RNAvigate: Efficient exploration of RNA chemical probing datasets

Patrick S. Irving and Kevin M. Weeks*
Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290
* correspondence, weeks@unc.edu

Abstract

Chemical probing technologies enable high-throughput examination of diverse structural features of RNA including local nucleotide flexibility, RNA secondary structure, protein- and ligand-binding, through-space interaction networks, and multi-state structural ensembles. These layers of RNA structural information are often most incisive for understanding RNA structure-function relationships when combined with each other and when evaluated under structure- and function-altering conditions. Analysis of these complex data has required, often time consuming and awkward, juggling of multiple intermediate data files and software to create visualizations that support RNA-centered hypotheses. Here, we present the RNA visualization and graphical analysis toolset RNAvigate, developed as an easy-to-use and flexible Python library. RNAvigate currently integrates seven chemical probing data formats, nine secondary and tertiary structure formats, and eleven plot types. These features enable efficient exploration of nuanced relationships between chemical probing data, RNA structures, and motif annotations across multiple experimental samples. Modularity supports integration of new data types and plotting features. Compatibility with Jupyter Notebooks facilitates reproducibility and organization of multistep analyses and makes RNAvigate an ideal, time-effective, and non-burdensome platform for sharing full analysis pipelines. RNAvigate streamlines implementation of chemical probing strategies and accelerates discovery and characterization of diverse RNA-centric functions in biology.

Introduction

Primary, secondary, tertiary, and quaternary structural features all effect RNA function. Primary structure, the nucleotide sequence of an RNA transcript, determines transcription start sites, splice sites, and transcription termination and polyadenylation sites. RNAs can contain post-transcriptional modifications including pseudouridine, 6-methyl adenosine, and 2'-O-methyl

nucleotides, which increase the primary sequence alphabet. Secondary structure is the pattern of both canonical and non-canonical base pairing and can involve long, stable helical regions, pseudoknots, and conformationally flexible regions. Tertiary structure, the three-dimensional arrangement of atoms in an RNA transcript involves through-space interactions between nucleotide groups and can be stabilized by binding of ions, small molecules, or macromolecules. Quaternary structure, the multi-molecular structure of an RNA, includes its protein and RNA partner interactions.

These diverse levels of RNA structure can be interrogated with a large and expanding high-throughput toolbox. Primary sequence features can be determined by sequencing-based methods that measure transcript complexity (1) and numerous post-transcriptional modifications (2). Chemical probing methods make use of structure-specific chemical probes that form covalent adducts with RNA and are widely used to characterize RNA secondary structure. Chemical adducts can be detected and quantified at a per-nucleotide level using high-throughput sequencing methods including mutational profiling (MaP) or RT-STOP methods (3, 4). Single-molecule correlated chemical probing (smCCP) methods take advantage of the ability to detect multiple (potentially correlated) chemical events per molecule and provide deep insights into RNA secondary and tertiary structure, coordinated networks of protein-binding, and conformational ensembles (5). Proximity crosslinking methods use photo or chemical crosslinking to characterize RNA inter- and intramolecular interactions and RNA-protein binding (6–8). Both experimental and computational methods are becoming powerful tools for modeling higher-order RNA structure and macromolecular complexes (9, 10).

These strategies produce rich and complex datasets that report on interdependent structural elements in RNA. These datasets also pose two broad challenges for data exploration and hypothesis generation. First, it is typically important to explore data filtering and preprocessing steps, such as background-correction and normalization, to make decisions on how best to process and interpret raw data. Second, many studies require that multiple layers of data be visualized, in combination and as a function of experimental conditions. Solutions exist for examining individual aspects of RNA structure. Missing has been a flexible and easy-to-use data analysis pipeline that enables integration of multiple complex datasets to produce accessible visualizations in the multiple formats commonly used by the RNA community.

Jupyter notebooks are open-source interactive computing environments that work in any modern web browser and in many integrated development environments (11). These notebooks enable melding of analyses, visualizations, and observational notes. A single document thus creates an ideal environment for data exploration and hypothesis generation, a documented report of analysis findings, and a shareable and reproducible analysis pipeline. The widespread use and popularity of Jupyter notebooks has resulted in a well-developed ecosystem of open-source tools supporting collaboration on and publication of data analyses. However, creation of the types of plots used by the RNA community requires high-level knowledge of one of the programming languages used by Jupyter Notebooks (Python, Julia, and R).

Here we share a Python data visualization module, RNAvigate, which automates visualization for diverse data input and display formats. Experimental and structural data with different underlying sequences or reference positions are automatically aligned and repositioned for visual comparison. RNAvigate, when used as a data exploration tool within Jupyter notebooks, rapidly compares, filters, analyzes, and displays RNA structural information obtained from diverse sources. These features significantly streamline the analysis of complex datasets and the generation of impactful hypotheses regarding RNA structure-function interrelationships.

Results and Discussion

RNAvigate workflow

The RNAvigate workflow is organized into three steps (Figure 1). First, Inputs are curated from experiments, computations, and database searches. Second, Data Classes are created by providing RNAvigate with the input file names. Finally, visualizations are created using plotting functions built into RNAvigate.

Inputs and file formats

RNAvigate natively supports inputs of RNA structural information in many standard file and data formats. The simplest inputs are annotations of the primary sequence of RNA. These annotations describe RNA features (transcription start site, exon-intron junctions), functional regions, modified nucleotides, or binding sites for proteins, ligands, and oligonucleotides. RNAvigate also takes, as input, file formats output by structure prediction software (including RNAStructure and Vienna RNA (12, 13)) and secondary structure drawing software (XRNA, VARNA, StructureEditor, and R2DT (12, 14, 15)). Supported tertiary structure formats include

CIF and PDB. Inputted experimental data can include chemical probe reactivities, such as those from MaP experiments, single-molecule events detected by smCCP, and sequencing enrichments such as those from CLIP experiments. Inputted computational data can include pairing probabilities, covariation, and sequence conservation. File formats output from ShapeMapper2, PairMapper, RingMapper, DanceMapper, and RNPMapper (7, 16–18) are all read by RNAvigate. RNAvigate allows users to create new file formats and is extensible to not-yet-standardized or -envisioned data types.

Data classes

Data classes in RNAvigate are structured representations of input data in the form of Python objects, which create functionality for the data to be analyzed, filtered, and visualized. Input data files are converted to a data class when its RNAvigate keyword is assigned. This step also organizes data in a standardized way across experimental samples.

Inputs are converted into one of five data classes: annotations, secondary structures, tertiary structures, profiles, or interactions (Figure 1). Annotations contain regions of interest, single and grouped nucleotides of interest, and primer-binding sites. Secondary structures contain base-pairing information and optional drawing coordinates for secondary structure models. Tertiary structures contain atomic coordinates for tertiary structure models. Profiles contain any type of experimental or computational measurement made on a *per-nucleotide* basis. Examples of per-nucleotide data include chemical probing reactivities, structure-based entropies, and sequence conservation. Interactions contain experimental or computational measurements made *between* two nucleotides. These measurements include sequence covariation, base-pairing probabilities, single-molecule correlations, and crosslinking information. These five data classes are flexible and extensible abstractions and can represent most types of RNA structural data.

Visualization options

RNAvigate automates many standard analyses and visualizations with easy-to-use, flexible functions. Two standard visualizations commonly used by the RNA community are per-nucleotide graphs and connectivity diagrams. Per-nucleotide graphs display nucleotide position along the x-axis and a measured value on the y-axis. RNAvigate can output two types of per-nucleotide graphs: bar graphs for displaying a single dataset and stepped-line graphs (skyline plots) for comparing multiple datasets (Figure 1).

RNAvigate can create four types of connectivity diagrams: arc plots, circle plots, secondary structure diagrams, and interactive 3D molecular renderings. Connectivity diagrams display through-space intra-nucleotide interactions such as base-pairs, crosslinks, or correlated single-molecule events. Arc plots arrange nucleotides along the x-axis and show interactions as semi-circles. Circle plots arrange nucleotides around a circle with interactions drawn as parabolas. Secondary structure diagrams have custom arrangements of nucleotides and display interactions as lines. Finally, interactive 3-D molecule renderings arrange nucleotides according to atomic coordinates and display interactions as thin cylinders (Figure 1).

Other visualizations automated by RNAvigate include 3D-distance histograms, heatmaps, and contour plots (8, 16). Automated analyses include identification of regions with low SHAPE reactivity and low Shannon entropy (lowSS regions) (17), linear regressions, deltaSHAPE (18), and windowed area under receiver-operator characteristic curves (19).

Example workflows

The following sections provide examples for using RNAvigate to analyze RNA structure and to generate hypotheses regarding underlying RNA-mediated function. Figures 3-5 replicate published figures and illustrate the ability of RNAvigate to create high-content illustrations efficiently. The raw data, full code, and Jupyter notebooks used to generate these analyses and figures are fully and freely available via Github and Binder, which allow access to and interaction with these notebooks with no required software download or installation (see Methods and Data statement).

smCCP analysis

Single molecule correlated chemical probing (smCCP) strategies represent a revolution in RNA structure probing and allow multiple classes of through-space interactions to be measured in a simple chemical probing experiment (5). Whereas smCCP experiments are generally straightforward to perform, their analysis and interpretation can be complex. RNAvigate dramatically streamlines the required analysis. We will illustrate one such analysis in a detailed way in this first example, showing all RNAvigate code explicitly. Similar use principles then apply to all examples shared here.

Bacterial transfer-messenger RNA (tmRNA) is responsible for rescuing stalled ribosomes. This RNA has a complex secondary structure with an overall large open ring structure that includes

four pseudoknots. Secondary structures for tmRNAs are poorly predicted by computational methods alone (20). We used RNAvigate to facilitate analysis of a publicly available data set of in-cell and cell-free dimethyl sulfate (DMS) treated *E. coli* RNA (20).

We collected raw sequencing reads from a sequence read archive, the literature-accepted secondary structure in the form of a .ct file, and the cryo-EM structure (21) from the RCSB. Raw sequencing reads were then analyzed with ShapeMapper2, PairMapper and RingMapper. These data informed minimum free energy-based secondary structure modeling using RNAStructure.

The RNAvigate Python module was imported into a Jupyter Notebook and given the alias **rnav** (Figure 2A). Input file names were provided for each sample using **rnav.Sample()**. This important step accomplishes two things: the data contained in these files are converted into data classes, and these data classes are given a short keyword for easy access, which is associated with each experimental sample. Data for cell-free and in-cell experimental samples were assigned to variable names **cf** and **ic**, respectively (Figure 2A). We illustrate here three classes of smCCP data analysis for the tmRNA; each plot is created with a single command, optionally with modifying arguments (Figure 2B-D).

We first visualized overall reactivity profiles under in-cell and cell-free conditions, compared using a skyline plot (Figure 2B). Per-nucleotide reactivity rates were highly similar overall, and slightly higher under cell-free conditions, reflective of higher structure for this non-coding RNA in cells, a common observation. Next, an arc plot was used to compare modeled and accepted structures and to visualize both DMS-MaP per-nucleotide reactivities and PAIR-MaP through-space correlations (PAIRs) that informed the structure model. This visualization revealed that in-cell data produced a more accurate model of the secondary structure than did data obtained under cell-free conditions. The RNAvigate analysis showed that increased accuracy is driven by more abundant and accurate PAIRs for the in-cell dataset, as the per-nucleotide reactivity profiles are very similar (Figure 2C). Finally, RING-MaP through-space correlations (RINGs) were visualized in the context of the cryo-EM structure. These RINGs reveal extensive structural communication between the adjacent pseudoknots along the outer ring of the structure and an absence of structural communication across the circle-like opening (Figure 2D).

Analysis of regions of low SHAPE reactivity and low Shannon entropy

One especially informative analysis of long RNAs is to identify regions that have both low SHAPE reactivities, indicative of extensive base-pairing, and also low Shannon entropy, which indicates a well-defined structure. These low SHAPE-low Shannon entropy (lowSS) regions are correlated with functional regions in long RNAs (17, 22, 23). RNAvigate makes identification and visualization of lowSS regions straightforward. Here, we use RNAvigate to reproduce a useful analysis based on a publicly available dataset obtained from SHAPE probing of Dengue virus serotype 2 (DENV2) genomic RNA (17).

Dengue virus has a positive-strand single-stranded RNA genome of about 11 kilobases. Highly conserved structures in the 5' and 3' untranslated regions regulate viral processes and are well characterized (24). In contrast, identification of functional structures in less conserved regions has presented challenges. Analysis of lowSS regions made it possible to define novel regions of functional importance and to identify regions likely to have complex higher-order (tertiary) structures (17, 25).

SHAPE reactivity data were used to inform a minimum free energy structure model and base-pairing probabilities using the fold and partition programs within RNAstructure (12), respectively. These results and RNAvigate were imported into a Jupyter Notebook. RNAvigate was used to calculate 51-nucleotide windowed median SHAPE reactivities and median Shannon entropies from the provided pairing probabilities (Figure 3, *top* and *middle*). The program then identified lowSS regions where the windowed median SHAPE reactivity was below 0.4 and the median Shannon entropy was below 0.15.

RNAvigate enables a useful visualization of this analysis as stacked linear profiles of SHAPE reactivity and Shannon entropy, and of predicted minimum free energy structures and pairing probabilities displayed as arc plots. lowSS regions are highlighted and numbered (Figure 3); plots share the same x-axis (nucleotide position) for easy comparison. Visualization of the first 2000 nucleotides of the DENV2 genome revealed eight lowSS regions. This straightforward application of RNAvigate recapitulates a published analysis (17) that required significantly greater effort to create.

Analysis of time-resolved probing

Many dynamic RNA processes are best analyzed in a time-resolved way, which creates complex datasets across multiple experimental samples. In one recent example, we developed

a fast-reacting alkylating reagent, trimethyloxonium (TMO), and used it to follow the time-resolved folding of the catalytical core of a ribonuclease P (RNase P) enzyme (26). The folding pathway was examined in smCCP experiments and involved creating multiple structure-based maps and heatmaps of through-space RINGs. This complex analysis is now rendered straightforward with RNAvigate.

B. stearothermophilus RNase P is a long non-coding RNA that catalyzes cleavage of 5' leader sequences of precursor tRNAs, an essential step for tRNA maturation. TMO reacts with RNA via a chemical mechanism similar to that of DMS, but reacts 90 times faster. TMO was used to perform time-resolved probing of RNase P, and smCCP analysis revealed the through-space interaction networks that form and change during the folding of this large RNA (26). The in vitro-transcribed RNase P RNA was treated with TMO, initially in the absence of Mg^{2+}, and then at multiple time points after addition of Mg^{2+}. Raw sequencing reads corresponding to MaP analyses at each time point and a custom-drawn secondary structure were obtained from the original publication. The raw data were analyzed using ShapeMapper2 and RingMapper. A three-dimensional model of the RNA (based on PDB 3DHS (27)) was obtained from the authors. RNAvigate was used to import the resulting files into a Jupyter notebook.

For the RNAvigate analysis, a detailed secondary structure was generated to orient the reader to the functional domains, important interactions, and helices present in RNase P, particularly the catalytic core, the P2-P5 pseudoknot, and the L5.1-L15.1 loop-loop interaction (Figure 4A). Next, the strongest RINGs with positive correlations were displayed on the secondary structure drawing. This superposition revealed large changes in the structural communication network as folding progresses. The L5.1-L15.1 loop-loop interaction forms very quickly and is present in an early partially folded state. In the fully folded state, interactions within the catalytic core are more pronounced, revealing this structure forms more slowly (Figure 4B).

The spatial density of positively correlated RINGs, plotted as heatmaps, across the three states revealed additional interactions where through-space RING correlations change dramatically during the folding process (Figure 4B). This visualization, like the RINGs displayed on the secondary structure above, identified rapid formation of the L5.1-L15.1 loop-loop interaction upon addition of Mg^{2+}. These heatmaps also reveal that interactions in the P4-P7 region disappear and interactions in P2-P5,P7 region form quickly after addition of Mg^{2+}, indicating interdependency between the pseudoknot helices. RING densities also reveal a slower forming

interaction between P4 and P19. Finally, RINGs from the fully folded state were filtered to isolate likely tertiary interactions and plotted on the 3D structure. These RINGs illustrate the density of interactions in the catalytic core and suggest this fully folded state closely matches the known structure (Figure 4C). Complex, time-resolved data are now analyzed efficiently using RNAvigate.

Deconvolution of a conformational ensemble

smCCP strategies can now resolve complex conformational ensembles of RNA, including in the cellular environment. DANCE-MaP analysis uses machine learning to deconvolute smCCP data from a DMS-MaP experiment to define the dominant states in a conformational ensemble and can then assign through-space PAIRs and RINGs to each state. This single experiment and subsequent analyses produce multiple layers of data for each state including population percentages, reactivity profiles, PAIRs, RINGs, predicted secondary structures, and pairing probabilities (28). RNAvigate makes analysis of this layered information broadly accessible.

The adenine riboswitch region of the *V. vulnificus add* mRNA populates two dominant conformations. Adenine binds to the ON state and shifts the equilibrium toward this state. Formation of the ON state results in production of adenine deaminase because, in this conformation, the Shine-Dalgarno (SD) sequence of the mRNA is single-stranded and accessible to the ribosome. In the OFF state, the Shine-Dalgarno sequence is sequestered in a duplex. Conventional per-nucleotide structure probing methods measure population averages and are misleading for this riboswitch, because the RNA samples multiple well-populated conformations (28).

RNAvigate enables efficient visualization of the multiple dominant structures of an RNA structural ensemble. Data files from the DANCE-MaP analysis of adenine riboswitch in the absence of adenine (28) were obtained from the GEO (GSE182552). These single-experiment data files contain reactivity profiles, PAIRs and RINGs for each of the dominant conformations. These data were used with foldClusters, a module of DanceMapper, to inform minimum free energy structure models for each conformation. These models were drawn with custom layouts using StructureEditor (12) to match the original publication. A three-dimensional structure for the translation ON state was obtained from the PDB (4TZX) (29). File names were then provided to RNAvigate within a Jupyter notebook and used to create the visualizations in this example.

As visualized using arc plots and consistent with the original publication (28), PAIRs accurately and sensitively detected base-paired nucleotides, recapitulating the accepted secondary structures of, and the differences between, these two states (Figure 5A). In the translation ON state, RINGs captured a known loop-loop tertiary interaction (Figure 5B,C, in *red*). In the context of the crystal structure, through-space RING interactions link two interacting loops that are close in space (Figure 5C). RNAvigate enabled efficient analysis of an information-rich DANCE-MaP experiment and enabled straightforward visualization of the multiple dominant states present in an RNA structural ensemble.

Perspective

RNA structure-function interrelationships are complex. Powerful and orthogonal computational and experimental strategies have been developed to identify the hierarchical and interdependent elements of RNA structure. One challenge is to understand, integrate, and interpret these rich layers of information. The visualization and analysis tools of RNAvigate facilitate analysis of diverse structural features of RNA. RNAvigate in conjunction with Jupyter Notebooks enables rapid analysis of and thorough documentation of data quality control, data exploration, and hypothesis generation. RNAvigate then directly generates publication-quality figures.

Data sharing and reproducibility are critically important aspects of any data analysis and have become major foci of both funders and researchers. RNAvigate, when used within a Jupyter notebook, provides a convenient, cost-effective, non-burdensome platform for documentation and for transparent and reproducible sharing of analyses of RNA experimental and computational data. Full pipelines for analysis and figure generation can then be recreated using open-source web-based tools, without the requirement to download any data or install any software, as demonstrated by the Jupyter notebooks accompanying this manuscript.

RNAvigate is modular and readily extensible, and thus provides a framework to support analysis and visualization of diverse experimental and computational data types. We anticipate that RNAvigate will enhance development, implementation and sharing of wide-ranging RNA structure, interaction, and annotation strategies to discover and characterize diverse RNA-centric functions in biology.

Methods and Data

Data files and Jupyter notebooks containing all data and methods described here are included in the supplemental data and are available on Github (https://github.com/Weeks-UNC/RNAvigate_figures) and Binder (https://mybinder.org/v2/gh/Weeks-UNC/RNAvigate_figures/HEAD). Notebooks on Github have been made interactive using Binder (part of the Jupyter Project). Binder is a free web service that launches and hosts a Jupyter environment containing a notebook and its related data files, facilitating the shareability of Jupyter notebooks. Users can reproduce these analyses and explore the data using RNAvigate in a web browser, with no required downloads or software installations.

Code availability

RNAvigate python module is available on Github (https://github.com/Weeks-UNC/RNAvigate). RNAvigate documentation is available on Read the Docs (https://rnavigate.readthedocs.io).

Acknoledgements

This work was supported by the NIH (R35 GM122532 and R01 HL111527) to K.M.W.

References

1. Hrdlickova,R., Toloue,M. and Tian,B. (2017) RNA-Seq methods for transcriptome analysis. *WIREs RNA*, **8**, e1364.

2. Wiener,D. and Schwartz,S. (2021) The epitranscriptome beyond m6A. *Nat. Rev. Genet.*, **22**, 119–131.

3. Strobel,E.J., Yu,A.M. and Lucks,J.B. (2018) High-throughput determination of RNA structures. *Nat. Rev. Genet.*, **19**, 615–634.

4. Weeks,K.M. (2021) SHAPE Directed Discovery of New Functions in Large RNAs. *Acc. Chem. Res.*, **54**, 2502–2517.

5. 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.

6. Lin,C. and Miles,W.O. (2019) Beyond CLIP: advances and opportunities to measure RBP–RNA and RNA–RNA interactions. *Nucleic Acids Res.*, **47**, 5490–5501.

7. Weidmann,C.A., Mustoe,A.M., Jariwala,P.B., Calabrese,J.M. and Weeks,K.M. (2021) Analysis of RNA–protein networks with RNP-MaP defines functional hubs on RNA. *Nat. Biotechnol.*, **39**, 347–356.

8. Christy,T.W., Giannetti,C.A., Houlihan,G., Smola,M.J., Rice,G.M., Wang,J., Dokholyan,N.V., Laederach,A., Holliger,P. and Weeks,K.M. (2021) Direct Mapping of Higher-Order RNA Interactions by SHAPE-JuMP. *Biochemistry*, **60**, 1971–1982.

9. Zhang,J., Fei,Y., Sun,L. and Zhang,Q.C. (2022) Advances and opportunities in RNA structure experimental determination and computational modeling. *Nat. Methods*, **19**, 1193–1207.

10. Li,B., Cao,Y., Westhof,E. and Miao,Z. (2020) Advances in RNA 3D Structure Modeling Using Experimental Data. *Front. Genet.*, **11**.

11. Beg,M., Taka,J., Kluyver,T., Konovalov,A., Ragan-Kelley,M., Thiéry,N.M. and Fangohr,H. (2021) Using Jupyter for Reproducible Scientific Workflows. *Comput. Sci. Eng.*, **23**, 36–46.

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

13. Lorenz,R., Bernhart,S.H., Höner zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.

14. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.

15. Sweeney,B.A., Hoksza,D., Nawrocki,E.P., Ribas,C.E., Madeira,F., Cannone,J.J., Gutell,R., Maddala,A., Meade,C.D., Williams,L.D., *et al.* (2021) R2DT is a framework for predicting and visualising RNA secondary structure using templates. *Nat. Commun.*, **12**, 3494.

16. Christy,T.W., Giannetti,C.A., Laederach,A. and Weeks,K.M. (2021) Identifying proximal RNA interactions from cDNA-encoded crosslinks with ShapeJumper. *PLOS Comput. Biol.*, **17**, e1009632.

17. 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., *et al.* (2018) Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl. Acad. Sci. U. S. A.*, **115**, 11513–11518.

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

19. Lan,T.C.T., Allan,M.F., Malsick,L.E., Woo,J.Z., Zhu,C., Zhang,F., Khandwala,S., Nyeo,S.S.Y., Sun,Y., Guo,J.U., *et al.* (2022) Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nat. Commun.*, **13**, 1128.

20. 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.

21. Rae,C.D., Gordiyenko,Y. and Ramakrishnan,V. (2019) How a circularized tmRNA moves through the ribosome. *Science*, **363**, 740–744.

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

23. Boerneke,M.A., Ehrhardt,J.E. and Weeks,K.M. (2019) Physical and Functional Analysis of Viral RNA Genomes by SHAPE. *Annu. Rev. Virol.*, **6**, 93–117.

24. Gebhard,L.G., Filomatori,C.V. and Gamarnik,A.V. (2011) Functional RNA Elements in the Dengue Virus Genome. *Viruses*, **3**, 1739–1756.

25. Boerneke,M.A., Gokhale,N.S., Horner,S.M. and Weeks,K.M. (2023) Structure-first identification of RNA elements that regulate dengue virus genome architecture and replication. *Proc. Natl. Acad. Sci.*, **120**, e2217053120.

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.

27. Kazantsev,A.V., Krivenko,A.A. and Pace,N.R. (2009) Mapping metal-binding sites in the catalytic domain of bacterial RNase P RNA. *RNA*, **15**, 266–276.

28. 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.

29. Zhang,J. and Ferré-D'Amaré,A.R. (2014) Dramatic Improvement of Crystals of Large RNAs by Cation Replacement and Dehydration. *Structure*, **22**, 1363–1371.

Figure Legends

Figure 1. Workflow for data exploration using RNAvigate.

Figure 2. A complete RNAvigate workflow: single molecule correlated chemical probing (smCCP) of tmRNA. In this example, full RNAvigate code is shown for each figure. (*box*) Essential coding steps in an RNAvigate analysis explained with code-like examples. Keywords and function names here differ from real code. (A) Importing RNAvigate module and loading in-cell and cell-free DMS probing data. These steps create a Python object that contains all data in a structured format. (B) Skyline plot comparing DMS-induced in-cell and cell-free mutation rates per nucleotide. (C) Arc plots illustrating multiple layers of experimental and structural data including cell-free and in-cell structures, data-informed structure predictions, and PAIRs. (D) Visualization of three-dimensional tmRNA structure overlayed with cell-free and in-cell RINGs. Data are from ref. (20).

Figure 3. Identification and visualization of low-SHAPE–low-Shannon entropy (lowSS) regions in the DENV2 genome using RNAvigate. Median (*top*) SHAPE reactivity and (*middle*) Shannon entropies over 51-nucleotide windows across the DENV2 genome. (*bottom*) Modeled minimum free energy structure (gray) and pairing probabilities (see scale) illustrated as arc plots. Regions with median SHAPE reactivity below 0.4 and median Shannon entropy below 0.15 are defined as lowSS regions and are shaded in gray and numbered. Full RNAvigate code used to generate this analysis and visualization shown at right; data are from ref. (17).

Figure 4. Time-resolved probing and folding of the RNase P RNA, visualized using RNAvigate. (A) Secondary structure of the catalytic core of RNase P RNA. Regions involved in major time-dependent changes are highlighted. (B) (*top*) Through-space RINGs as a function of folding time shown on the secondary structure. Newly formed RINGs at each timepoint are shown in green; RINGs retained from the prior timepoint are yellow. (*bottom*) Heatmaps of RING density at each time point. (C) Three-dimensional structure of the RNase P RNA with superimposed RINGs (in green). Data are from ref. (26).

Figure 5. DANCE-MaP analysis of the *add* riboswitch ensemble, visualized using RNAvigate. (A) PAIRs, RINGs (each shown as arc plots), and reactivity profiles for DANCE-identified ON and OFF states. (B) Secondary structure drawings (conventional style) for each component,

with PAIRs (blue) and RINGS (red) superimposed. Shine-Dalgarno sequence (SD) is highlighted. (C) Structure of *add* riboswitch in the ligand-bound state, overlayed with RINGs uniquely identified in this state. Note that these data were all obtained in the absence of adenine ligand but, nonetheless, clearly show that the ade riboswitch populates the ON state. Data are from ref. (20).

Figure 1

Inputs:

obtained from experiments, computations, and database searches

Annotations

mRNA features
functional regions
protein and ligand binding sites
modified nucleotides

Structural data

secondary structure models
secondary structure drawings
tertiary structure models

Experimental data

chemical probe reactivities
single-molecule correlations
sequencing enrichments

Computational data

base-pairing probabilities
sequence covariation
sequence conservation

Data classes:

obtained by providing RNAvigate with input file names

Annotations

regions, groupings, or single nucleotides

Secondary structures

base-pairing and drawing coordinates

Tertiary structures

3D structure coordinates

Profiles

per-nucleotide information

Interactions

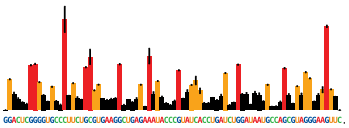
inter-nucleotide information

Visualizations:

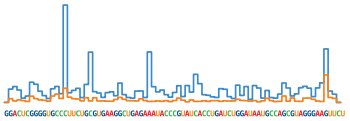
obtained from RNAvigate plotting functions

11+ plot types, including:

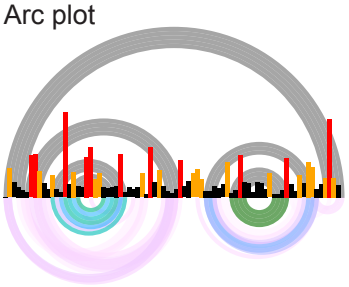
SHAPE profile



Skyline (step)

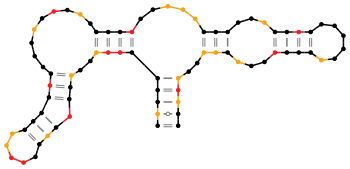


Arc plot



Circle plot

Structure drawing



3D structure

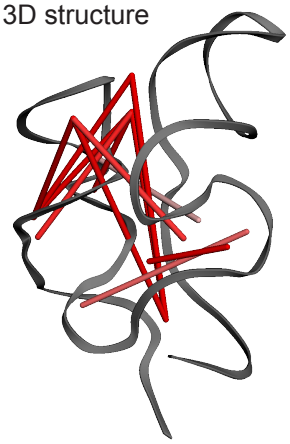


Figure 2

A Create samples and provide data inputs

```
import rnavigate as rnav

common_data = rnav.Sample(
    ct="tmRNA.ct",
    pdb={"filepath": "6q9A_tmRNA.pdb",
         "chain": "4",
         "fasta": "tmRNA.fa"})
cf = rnav.Sample(
    sample="Cell-free",
    inherit=common_data,
    dmsmap="cellfree_tmRNA_profile.txt",
    ringmap="cellfree_tmRNA_rings.txt",
    pairmap="cellfree_tmRNA-pairmap.txt",
    compct="cellfree_tmRNA.all.f.ct")
ic = rnav.Sample(
    sample="In cell",
    inherit=common_data,
    dmsmap="incell_tmRNA_profile.txt",
    ringmap="incell_tmRNA_rings.txt",
    pairmap="incell_tmRNA-pairmap.txt",
    compct="incell_tmRNA.all.f.ct")
```

Code structure

Import the *RNAvigate python module*

```
import rnavigate as rnav
```

Create *RNAvigate samples* by providing *data file inputs* with *data class keywords*

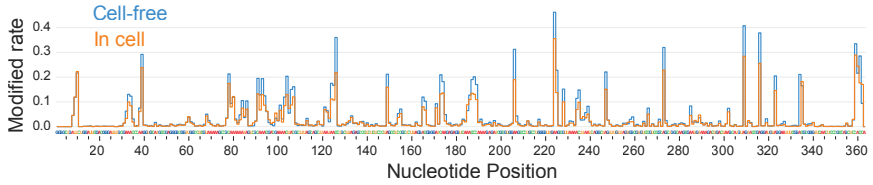
```
sample_name = rnav.Sample(
    data_class_keyword="input_file.txt")
```

Create visualizations by providing *plotting functions* with *samples* and *data class keywords*

```
rnav.plotting_function(
    samples=[sample_name],
    plotting_keyword="data_class_keyword")
```

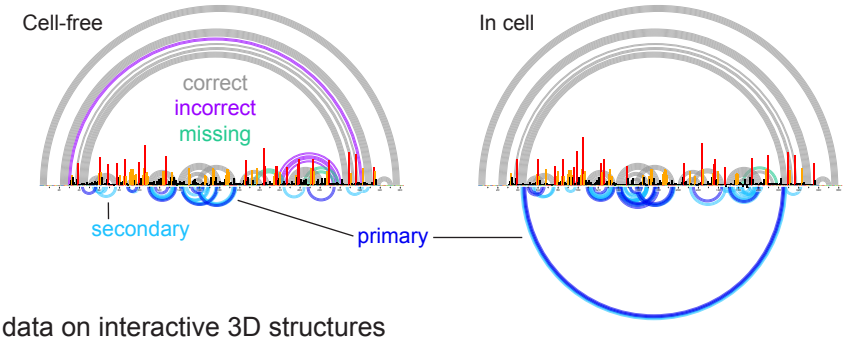
B Compare per-nucleotide data between samples using skyline plots

```
rnav.plot_skyline(
    samples=[cf, ic],
    profile="dmsmap",
    columns="Modified_rate")
```



C Show multiple layers of experimental and structural data on arc plots

```
rnav.plot_arcs(
    samples=[cf, ic],
    ct="ct",
    comp="compct",
    profile="dmsmap",
    interactions="pairmap")
```



D Plot per-nucleotide and internucleotide data on interactive 3D structures

```
rnav.plot_mol(
    samples=[cf, ic],
    structure="pdb",
    interactions="ringmap",
    interactions_filter={
        "MI_ge": 0.000025,
        "ss_only": True,
        "positive_only": True,
        "Statistic_ge": 23,
        "min_cd": 6,
        "Zij_ge": 2}
    nt_color="grey",
    hide_cylinders=True)
```

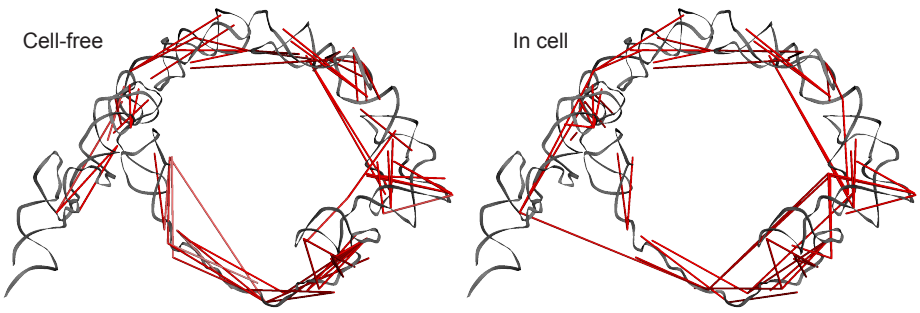
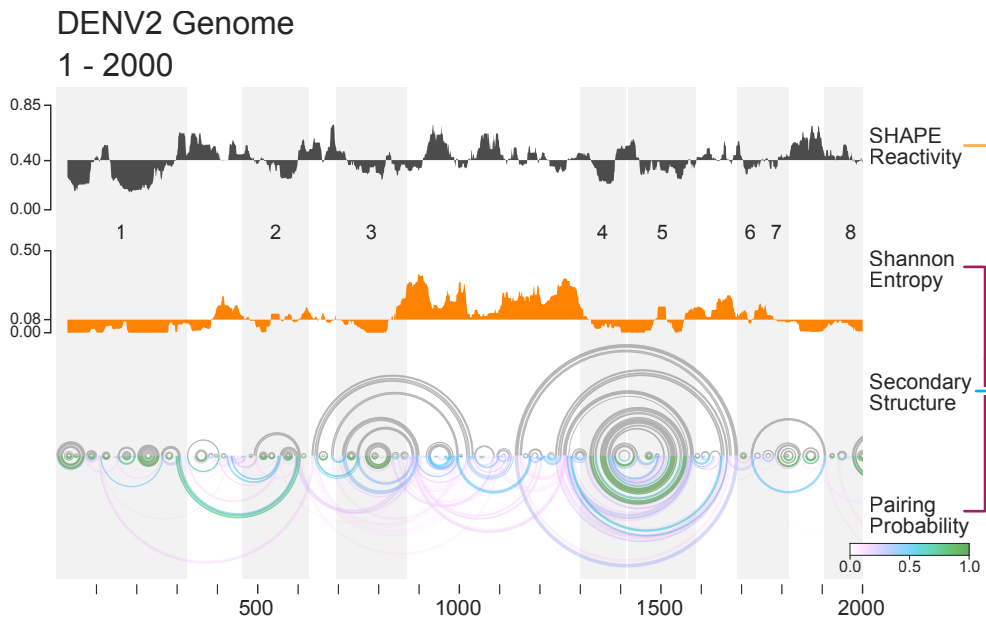


Figure 3



Code

```
import rnavigate as rnav
denv = rnav.Sample(
    sample="DENV2 Genome",
    shapemap="DENV2.map",
    pairprob="DENV2.dp",
    ct="DENV2_MFE.ct")
analysis = rnav.LowSS(
    sample=denv,
    region=[1, 2000])
```

Figure 4

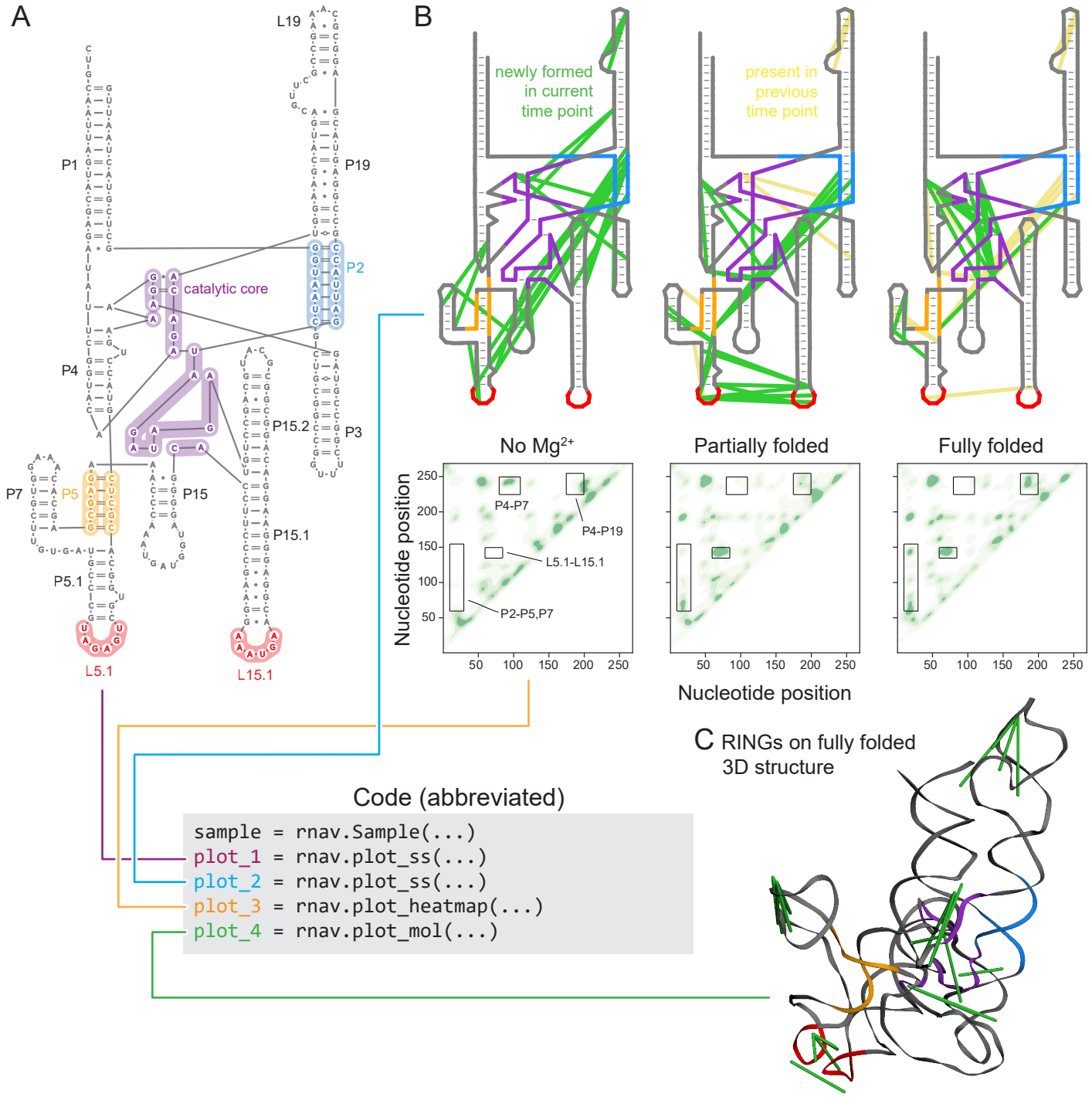
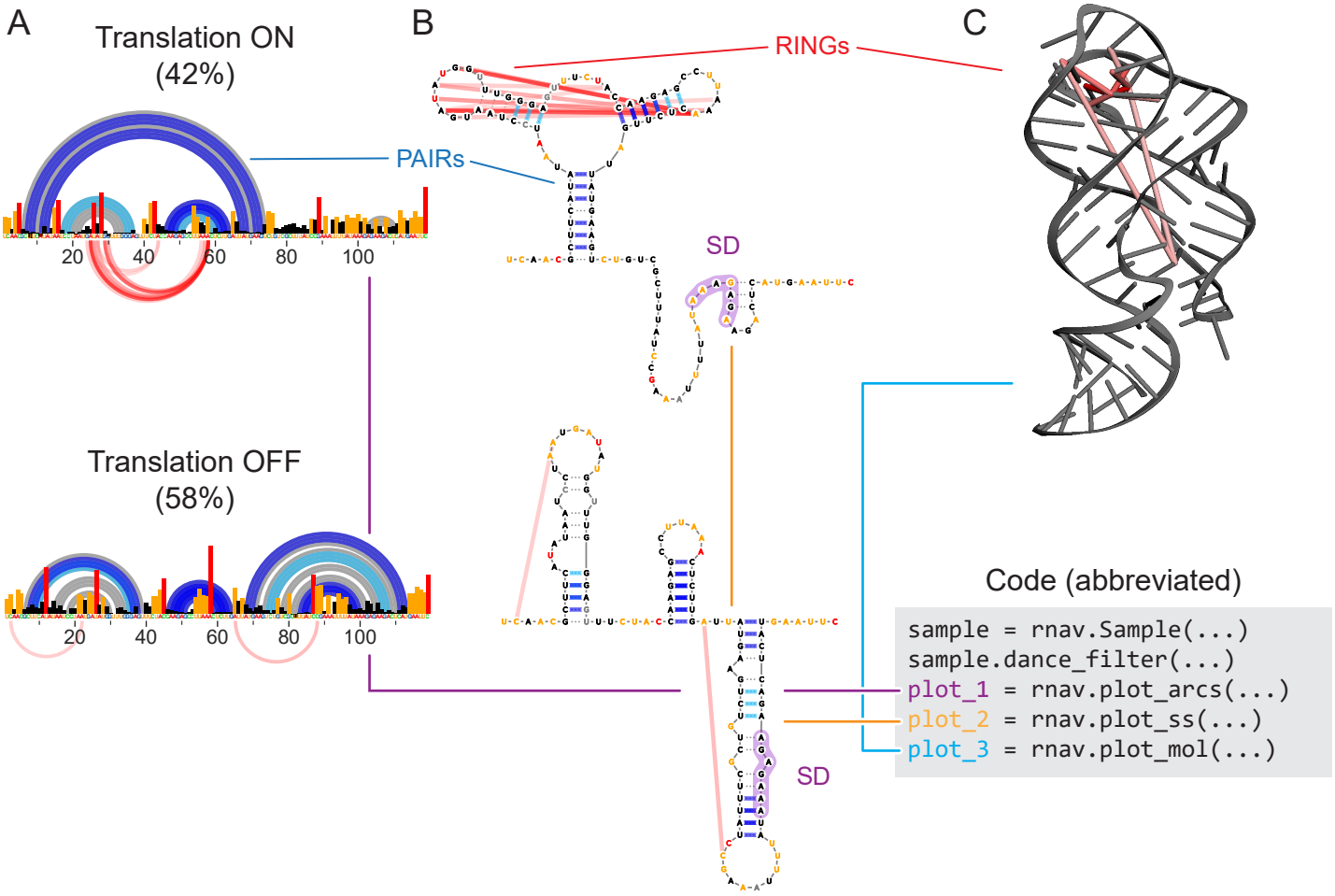

A
L19
P1
P19
P2
catalytic core
P4
P15.2
P3
P7
P5
P15
P15.1
P5.1
L5.1
L15.1
B
newly formed in current time point
present in previous time point
No Mg2+
Partially folded
Fully folded
Nucleotide position
P4-P7
P4-P19
L5.1-L15.1
P2-P5,P7
Nucleotide position
C RINGs on fully folded 3D structure
Code (abbreviated)
sample = rnav.Sample(...)
plot_1 = rnav.plot_ss(...)
plot_2 = rnav.plot_ss(...)
plot_3 = rnav.plot_heatmap(...)
plot_4 = rnav.plot_mol(...)

Figure 5


A
Translation ON
(42%)
20
40
60
80
100
Translation OFF
(58%)
20
40
60
80
100
B
RINGs
PAIRs
SD
SD
C
Code (abbreviated)
sample = rnav.Sample(...)
sample.dance_filter(...)
plot_1 = rnav.plot_arcs(...)
plot_2 = rnav.plot_ss(...)
plot_3 = rnav.plot_mol(...)