure state or from multiple states. DANCE-MaP provides populations, per-nucleotide reactivities, PAIRs, and RINGs for each identified structure state (1, 5). For the SAM-III riboswitch native sequence, DANCE identified two states, both in the absence and presence of SAM (Figure 2). Again, we modeled structures using per-nucleotide reactivities and PAIRs, this time for each structural state in the ensemble identified by DANCE. Both the ON and OFF states were populated under all conditions, emphasizing that the SAM riboswitch functions as a rheostat, rather than a switch. In the absence of SAM, the two states were present in roughly equal populations (Figure 2A). In the presence of SAM, the translation OFF state predominated (86%) (Figure 2B). These ON-OFF state ratios are consistent with independent studies of the SAM-III riboswitch (15).

To further validate DANCE-MaP modeling, we designed four SAM-III mutant constructs that isolate a single structural state. Two mutants lock the riboswitch in the translation ON state and two lock it in the translation OFF state. The two mutants locking the RNA in each conformation were based on distinct rationales: one was designed to stabilize the target state and one designed to destabilize the competing state (Figure 3). Construct designs included minimal

changes to the native sequence and avoided changing nucleotides involved in stabilizing higher-order RNA structure or involved in direct interactions with the SAM ligand. For each construct, DANCE-MaP identified a single conformation, corresponding to the desired structure, and this structure formed both in the presence and absence of SAM (Figure 3). DANCE-MaP thus correctly characterizes both multi-state ensembles and well-defined single-conformation RNAs.

Identification of RNA tertiary structure using smCCP

We next focused on identifying a universal set of RING filters that isolate close-in-space tertiary interactions. An optimal filtering strategy would distinguish close-in-space correlations from all measured inter-nucleotide correlations (Figure 4A). Existing strategies generally assign RINGs to tertiary interactions if they involve single-stranded nucleotides within a stringent minimum contact distance (the shortest path distance in the secondary structure graph (16)), and if the measured correlation is significantly positive (mutations more likely than chance to co-occur on the same strand). Existing filters tend to be specific to a given system, yield variable numbers and quality of inter-nucleotide correlations, and are dependent on sequencing read depth. In some cases, filters have been optimized using the known structures of the modeled RNAs, and were overfit.

We explored alternative, more stable, strategies to filter for tertiary interactions, initially focusing on the MUT_{OFF1} mutant. We retained the requirements for nucleotides to be single-stranded and to be significantly positively correlated. We reduced the stringency of the minimum contact distance filter (from >15 to >5). We then applied two filters based on mutual information: raw mutual information, a measure of the amount of information shared between nucleotides, and the Z-score of mutual information, a measure relative to other RINGs computed at the same nucleotides (Scheme 1).

For filters relying on correlation significance only, the number, pattern, and usefulness of correlations varied with sequencing read depth. For the SAM-III riboswitch, the best performance using significance-based filters was achieved at a read depth of 800,000, and then became worse at higher read depths, to include correlations occurring between nucleotides located more than 25 Å apart (Figure 4B).

The updated mutual information-based filtering strategy (Scheme 1C) yielded a similar pattern of correlations, once a read depth of 200,000 or greater was achieved (Figure 4C). The pattern of RING correlations remained stable as read depth increased from 200,000 to 2,000,000 reads and is consistent with the accepted structure of the ligand-bound riboswitch (Figure 4C). This mutual information framework for selecting RING correlations provides accurate and stable insight into RNA folding and tertiary structure.

RING filters identify a structural fingerprint for ligand binding to the SAM-III riboswitch

We next used the mutual information RING filters framework to analyze smCCP data for the native sequence SAM-III riboswitch and for the four mutant constructs. For the native SAM-III sequence, we analyzed the ON and OFF states independently giving, together with the four single-state mutants, six total states (Figure 5). We did not detect RING correlations when we applied our filters to the translation ON state of the native RNA and none were detected for either MUT_{ON} construct in the presence or absence of SAM (Figure 5, *left*). This lack of correlations is consistent with the expectation that the translation ON state does not have significant higher-order tertiary structure. In contrast, in the presence of the SAM ligand, multiple correlations were observed for the translation OFF state of the native construct and for both MUT_{OFF} constructs (Figure 5, *lower right*). These correlations connected nucleotides known to form the SAM binding site. Intriguingly, mutual information-based RING correlations were only observed in the presence of the SAM ligand, consistent with ligand binding consolidating tertiary structure interactions in the core of the riboswitch aptamer domain.

Generality of RING filters for direct assessment of RNA tertiary structure

We applied our newly devised RING filters to RNAs that vary in size and complexity, including the TPP riboswitch aptamer domain (86 nts), switching-competent constructs for both the adenine and TPP riboswitches (119 and 163 nts), and the catalytic domain of RNase P (265 nts). For each RNA, we assessed three RING visualizations (Figure 6). First, we computed all possible nucleotide pairs in the accepted structures that are adjacent in three-dimensional space and that have a contact distance of 6 or greater. These computed nucleotide pairs represent a best case for close-in-space correlations that can theoretically be detected as filtered RING correlations. Second, we show correlations observed using (prior) significance-based and (new to this work) mutual information-based filters. Note that the number of correlations observed using the significance-based metric is unstable and can degrade with increasing sequencing reads (Figure 4B).

For the SAM-III riboswitch, the new mutual information-based filters identified interactions surrounding the SAM binding pocket that were undetected using previous filters (Figure 6A). When we analyzed the relatively simple case of the TPP riboswitch aptamer domain alone (which cannot switch between states), the mutual information-based filters developed here outperformed significance-based filters, yielding both more and shorter-distance correlations, and are independent of read depth (Figure 6B).

We then examined the more challenging cases of adenine and TPP riboswitches, both in the context of flanking sequences that allow each RNA to sample multiple states (their ON and OFF states). We deconvoluted the ensembles via DANCE analysis. In both cases, we correctly identified a compact set of correlations that overlap in the ligand-binding regions of each RNA (Figure 6C,D). For the adenine riboswitch, the new filters did define a longer-range correlation that connects the adenine binding site to the loop-loop interaction, likely reflective of indirect communication between these features. This interaction was not observed for the significance-based correlations in the original report of these data (5), but a lower or higher read depth experiment would have likely produced a different, misleading result.

Most importantly, for all three switch-capable riboswitch constructs (Figure 6 A,C,D), RING correlations were only observed for a single state, that comprising the folded, tertiary structure containing, aptamer domain. DANCE, coupled with mutual information criteria, can thus be used as a tertiary structure discovery tool, including to identify specific states in an ensemble that have higher-order structure.

Results for the RNase P RNA were particularly impressive in that a large fraction of the theoretically detectable correlations were detected, these correlations occurred across the RNA structure, and all correlations were close in space (Figure 6E). We hypothesize that the new mutual information-based filters will prove especially useful for larger RNAs with greater tendency to exhibit interactions which are close in three-dimensional space but far in primary sequence.

Discussion

smCCP represents a new frontier for understanding the structure and dynamics of RNA molecules in solution (1, 2). Correlation data, derived from co-occurring chemical modifications, obtained across single RNA strands, yield rich and detailed information regarding higher-order interactions and conformational complexity in RNA structure (Scheme 1). Two uses of smCCP had been well developed: direct measurement of base pairing interactions (PAIRs) (4, 17) and deconvolution of conformational ensembles (DANCE and independent algorithms) (2, 5–7). A critical challenge in prior work was that criteria for measurement of tertiary interactions was *ad hoc* and study dependent, even though analysis of through-space interactions (RINGs) was one of the first applications of smCCP (3, 8, 9).

Our initial analysis of the SAM-III riboswitch revealed the current strengths and limitations of smCCP analysis. Probing the SAM-III RNA in the absence and presence of the SAM ligand yielded per-nucleotide reactivity profiles compatible with the single structures modeled for each state. This analysis did not provide any clues that the RNA populated two underlying states, nor that one of these states contained a complex tertiary fold. This is the usual situation. DANCE analysis readily resolved the two-state ensemble of the SAM-III riboswitch. The ON and OFF states are both present in the presence and absence of SAM and the equilibrium is shifted toward the translation OFF state in presence of the cognate ligand. This successful deconvolution still missed clues that one conformation also harbored a well-defined tertiary structure.

Decisive detection of interactions clearly indicative of RNA tertiary structure has proven challenging as correlation significance is dependent on the number of sequencing reads collected in an experiment. Here, we developed a new framework for detecting through-space RING interactions, based on mutual information, that specifically reflect higher-order RNA structure. The new framework reduces both dependence on sequence read depth and on the read depth required to detect RINGs reflective of higher-order interactions. The RING filtering strategy was optimized with a simple internal control, as the ON and OFF state RNAs have the same sequence but only the OFF state has a higher-order tertiary structure. The resulting smCCP experiment remains remarkably simple to perform and can be implemented at modest sequencing depth.

We optimized the RING filtering framework on a single switching RNA, the SAM-III riboswitch. Extending RING smCCP analysis to five highly structured RNAs, ranging in size from 86 to 265

nucleotides, revealed that mutual information filters detect tertiary structure for both multi-state and large RNA systems. smCCP is now poised to enable efficient investigation of higher-order structural communication in a broad range of RNAs. This work sets the stage for using smCCP and the RING approach for *de novo* scanning of large RNA regions for likely internal tertiary structures.

Materials and Methods

In vitro transcription of the SAM-III RNA riboswitch, mutants, and other RNAs

In vitro transcription was performed as described (4, 5). Briefly, templates containing our native sequence or mutant designs, flanked by 5' and 3' structure cassettes (18), were synthesized (gBlocks; IDT), amplified by PCR (Q5 DNA polymerase), purified (PCR PureLink Micro-Kit; Invitrogen), transcribed *in vitro*, and purified (RNeasy MinElute columns; Qiagen). Template and primer sequences are provided in Table S1. RNA size and purity were confirmed (Bioanalyzer RNA 6000 Nano Kit; Agilent). Concentrations were determined by UV absorbance (NanoDrop spectrophotometer; ThermoFisher).

Design of SAM-III riboswitch mutants

Mutant constructs were designed to populate each of the accepted secondary structures for the ON and OFF states of the SAM-III riboswitch. *RNAstructure* (19) was used to model minimum free energy secondary structures, and to guide design of structural mutants. For the designed constructs, the calculated ΔG for the desired state was required to differ from that of the alternative state by at least 2-4 kcal/mol.

DMS probing and MaP reverse transcription

RNA constructs were denatured at 95 ˚C for 2 minutes, placed on ice, and then refolded [9 µL; 4 pmol RNA, 256 mM Bicine (pH 8.0), 4.3 mM $MgCl_2$, 85.5 mM NaCl] at 37 ˚C for 30 minutes. After this folding step, either 1 µL of 0.1 mM SAM ligand or 1 µL of nuclease-free water were added, and the RNA was allowed to fold at 37 ˚C for 10 additional minutes. To the folded RNA was added either 1 µL of 1.7 M DMS in ethanol or 1 µL of ethanol, and samples were allowed to react at 37 ˚C for 6 minutes (final probing conditions were: 3.3 pmol RNA, 209 mM Bicine (pH 8.0), 3.5 mM $MgCl_2$, 70 mM NaCl, 9 µM or no SAM ligand, 0.15 M DMS in ethanol or 9% ethanol). Reactions were quenched with 10 µL 20% 2-mercaptoethanol, and the RNA was purified (RNeasy MinElute columns; Qiagen). MaP reverse transcription was performed exactly

as described (4, 5) using 370 ng of either the unmodified or DMS-modified RNA. The resulting cDNAs were purified (MicroSpin G-50 columns; Cytiva). The adenine and TPP riboswitch RNAs and the RNase P RNA were probed under closely related conditions. Final concentrations included Bicine, pH 8.0 (at 37 °C) and 3-10 mM $MgCl_2$ (5).

Sequencing library construction

Sequencing libraries were generated using a two-step PCR approach (20). Step 1 PCR uses overhanging primers to add part of the Illumina sequencing adapter. Step 2 PCR adds sample barcodes and the full sequencing adapter. For the SAM-III riboswitch, 5 µL of purified cDNA from reverse transcription was input to step 1 PCR [98 ˚C for 30 s, 20 cycles of (98 ˚C for 8 s, 68 ˚C for 25 s, 72 ˚C for 30 s), and 72 ˚C for 2 min]. Step 1 PCR products were purified (Mag-Bind TotalPure NGS beads, 0.8x volume ratio; Omega Bio-tek). Next, 2.5 ng of purified step 1 PCR product was input to step 2 PCR [98 ˚C for 30 s, 10 cycles of (98 ˚C for 8 s, 65 ˚C for 25 s, 72 ˚C for 30 s) and 72 ˚C for 2 min]. Step 2 PCR products were purified (Mag-Bind TotalPure NGS beads; 0.8x ratio). Resulting libraries were sequenced with an Illumina MiSeq instrument using 2x150 paired-end sequencing (v2 chemistry). Step 1 and Step 2 primer sequences are given in Table S1.

Detection of sequence duplicates

Paired-end reads were merged using FLASh (21) and 5 nucleotides were trimmed from the beginning and end of each merged read (UMI-tools extract, (22)). From this set of merged and trimmed reads, duplicate sequences were quantified with BBMap dedupe.sh using exact matches only (ac=f) (23). Observed duplicated sequences were less than 10% for all RNAs. Severe duplication in raw sequencing data negatively impacts downstream smCCP analysis. When sequencing amplicons with identical ends to read depths of hundreds of thousands or millions, some sequences will appear as duplicates when, in fact, they originate from distinct RNA molecules. Removing these sequence duplicates also negatively impacts downstream smCCP analysis. In this study, duplicate detection was used only for quality control purposes and only for DMS-modified samples from which smCCP calculations were made.

Population average smCCP analysis

Population average per-nucleotide reactivities as well as per-molecule mutation signatures were obtained by providing raw sequencing files to *ShapeMapper2* v2.1.5 (24). The primer binding regions were ignored (--primers) and parsed mutations files were obtained for smCCP analysis

(--output-parsed-mutations). The output files were provided to *PairMapper* to obtain DMS-normalized reactivity profiles and PAIRs (4). Predicted minimum free-energy structures and base-pairing probabilities were obtained from *RNAstructure partition* and *Fold* programs, informed by DMS reactivities and PAIRs (--dmsnt and -x, respectively) (19). Data were visualized using RNAvigate within a Jupyter notebook (25) (provided in the Supporting Information).

DANCE deconvolution analysis

For all switching-competent riboswitch RNAs, ensembles were deconvoluted with *DanceMapper*. Non-switching RNAs used population averaged data. DANCE clustering was limited to a maximum of three components (--maxcomponents 3); in practice, DanceMapper identified two state models as the best fit for all switching constructs and single state models for all SAM-III mutants constructs. Data obtained for the SAM-III riboswitch native sequence and the four mutant constructs were downsampled to 1.5 million reads for all DANCE model fitting (--undersample 1500000). All reads were used for *DanceMapper* PAIR, and RING analyses (--pairmap and –ring, respectively). We used foldClusters.py (part of *DanceMapper*) to model the MFE structure (default option) and pairing probabilities (--prob) of each component following DANCE deconvolution using both per-state reactivities (default input) and per-state PAIRs (--bp). The resulting output files define all states for a deconvoluted ensemble, and their respective pairing probabilities. Data for the adenine riboswitch and for RNase P were obtained from refs. (5) and (26). Datasets were analyzed as described for the SAM-III riboswitch but with different numbers of sequencing reads (approximately 500,000 for the adenine riboswitch, two million for RNase P, 650,000 for the TPP riboswitch switching construct, and 440,000 for the TPP riboswitch aptamer domain).

RING correlation analysis

The *RingMapper* framework was employed to compute correlated modification events (4, 5). A customized version of *RingMapper* (available on GitHub) was used that outputs full contingency tables for each RING correlation to allow calculation of mutual information (MI). The new filters used in this study are: contact distance greater than 5, correlation significance greater than 23, mutual information score greater than 0.000025, and Z-score greater than 2. Computation of mutual information score, filtering of RING correlations, and visualizations of arc plots, secondary structures, and 3D structures was executed using *RNAvigate* (see Jupyter notebooks in the Supporting Information) (25).

Software availability

All software used in this study are freely available from our lab webpage (weekslab.org) and from https://github.com/Weeks-UNC. Full data analysis for all figures is provided in the form of *RNAvigate* Jupyter notebooks (see SI).

Data availability

Sequencing data are accessible through Gene Expression Omnibus accession number GSE278422 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE278422).

Supporting Information

Table providing DNA template and primer sequences, plus six RNAvigate Jupyter notebooks (as separate files) giving full visualization and data analysis frameworks for all figures.

Acknowledgements

This work is part of the Undergraduate Transcriptome Project, supporting undergraduate-driven independent inquiry, funded by the US NSF (MCB-2027701). This work was also supported by the US NIH (R35 GM122532 to K.M.W.). J.T.K. is a UNC Lineberger Integrated Training in Cancer Model Systems Fellow (T32CA009156) and an NIH Kirschstein Postdoctoral Fellow (F32GM143863).

Disclosure

K.M.W. is an advisor to and holds equity in ForagR Medicines, A-Form Solutions and Ribometrix. J.T.K. is an advisor to and holds equity in ForagR Medicines. All other authors declare no competing interests.

References

1. Mustoe,A.M., Weidmann,C.A. and Weeks,K.M. (2023) Single-molecule correlated chemical probing: A revolution in RNA structure analysis. *Acc. Chem. Res.*, 10.1021/acs.accounts.2c00782.

2. Arney,J.W., Laederach,A. and Weeks,K.M. (2024) Visualizing RNA structure ensembles by single-molecule correlated chemical probing. *Curr Opin Struct Biol*, **88**, 102877.

3. Homan,P.J., Favorov,O.V., Lavender,C.A., Kursun,O., Ge,X., Busan,S., Dokholyan,N.V. and Weeks,K.M. (2014) Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci.*, **111**, 13858–13863.

4. Mustoe,A.M., Lama,N.N., Irving,P.S., Olson,S.W. and Weeks,K.M. (2019) RNA base-pairing complexity in living cells visualized by correlated chemical probing. *Proc. Natl. Acad. Sci.*, **116**, 24574–24582.

5. Olson,S.W., Turner,A.-M.W., Arney,J.W., Saleem,I., Weidmann,C.A., Margolis,D.M., Weeks,K.M. and Mustoe,A.M. (2022) Discovery of a large-scale, cell-state-responsive allosteric switch in the 7SK RNA using DANCE-MaP. *Mol. Cell*, **82**, 1708-1723.e10.

6. Tomezsko,P.J., Corbin,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. and Rouskin,S. (2020) Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature*, **582**, 438-2,442A-442Q.

7. Morandi,E., Manfredonia,I., Simon,L.M., Anselmi,F., van Hemert,M.J., Oliviero,S. and Incarnato,D. (2021) Genome-scale deconvolution of RNA structure ensembles. *Nat. Methods*, **18**, 249–252.

8. Dethoff,E.A., Boerneke,M.A., Gokhale,N.S., Muhire,B.M., Martin,D.P., Sacco,M.T., McFadden,M.J., Weinstein,J.B., Messer,W.B., Horner,S.M. and Weeks,K.M. (2018) Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl. Acad. Sci. U. S. A.*, **115**, 11513–11518.

9. Larman,B.C., Dethoff,E.A. and Weeks,K.M. (2017) Packaged and free satellite Tobacco Mosaic Virus (STMV) RNA genomes adopt distinct conformational states. *Biochemistry*, **56**, 2175–2183.

10. Fuchs,R.T., Grundy,F.J. and Henkin,T.M. (2006) The SMK box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nat. Struct. Mol. Biol.*, **13**, 226–233.

11. Lu,C., Smith,A.M., Fuchs,R.T., Ding,F., Rajashankar,K., Henkin,T.M. and Ke,A. (2008) Crystal structures of the SAM-III/SMK riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nat. Struct. Mol. Biol.*, **15**, 1076–1083.

12. Wang,J.X. and Breaker,R.R. (2008) Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochem. Cell Biol.*, **86**, 157–168.

13. Salvail,H. and Breaker,R.R. (2023) Riboswitches. *Curr. Biol.*, **33**, R343–R348.

14. Smith,A.M., Fuchs,R.T., Grundy,F.J. and Henkin,T.M. (2010) The SAM-responsive SMK box is a reversible riboswitch. *Mol. Microbiol.*, **78**, 1393–1402.

15. Wilson,R.C., Smith,A.M., Fuchs,R.T., Kleckner,I.R., Henkin,T.M. and Foster,M.P. (2011) Tuning riboswitch regulation through conformational selection. *J. Mol. Biol.*, **405**, 926–938.

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

17. Mitchell,D.,III, Cotter,J., Saleem,I. and Mustoe,A.M. (2023) Mutation signature filtering enables high-fidelity RNA structure probing at all four nucleobases with DMS. *Nucleic Acids Res.*, **51**, 8744–8757.

18. Wilkinson,K.A., Merino,E.J. and Weeks,K.M. (2006) Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.

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

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

21. Magoč,T. and Salzberg,S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

22. Smith,T.S., Heger,A. and Sudbery,I. (2017) UMI-tools: Modelling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.*, 10.1101/gr.209601.116.

23. Bushnell,B. (2022) BBMap v38.96. *https://sourceforge.net/projects/bbmap/*.

24. Busan,S. and Weeks,K.M. (2017) Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA*, 10.1261/rna.061945.117.

25. Irving,P.S. and Weeks,K.M. (2024) RNAvigate: efficient exploration of RNA chemical probing datasets. *Nucleic Acids Res.*, 10.1093/nar/gkae089.

26. Ehrhardt,J.E. and Weeks,K.M. (2020) Time-Resolved, Single-Molecule, Correlated Chemical Probing of RNA. *J. Am. Chem. Soc.*, **142**, 18735–18740.

Figure Legends

Scheme 1. Overview of (A) single-molecule correlated chemical probing (smCCP) and (B, C) classes of information that can be measured and the algorithmic frameworks for measuring each class.

Figure 1. Translation ON and OFF states for the SAM-III riboswitch, partially visualized by population-average, per-nucleotide probing. (A) Secondary structures for the ON and OFF states (11). SAM binding shifts the equilibrium to favor the OFF state, in which the Shine-Dalgarno sequence (SD) is sequestered in a base-paired region. Secondary structures are shown as both base pairing diagrams and arc plots. (B) Minimum free energy secondary structures modeled using population average per-nucleotide DMS-MaP reactivities and PAIRs. Histograms show per-nucleotide reactivities. Gray arcs above the axis and blue arcs below the axis show modeled base pairs and PAIRs, respectively.

Figure 2. DANCE deconvolution of the conformational ensemble for the SAM-III riboswitch in the absence (A) and presence (B) of SAM ligand. Secondary structure models are based on DMS-MaP and PAIR-MaP pseudo-free energy restraints (4). Histograms display DMS reactivity. Gray and blue arcs represent modeled base pairs and PAIRs, respectively.

Figure 3. Designs and conformations for the native sequence and single-state mutants of the SAM-III riboswitch. Mutated regions in each construct are highlighted. Table shows percent in each state, in the absence and presence of ligand, determined by DANCE.

Figure 4. Performance of RING filters. (A) Structure of SAM-III riboswitch (PDB ID: 3e5c) (11) with lines connecting pairs of single-stranded residues with contact distances greater than 5 (*left*) and the subset of these pairs that are less than 15 Å apart (*right*). Nucleotide pairs are linked at the DMS reactive atom of each nucleotide and are colored by distance. Spheres indicate structures in the conformation-switching RNA, not present in the crystalized RNA. Visualization of RINGs using (B) significance-based versus (C) mutual information-based strategies (see Scheme 1). For each read depth, RINGs are overlaid on the accepted structure; histograms quantify RINGs that passed significance or mutual information filters. Analyses were performed using data from the $MUT_{OFF\ 1}$ RNA probed in the presence of SAM.

Figure 5. Visualization of RINGs from native sequence (two states) and mutant constructs (one state each). (A) Filtered RINGs superimposed on the structure models for the ON and OFF states (PDB IDs 6c27 and 3e5c, respectively) (11) of the SAM-III riboswitch. Structures at top and bottom correspond to RINGs obtained in the absence and presence of ligand, respectively. The SAM ligand (blue) is shown in the SAM-bound OFF structures. Note: $MUT_{ON\ 1}$ and $MUT_{ON\ 2}$ were analyzed and had the same (lack of RING) correlations as the natively deconvoluted states, and are shown as single representative structures for simplicity. (B) Representative view of the ligand binding site and tertiary structure RINGs.

Figure 6. Performance of RING filters on independent RNAs. Performance shown for (A) SAM-III riboswitch (same data as in Figure 5), (B) TPP riboswitch aptamer domain; structure-switching constructs for the (C) adenine and (D) TPP riboswitches; and (E) the catalytic domain of a bacterial RNase P. In each series, the left panel shows pairwise combinations of single stranded, distant (contact distance >5), and close-in-space (<15 Å) nucleotides, corresponding to a best case RING-MaP result. The middle and right panels show RINGs identified with significance-based and mutual information-based filters, respectively.

Figure 1

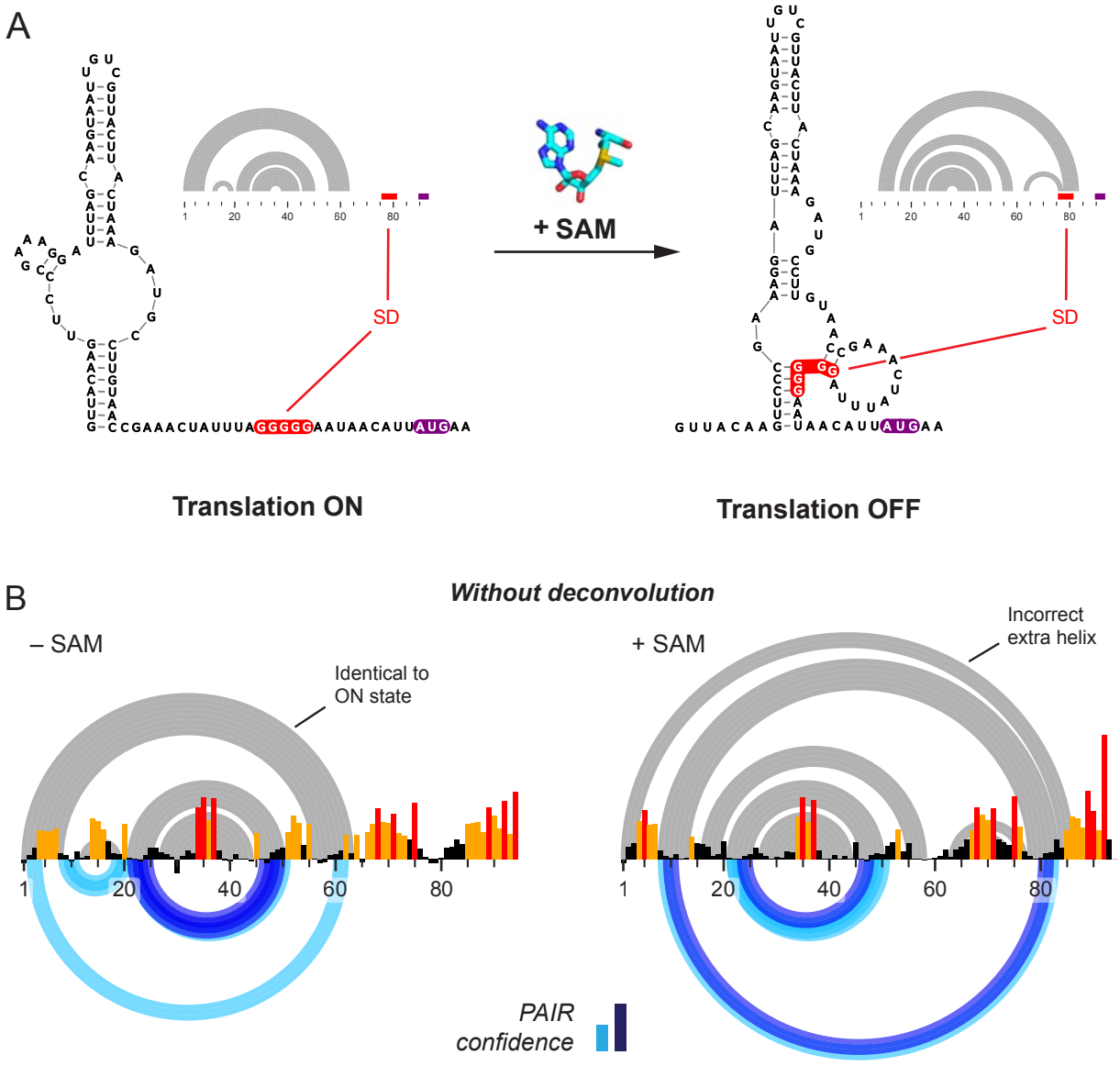

A
SD
+ SAM
SD
Translation ON
Translation OFF
B
Without deconvolution
– SAM
Identical to
ON state
+ SAM
Incorrect
extra helix
PAIR
confidence

Figure 2

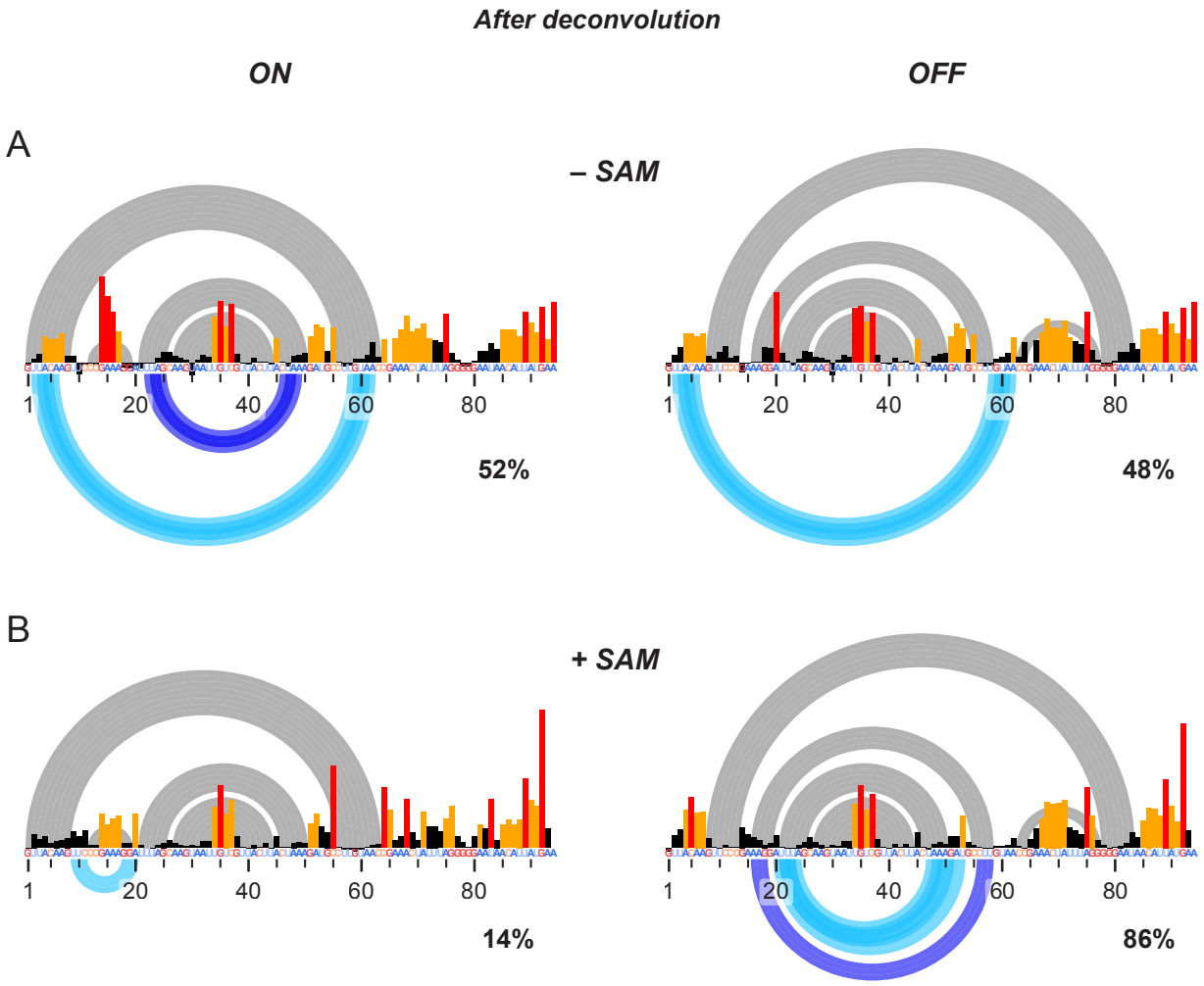
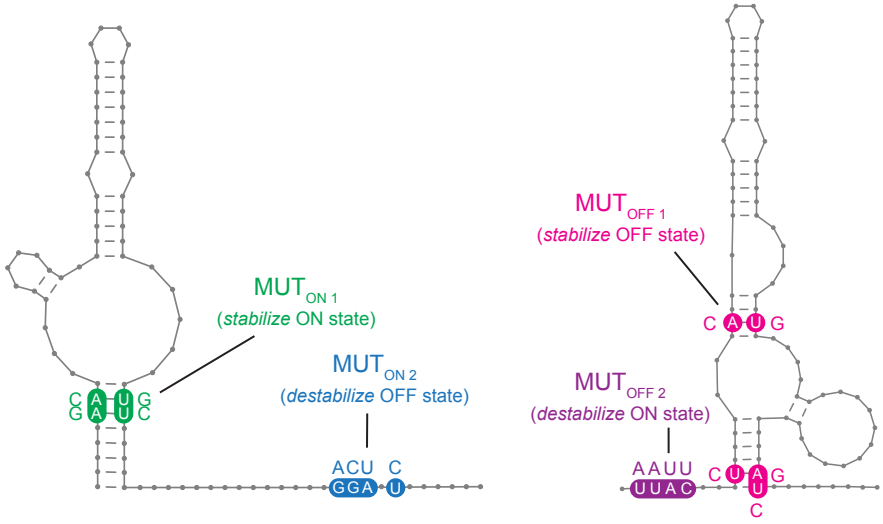


Figure 3



	SAM	ON state	OFF state
Native	–	52%	48
	+	14%	86
MUT ON 1	–	100	
	+	100	
MUT ON 2	–	100	
	+	100	
MUT OFF 1	–		100
	+		100
MUT OFF 2	–		100
	+		100

(print at single column width)

Figure 4

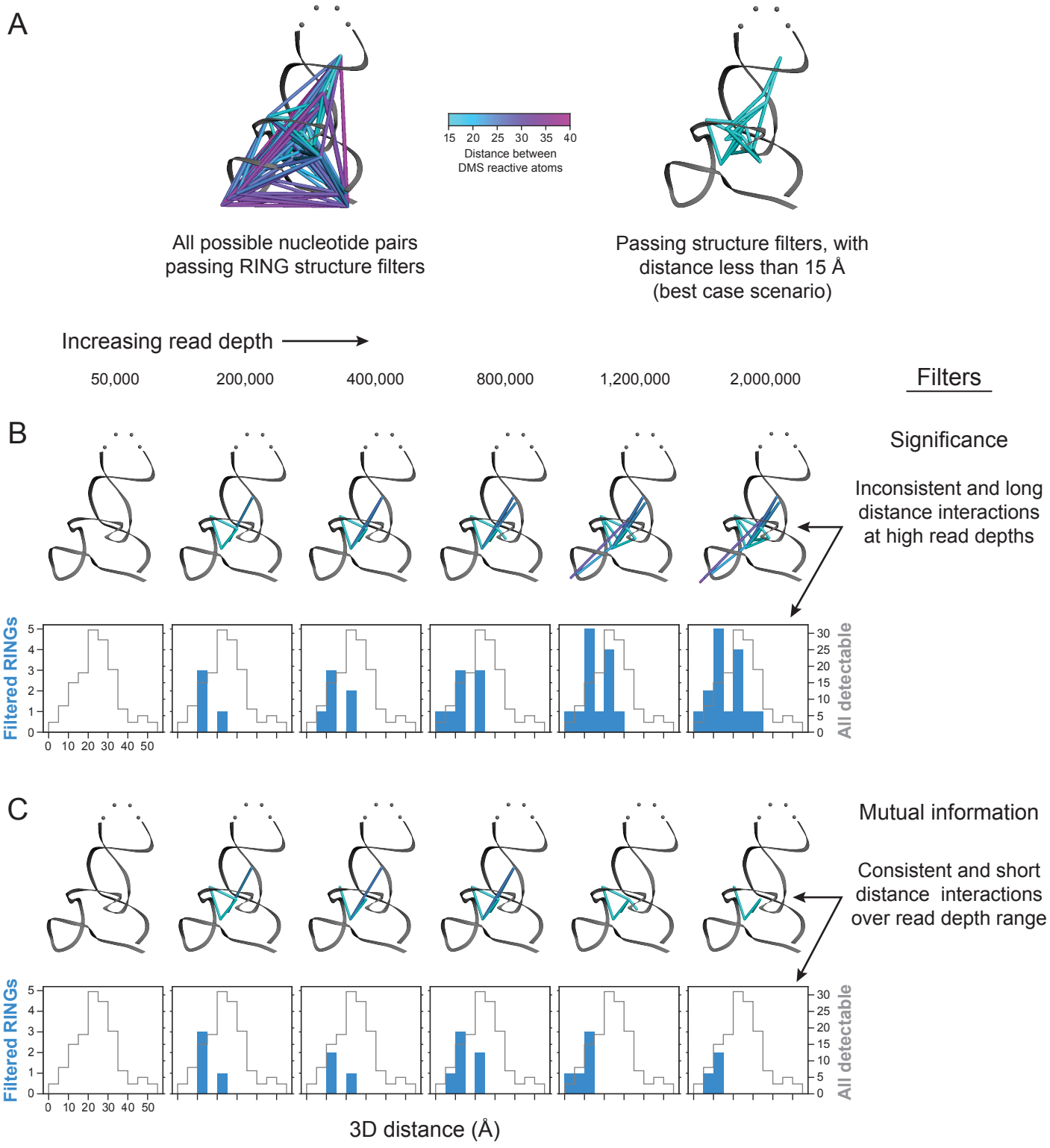

A
15 20 25 30 35 40
Distance between
DMS reactive atoms
All possible nucleotide pairs
passing RING structure filters
Passing structure filters, with
distance less than 15 Å
(best case scenario)
Increasing read depth
50,000
200,000
400,000
800,000
1,200,000
2,000,000
Filters
B
Significance
Inconsistent and long
distance interactions
at high read depths
Filtered RINGs
All detectable
C
Mutual information
Consistent and short
distance interactions
over read depth range
Filtered RINGs
All detectable
3D distance (Å)

Figure 5

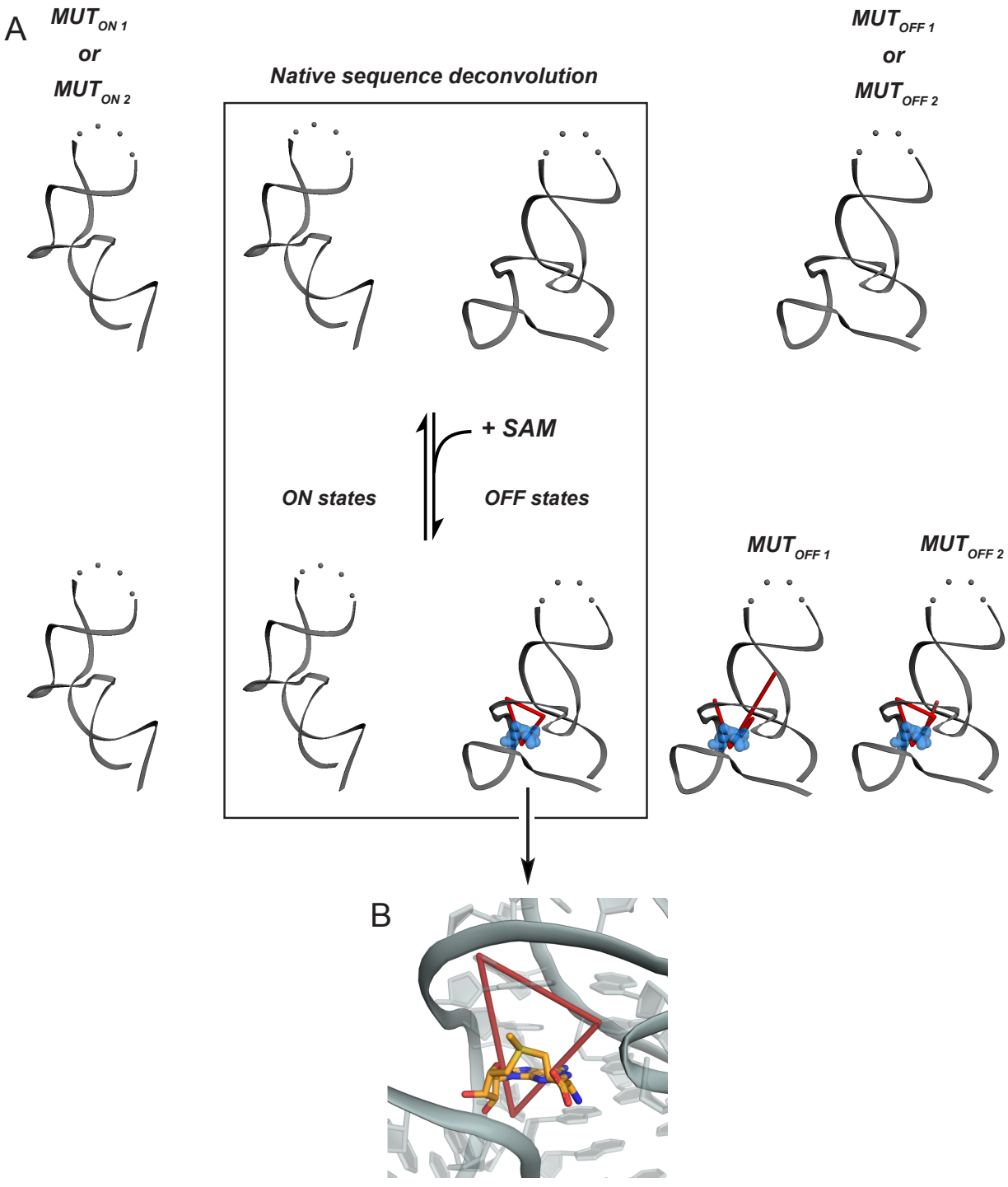
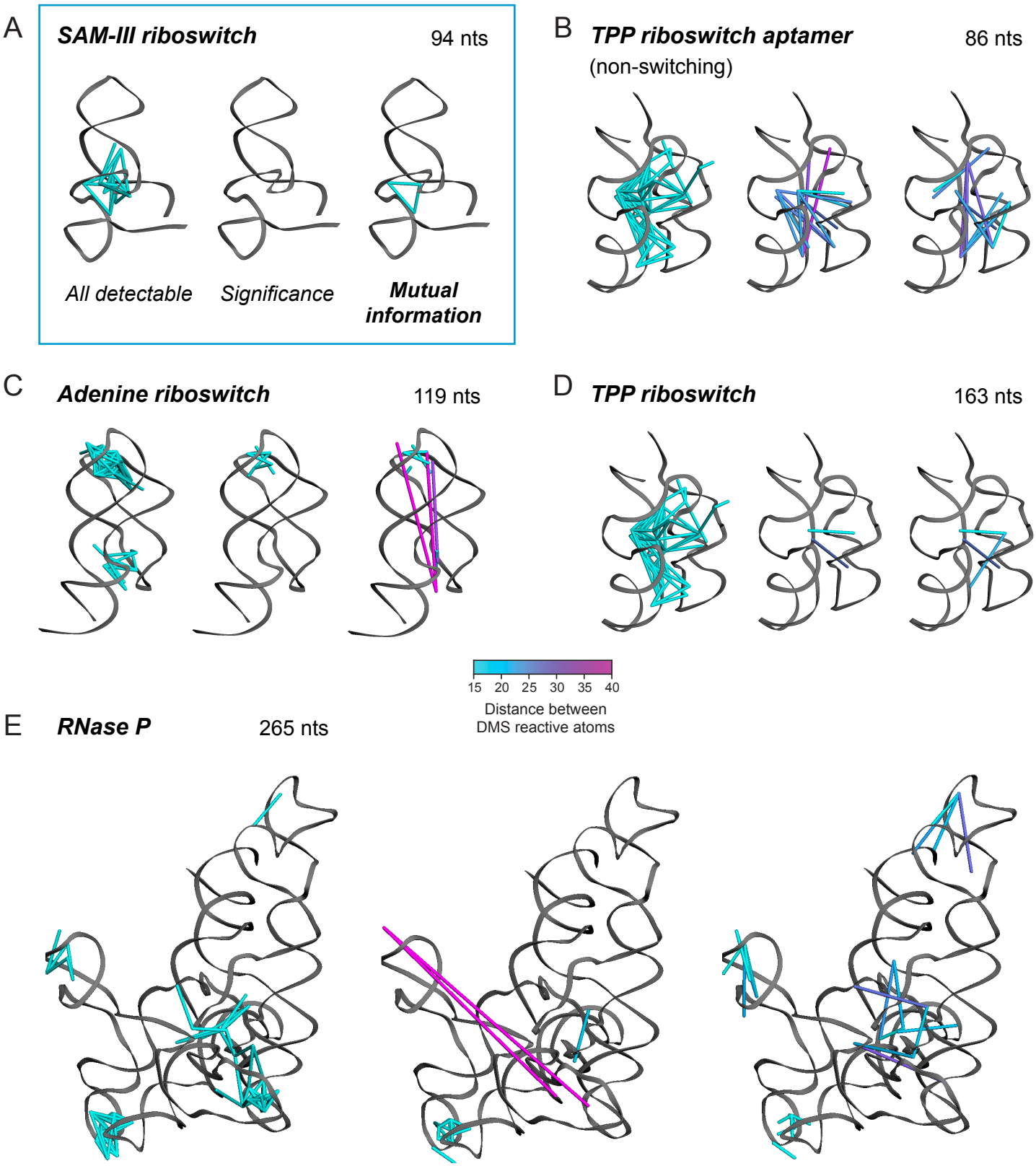
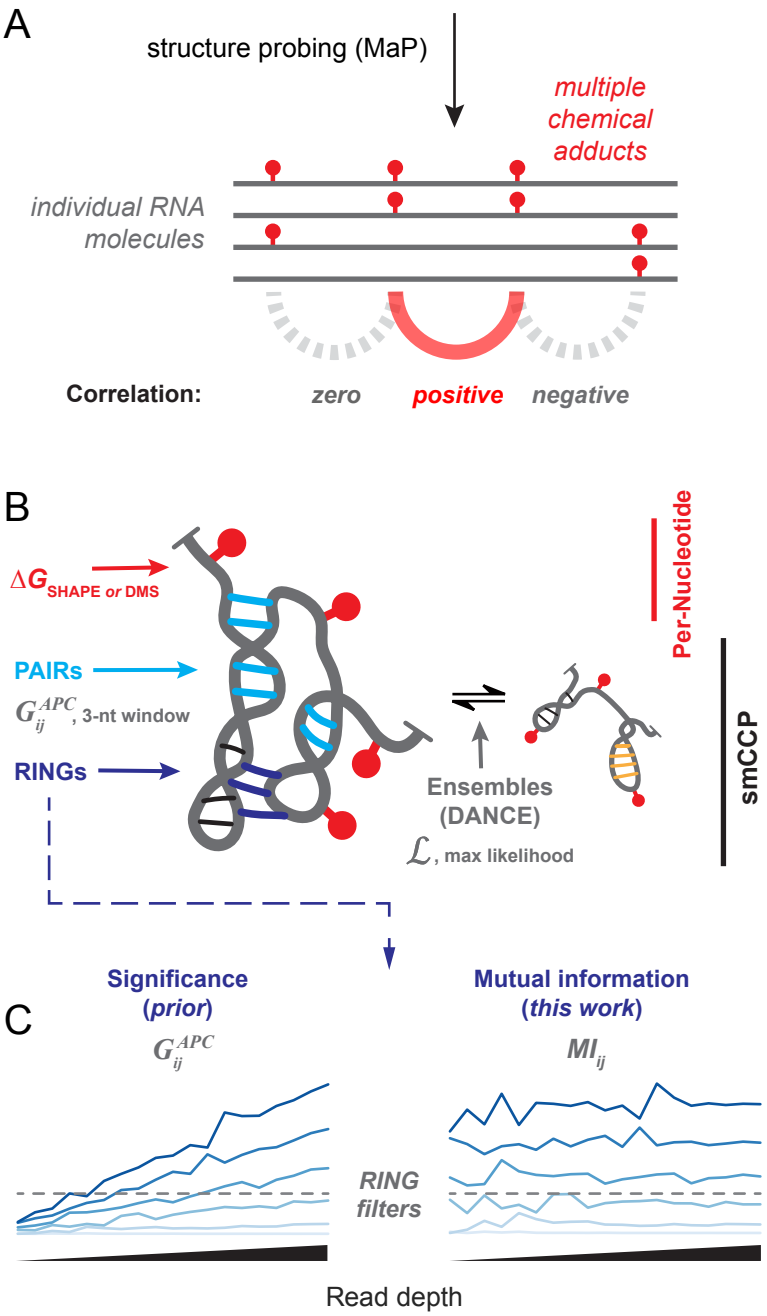

A
MUT ON 1
or
MUT ON 2
Native sequence deconvolution
MUT OFF 1
or
MUT OFF 2
+ SAM
ON states
OFF states
MUT OFF 1
MUT OFF 2
B

Figure 6


A
SAM-III riboswitch
94 nts
All detectable
Significance
Mutual information
B
TPP riboswitch aptamer
(non-switching)
86 nts
C
Adenine riboswitch
119 nts
D
TPP riboswitch
163 nts
15 20 25 30 35 40
Distance between DMS reactive atoms
E
RNase P
265 nts

Scheme 1



(print at single column width)