

# IPANEMAP Suite: a pipeline for probing-informed RNA structure modeling

Pierre Hardouin<sup>1,†</sup>, Nan Pan<sup>2,†</sup>, Francois-Xavier Lyonnet du Moutier<sup>1,†</sup>, Nathalie Chamond<sup>1</sup>, Yann Ponty<sup>2</sup>, Sebastian Will<sup>2,\*</sup>, Bruno Sargueil<sup>1,\*</sup>

<sup>1</sup>CNRS UMR 8038, CiTCoM Cibles Thérapeutiques et Conception de Médicaments, Université Paris Cité, 4 avenue de l'Observatoire, 75270 Paris, France

<sup>2</sup>CNRS UMR 7161, LIX, Ecole Polytechnique, 1 rue Estienne d'Orves, 91120 Palaiseau, France

\*To whom correspondence should be addressed. Email: [bruno.sargueil@parisdescartes.fr](mailto:bruno.sargueil@parisdescartes.fr)

Correspondence may also be addressed to Sebastian Will. Email: [sebastian.will@polytechnique.edu](mailto:sebastian.will@polytechnique.edu)

<sup>†</sup>The first three authors should be regarded as Joint First Authors.

## Abstract

In addition to their sequence, multiple functions of RNAs are encoded within their structure, which is often difficult to solve using physico-chemical methods. Incorporating low-resolution experimental data such as chemical probing into computational prediction significantly enhances RNA structure modeling accuracy. While medium- and high-throughput RNA structure probing techniques are widely accessible, the subsequent analysis process can be cumbersome, involving multiple software and manual data manipulation. In addition, the relevant interpretation of the data requires proper parameterization of the software and a strict consistency in the analysis pipeline. To streamline such workflows, we introduce IPANEMAP Suite, a comprehensive platform that guides users from chemically probing raw data to visually informative secondary structure models. IPANEMAP Suite seamlessly integrates various experimental datasets and facilitates comparative analysis of RNA structures under different conditions (footprinting), aiding in the study of protein or small molecule interactions with RNA. Here, we show that the unique ability of IPANEMAP Suite to perform integrative modeling using several chemical probing datasets with phylogenetic data can be instrumental in obtaining accurate secondary structure models. The platform's project-based approach ensures full traceability and generates publication-quality outputs, simplifying the entire RNA structure analysis process. IPANEMAP Suite is freely available at <https://github.com/Sargueil-CiTCoM/ipasuite> under a GPL-3.0 license.

## Introduction

The integration of chemical structure probing experiments significantly improves the accuracy of computations for predicting RNA secondary structure. The information from such experiments, by and large, position-wise pairing propensities, is integrated into folding algorithms as soft constraints [1–3]. Typically, these propensities are determined as reactivities with specific chemical probes. For example, Deigan *et al.* [1] transform the reactivities into pseudo-energies, which specifically penalize or reward the pairing of the individual nucleotides in predicted structures. Different types of probing experiments were proposed, varying in their chemistry and readout of the structure-informing reaction rates of these probes with RNA nucleotides. Concerning the former, one distinguishes two main types of chemical probes: first, DMS (dimethyl sulfate) or CMCT [cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate], which probe the availability of H-bond donor or acceptor sites located on the Watson–Crick face of the base [4]; second, SHAPE probes that report on the flexibility of the ribose, reacting with its 2' hydroxyl group [5–8].

Eventually, both types of probes induce covalent modifications of the nucleoside with structure-dependent frequencies. These modifications can be read out and converted to reactivity scores. The modifications are mapped using reverse tran-

scriptase (RT), where they lead to the premature termination of the reverse transcription.

Such experiments are analyzed on capillary electrophoresis (CE) systems.

Several pipelines have been developed to analyze the electrophoresis results yielding a reactivity profile for the RNA of interest [9–13]. Reactivity profiles can then be fed to RNA secondary structure prediction software, such as RNAfold [14, 15], RNAstructure [16, 17], or Shapeknot [18], to guide the prediction.

However, the existing pipelines were developed to utilize the information from single probing experiments (including replicates), while recent results promise strong benefits from combining results of multiple probing experiments in the modeling of one RNA [19–23]. For example, experiments can vary in the use of chemical probes and specific experimental conditions (salts, temperature, etc). Since they can provide complementary information, integrating multiple reactivity profiles at the same time has the potential to substantially improve the modeling accuracy.

For instance, our recently developed bioinformatics tool IPANEMAP [23] was shown to match the performance and even outcompete available structure modeling tools due to clean integration of multiple probing data.

The development of such an innovative modeling workflow eased and enhanced the accuracy of experimentally driven

Received: October 23, 2024. Revised: February 7, 2025. Editorial Decision: February 24, 2025. Accepted: March 5, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of NAR Genomics and Bioinformatics.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

RNA secondary structure modeling. However, using them requires the local installation of several in-house developed software or scripts with different prerequisites and to run each script manually through the use of command lines. In addition, going from the raw CE files to the RNA secondary structure drawing colored with reactivities involves cumbersome file manipulations from one software to another, often requiring format adaptation and calculations. The whole process ends up being very time-consuming, complex, and error-prone, especially for inexperienced users. To prevent such problems, convenient web servers, such as RNAProbe [24] or RNATHOR [25], have been developed, but none support an initial processing of CE data, or offer the possibility to integrate datasets obtained with different probes.

We developed IPANEMAP Suite to fill this gap while ensuring a full traceability and reproducibility all along the process. It is designed as a fully integrated and highly configurable pipeline organizing the user journey from CE files to the RNA secondary structure model representation.

IPANEMAP Suite is grounded on a new version of IPANEMAP that offers the possibility to exploit the covariation information contained in a multiple sequence alignment (MSA) of a phylogeny of the RNA of interest, further improving its prediction performance. CE file analysis is provided by QuShape [12], but structure probing data from other CE data processing software [9–13], or from high-throughput “Map” or “Seq” technologies, such as SHAPE-MaP [26], SHAPE-Seq [27, 28], and DMS-MaPseq [29], can be readily fed into the pipeline. Finally, it is known that many functions performed by RNA hinge critically on its capacity to interact specifically with proteins or small molecule metabolites and/or to adapt its structure to environmental variations, such as ionic conditions, temperature, or pH. RNA structural flexibility and/or intermolecular interactions can efficiently be monitored with probing technologies [20, 30, 31]. Indeed, the susceptibility of each nucleotide to chemical probes can be determined under different experimental conditions or in the absence/presence of a ligand. Interpreting such data requires a systematic experimental and analytical workflow as well as the application of objective and statistical criteria. IPANEMAP Suite implements a “footprinting” framework allowing to compare and interpret chemical reactivity maps obtained in different conditions or in the presence of an RNA ligand.

## Materials and methods

### Structure probing

RNA structure probing was carried out as previously described in [32, 33].

### Input file format

IPANEMAP Suite accepts input as raw ABI and FSA files from sequencers (Applied Biosystems), as well as output from the CEQ8000/genomeLab (SCIEX) to derive reactivity profiles. Moreover, it directly imports various reactivity files from QuShape or other analysis software, including those from HTS technologies, such as SHAPE-MaP, SHAPE-seq, and DMS-MaPseq. Files accepted are tab-delimited text, including the nucleotide number in the first column and its reactivity in the second. Sequences and MSA files are read in fasta format (.fas). Although input files can be named arbitrarily, following

a defined naming convention facilitates the integration into the pipeline (see Supplementary data for a tutorial).

### Removal of outliers and normalization

Outliers corresponding to intrinsic RT stops are removed as previously described in [33, 34]. Two normalization methods are implemented in IPANEMAP Suite: The first “simple” method takes the 2% of the most reactive peaks of all reactivities (modified and unmodified conditions) and removes them; they empirically correspond to intrinsic stops. Then, the 8% of the most reactive peaks are averaged, and the reactivities are normalized by this average. The second “interquartile” method uses parametric removal: peaks whose value is higher than the 75th percentile by at least 1.5 times the interquartile range are considered as outliers and removed. The 10% of the remaining higher values are averaged and the average is used to normalize all reactivities. Based on empirical observations, the “simple” method (used by default) is recommended for molecules shorter than 300 nt, and the interquartile method is typically reserved for longer molecules [34]. The normalization method can be parameterized by the user, and IPANEMAP Suite enforces the use of only a single method per project in order to ensure consistency. Finally, negative values down to  $-0.3$  are set to zero, while inferior negative values are considered as undetermined and set to  $-10$  to be interpreted as missing by subsequent analyses.

### Data aggregation and consistency checks

One of the key features of the IPANEMAP Suite pipeline is its ability to use multiple datasets as constraints for predictive modeling. When multiple datasets are available for the same RNA under identical probing conditions, IPANEMAP Suite treats them as replicates and aggregates their reactivity data.

Our aggregation strategy is designed to work well with the typically low numbers of replicates. Therefore, it resembles steps of a typical manual aggregation to produce interpretable results.

In detail, IPANEMAP Suite proceeds position by position, as follows: First, positions with too few valid values are labeled as “undetermined”.

For every other position, IPANEMAP Suite computes the mean and standard deviation of its reactivities. By default, positions are accepted and the mean is assigned as an aggregation result, if one of two criteria is satisfied:

- the standard deviation of the reactivities of the replicates at this position is less than or equal to `min_std` (defaulting to 0.15);
- all pairwise means of reactivities fall in the same reactivity range as the total mean, distinguishing the following reactivity ranges: below `reactivity_medium`, between `reactivity_medium`, and `reactivity_high`, and above `reactivity_high` (with defaults `reactivity_medium` = 0.4 and `reactivity_high` = 0.7).

Otherwise, no reactivity is used at this position in the downstream modeling process. Such positions are either marked as “nonconsistent” or “warning”, and are shown in the color-coded structure representations. For this purpose, we flag reactivities as “warning”, if all pairwise means deviate from the overall mean but still fall within a consistent range.

While this specific data treatment works well for us with default values `min_std` = 0.15, `reactivity_medium` = 0.4, and

reactivity\_high = 0.7, these parameters can be modified in the project configuration. This allows the user to turn off the data treatment and use all aggregated reactivities for structure modeling.

Independently, at this stage, it is possible to assay the overall consistency between two replicates by calculating pairwise Pearson and Spearman correlation coefficients. This allows for a fast identification of failed experiments to exclude them from the project.

### Concatenation of data obtained with different primers on the same RNA

IPANEMAP Suite seamlessly supports the use of multiple primers per RNA as typically performed when probing long RNAs. In this case, the originally separate reactivity profiles of the same probing condition are properly aggregated and merged into single profiles of the RNA. This alleviates tedious file manipulation and a potential source of error.

### IPANEMAP: multiple probing-informed prediction

RNA secondary structures are predicted using IPANEMAP (v0.1.9) [23], based on the dataset(s) generated in previous steps. A single project can initiate arbitrarily many IPANEMAP predictions with different combinations of input data and configurations. For each prediction, the IPANEMAP Suite generates two types of RNA secondary structure outputs: a standard dot-bracket file allowing you to draw the secondary structure with most of the secondary structure visualization software and a VARNA session file allowing the visualization of the secondary structure colored according to the reactivity of each nucleotide with the probe. Note that if several probes or conditions are used by IPANEMAP, a VARNA file colored with each reactivity set is generated.

### IPANEMAP core functionality

IPANEMAP was specifically designed to integrate several reactivity profiles from experiments under different conditions into the modeling of RNA secondary structures. It performs this task by sampling structures informed by each of the input reactivity data, followed by a clustering of the structures. It predicts structures based on their occurrence in large, favorable clusters supported by multiple conditions [23]. Roughly, IPANEMAP extends and generalizes the structure sampling and clustering-based prediction approach of Sfold [35] to probing-informed prediction under potentially multiple probing conditions.

### Integration of phylogenetic data

For IPANEMAP Suite, the tool IPANEMAP (v0.1.9) was extended to make use of phylogenetic data, in order to additionally support structure modeling. The use of this option needs to be declared in the configuration step, and the MSA fasta file must be placed manually in the “resources” folder as for sequence files. This option was implemented in IPANEMAP using RNAalifold [14, 36], which predicts consensus structures of the given alignment. These predictions are then treated analogously to predictions based on a different reactivity profile. Effectively, the alignment is integrated in the same way as data from additional probing condition.

### Footprinting analysis

The probing profiles obtained in two different conditions are compared as described in [37, 38]. In brief, the reactivities for each nucleotide in two different conditions are compared using their difference, their relative difference (difference/sum), and a *t*-test.

In more detail, the reactivities averaged from three replicates in two conditions,  $R_1$  and  $R_2$ , were used to calculate the absolute difference  $\Delta R = |R_1 - R_2|$  and the relative difference  $\Delta R / (R_1 + R_2)$ . The reactivity difference between two conditions is considered significant when both variables exceed a threshold of 0.2, and the *P*-value from a two-sided *t*-test is inferior to .05. Nucleotides with undetermined reactivity are excluded from the analysis. Such an approach excludes small reactivity changes between weakly reactive nucleotides as well as large yet unmeaning reactivity changes between highly reactive nucleotides.

### Visualization of candidate structures

IPANEMAP Suite uses VARNA (v0.94.1) to draw RNA secondary structures in the visualization of IPANEMAP predictions and footprinting results. The color annotation capabilities of VARNA were extended to annotate nucleotides according to their reactivity class (undefined, low, medium, and high). In footprinting analyses, this feature allows us to highlight differentially reactive positions in the context of predicted structures. Moreover, within structural comparison, it allows for the representation of bases that share the same or differ in structure between two models, as well as paired bases with different partners in the two structures.

## Results

### Detailed overview of the IPANEMAP Suite

IPANEMAP Suite is a one-stop solution to model RNA secondary structure from output files of CE of a probing experiment to the representation of the secondary structure colored according to the reactivity of each nucleotide. It may also integrate probing data from HT probing technologies, such as ShapeMap, ShapeSeq, or DMS-MaPseq.

IPANEMAP Suite moreover avoids dependency and potential compatibility problems, facilitates maintenance, and ensures reproducibility of the results. IPANEMAP Suite has been designed to be user-friendly and prevent tedious file manipulation from one software to another. It is project-oriented and allows for full traceability all along the project, preventing any mix-up or misinterpretation. A project may include different RNA sequences, probed in different experimental conditions performed by one or several experimentalists. It allows modeling RNA structure using different sets of data obtained with different probes or in different conditions following the process of the IPANEMAP workflow [19, 23]. IPANEMAP Suite additionally implements functionality to incorporate the phylogenetic information contained in a MSA in the modeling process. It outputs two RNA secondary structure models for each RNA sequence and for each series of conditions chosen by the user in the configuration interface. Granted that the two secondary structure models belong to two different structural clusters, they may represent alternative foldings. IPANEMAP Suite may also analyze and highlight the significant differences between two probing conditions if the footprint option is run. Finally, the function “comparison” col-



ors secondary structure models for a fast visual comparison of two models of the same RNA. The software is packaged using Conda to simplify its installation on Linux, or Windows (using WSL), or MacOS. All requirements and steps to install and run IPANEMAP Suite are described in detail at <https://sargueil-citcom.github.io/ipasuite-docs>. In the following, we describe the main stages performed by IPANEMAP Suite step by step; see Fig. 1 for a brief overview.

### Step 1: initialization of projects and configuration

#### *Initialization of IPANEMAP Suite projects*

Projects are created and initialized using a single command (`ipasuite init myproject_name`; see Fig. 1). Upon initialization, files and folders are necessary to run the project (see [Supplementary Fig. S2](#)). Raw data files, sequence fasta files, MSA, and/or reactivity files may then be added in the resources folder (see [Supplementary Fig. S2](#)).

#### *User-friendly configuration*

IPANEMAP Suite offers a user-friendly, web-based interface for configuring sequences, probing reagents, probing conditions, IPANEMAP settings, and footprinting analysis options. IPANEMAP Suite offers the possibility to automatically concatenate results obtained from the same sequence using different primers. This facilitates the work with long RNAs and eliminates a source of error. Any information can be updated throughout the progression of the project, allowing the analysis pipeline to operate with the initial results and the integration of further experiments at any time. To this end, the preparation and configuration can be repeated and updated. To maintain an overview of the project, the information on every single experiment is kept together in a central table (file `samples.tsv`). One can further annotate the table entries with arbitrary additional information, such as the identity of the experimentalist (see [Supplementary data](#)).

### Step 2: deriving reactivities from SHAPE-CE by QuShape

The CE raw data files are analyzed with the software QuShape [12]. This state-of-the-art tool largely facilitates the derivation of reactivity profiles based on the fluorescence profiles coming from CE sequencers. Importantly, files previously analyzed by QuShape may be added to benefit from the downstream pipeline.

### Step 3: normalization and removal of outliers

Outliers, mostly representing RT intrinsic stops, are removed, set to  $-10$ , and the reactivities are normalized according to [34] (see the “Materials and methods” section). In brief, two normalization methods can be applied; for RNA under 300 nt, the highest 2% are considered as outliers, and reactivity values are normalized to the average of the 8% highest values. For longer RNA, values superior to the 75th percentile by over 1.5 the interquartile are removed, and the 10% remaining are averaged and used to normalize the reactivity. Importantly, reactivity files from any workflow, and most notably from high-throughput processes, such as SHAPE-MaP or DMS-MaPseq, can be fed into the pipeline at this stage. However, for consistency, one should check whether all the reactivity files used in the same project follow the same outlier removal and normalization rules.

### Step 4: aggregation of replicates

IPANEMAP Suite then automatically gathers reactivity data obtained for the same RNA in the same probing conditions, thus aggregating data from replicates. It also concatenates reactivity values from elongations performed with different primers on the same RNA in the same condition (see the “Materials and methods” section). The average reactivity value for each nucleotide and its standard deviation are automatically calculated. If only one experiment is performed, the workflow will run with single data. The overall consistency of the replicates may be checked by calculating pairwise Pearson and Spearman correlation coefficients. The results are stored in the file but do not have any automated consequences. Nonconsistent replicates may be annotated as “discarded” in the `samples.tsv` file to ignore them in subsequent runs.

### Step 5: modeling and analysis

#### *RNA structure modeling by IPANEMAP*

Models of the RNA secondary structure are generated using our software IPANEMAP. This tool uniquely allows the use of data from multiple probing experiments under different conditions and/or using different probes, as well as evolutionary information from MSAs. As described in the “Material and methods” section as well as in much more detail in the original publication [23], IPANEMAP aims to predict structures in best agreement with all reactivity and alignment input information as well as the thermodynamic RNA nearest neighbor model [39].

Using the pipeline’s graphical interface, the user can configure—and experiment with—arbitrarily many IPANEMAP prediction runs based on different combinations of datasets. As direct output, IPANEMAP predicts (see the “Materials and methods” section) best and second-best RNA secondary structures in dot-bracket format, which are furthermore visualized as reactivity-annotated structure drawings using VARNA.

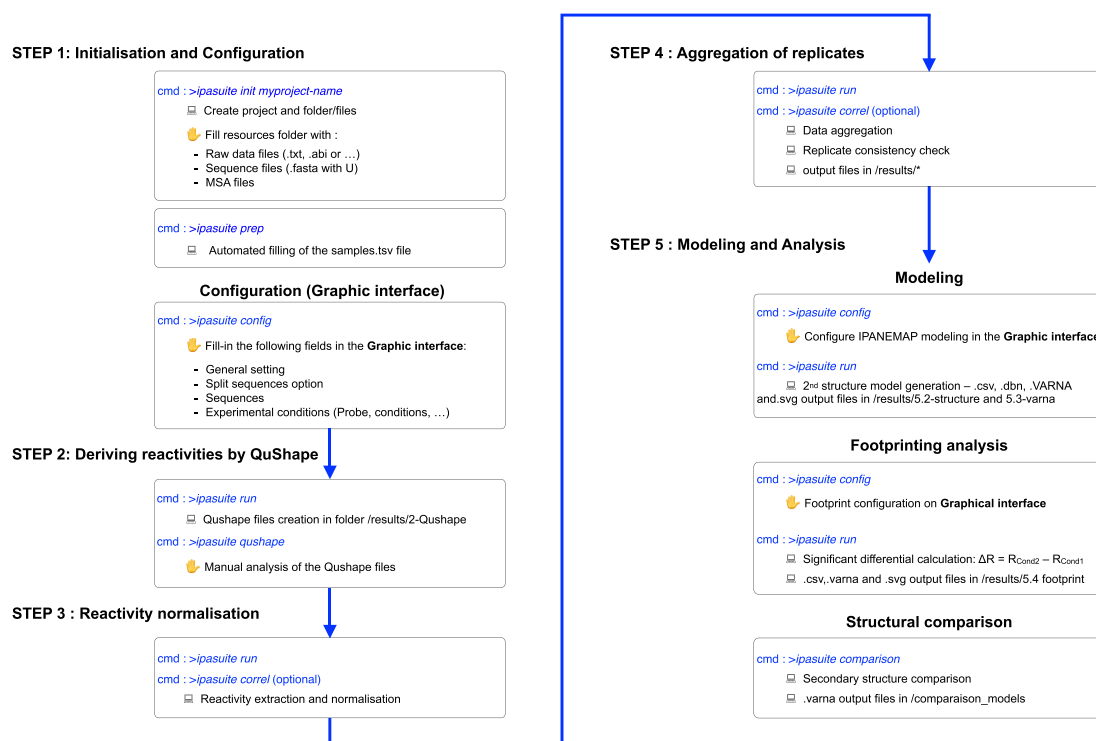
#### *Footprinting analysis*

IPANEMAP Suite compares reactivities of different conditions and produces intuitively interpretable visualizations (see the “Materials and methods” section). Again, the pipeline allows configuring arbitrarily many comparisons and concrete visualization through its graphical configuration interface. Significant differences are reported in publication-quality files in the form of histograms and secondary structure representations. See also Fig. 4 from our examples.

A typical use case of footprinting conditions would be to determine a protein or a small molecule binding site on an RNA, or to evaluate the RNA structural rearrangement upon changing experimental conditions, such as, temperature, ionic conditions, and pH. It can also be used to understand the impact of modifications, such as punctual mutations or trimming of the sequence.

#### *Secondary structure model comparison*

Finally, we provide functionality to visually compare two structure models, which is especially instrumental in the case of medium-sized or long RNAs. Given two secondary structures in dot-bracket notation (dbn files), our comparison tool generates two output files visualizing the structural differences by color annotations. The first highlights the nucleotides with



**Figure 1.** The project-centered workflow of IPANEMAP Suite: a visually intuitive interface for configuring the entire chemical probing data analysis pipeline within IPANEMAP Suite. Users can easily customize each step of the process, from data import and preprocessing to analysis and visualization.

a common structure, while the second distinguishes three categories that are color-encoded: the nucleotides involved in the same structures (blue by default), those involved in different structures but paired in both models (red by default), and those in a different pairing status in both models (white by default).

### IPANEMAP Suite use cases and example results

IPANEMAP Suite provides valuable tools for RNA structure investigation and generates numerous plots and figures to streamline data analysis and interpretation, minimizing the need for time-consuming manual plotting.

To showcase the fundamental, unique key features of the pipeline, namely the modeling based on multiple probing data, two “benchmark” RNAs with crystal structure models were probed under various conditions. Namely, we probed the SAM-IV aptamer from *Mycobacterium* and the *yitJ* SAM-I riboswitch from *B. subtilis* with previously solved X-ray crystallographic structures (PDB IDs: 6WLR [43] and 4KQY [42], respectively). Please note that datasets for those examples are available at <https://github.com/Sargueil-CiTCOM/ipasuite>.

### Integration of multiple sets of data improve secondary structure prediction

The *yitJ* SAM-I aptamer was probed with 1M7 (Fig. 2A, top) and DMS (Fig. 2, bottom) in the absence of the ligand. Three RNA structure models were generated based on

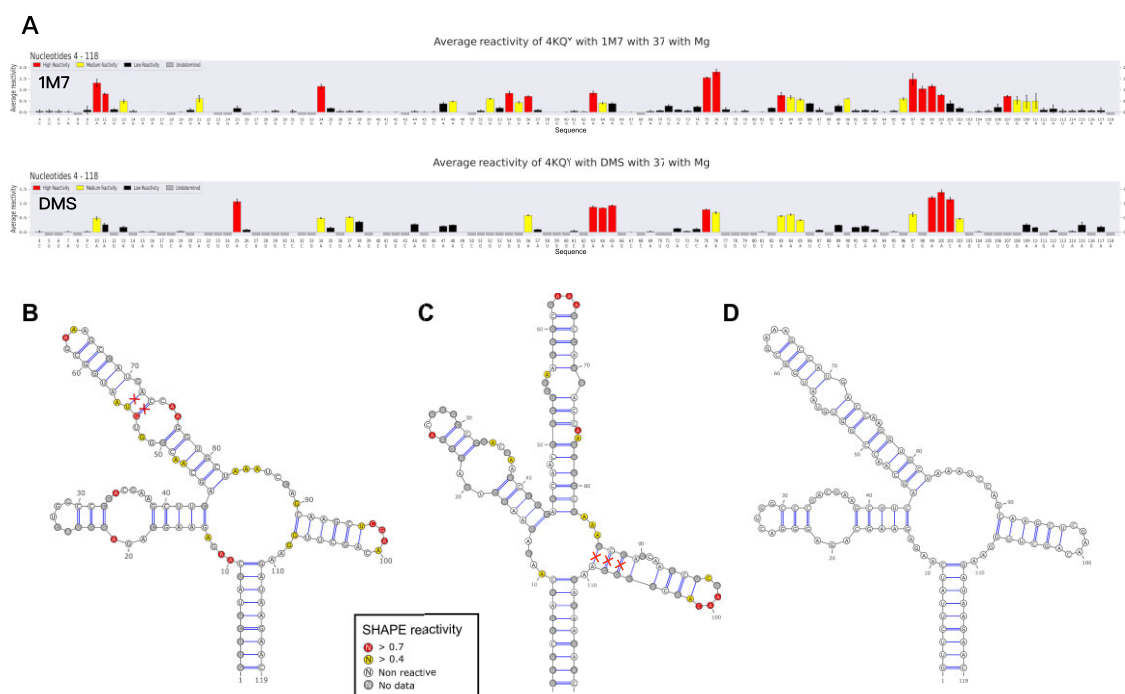
- 1M7 (Fig. 2B),
- DMS probing data (Fig. 2C) separately, and
- using the multi-probing integration of IPANEMAP (Fig. 2D).

Here, IPANEMAP Suite’s project-based design allows us to easily produce the different models from the same project by configuring different IPANEMAP modeling setups in the graphical interface.

Figure 2 illustrates that neither of the models obtained using only one probing profile is correct. Due to the unique multi-probing features of IPANEMAP Suite, we show here that the integration of probing data using two different probes fundamentally improves the prediction of the *yitJ* SAM-I riboswitch, as the model obtained with both datasets exactly agrees with the secondary structure derived from the crystal structure.

### Further refining structural models by integrating phylogenetic data

The *Mycobacterium* SAM-IV aptamer was subjected to DMS and 1M7 SHAPE probing. IPANEMAP was run using the 1M7 reactivity profile alone (Fig. 3A), the DMS reactivity profile alone (Fig. 3B), 1M7 and DMS datasets conjointly (Fig. 3C), or 1M7 and DMS datasets in conjunction with phylogenetic data (Fig. 3D). Different runs of IPANEMAP are easily configured in the graphic interface of IPANEMAP Suite, and the models obtained may be readily compared in regard to the secondary structure extracted from the PDB 3D structure (Fig. 3C and D). The crystal structure (6WLR) features four stems termed P1, P2, P3, and P4. The model obtained with 1M7 only predicts P3, P4, and two short spurious hairpins (Fig. 3A); the DMS-informed model features P2, P3, P4, and one short false stem (Fig. 3B). While the model obtained using both 1M7 and DMS reactivities (Fig. 3C) is similar to the one obtained with 1M7 only (Fig. 3A), the model using both datasets and the covariation information is very close to the secondary structure derived from the crystal structure



**Figure 2.** Structure modeling based on multiple probing datasets. **(A)** 1M7 (top) and DMS (bottom) reactivity profiles of *Bacillus subtilis* yitJ riboswitch. Secondary structure model colored with reactivities, obtained with 1M7 **(B)** and DMS **(C)** reactivity constraints. The red crosses mark base pairs that are predicted using 1M7 or DMS data but are not observed in the crystal structure. **(D)** Secondary structure model based on the integration of two probe datasets, which exactly matches the crystal structure. The respective prediction performances of panels (B), (C), and (D) were calculated as 0.93, 0.92, and 1.00 in terms of the widely used F1 score [40], the harmonic mean of recall and precision of base pairs with respect to the crystal structure. (See [Supplementary data](#) for the configuration of the structure modeling based on multiple profiles.)

(Fig. 3D). The same modeling strategy was applied to various benchmark RNAs of various lengths and functions, and belonging to various organisms: *Thermatoga maritima* ribonuclease P (347 nucleotides, PDB ID: 3Q1Q), *Sacharomyces cerevisiae* U1 snRNA (568 nucleotides, PDB ID: 7OQC), *Syntrophothermus lipocalidus* PRPP riboswitch (107 nucleotides, PDB ID: 6DNR), and *Escherichia coli* tRNAPhe (76 nucleotides, PDB ID: 3LOU). The accuracy (F1 score) for each prediction is reported in Table 1, and the structure models obtained with 1M7 and DMS datasets conjointly with phylogenetic data are presented in [Supplementary Fig. S1](#) compared to the secondary structure derived from the crystal structure. Although the relative accuracy of the modeling using the various constraints may vary, the prediction, including both datasets and the phylogenetic information, is always the most similar to the secondary structure derived from the crystal structure. Furthermore, [Supplementary Fig. S1](#) shows a remarkable accuracy of the models obtained for benchmark RNAs, including long and complex ones, such as RNase P or U1 snRNA. This demonstrates that IPANEMAP's unique ability to integrate several datasets and phylogenetic information can be decisive in obtaining accurate models.

### Footprinting analysis of SAM-I aptamer structural switch upon binding its ligand

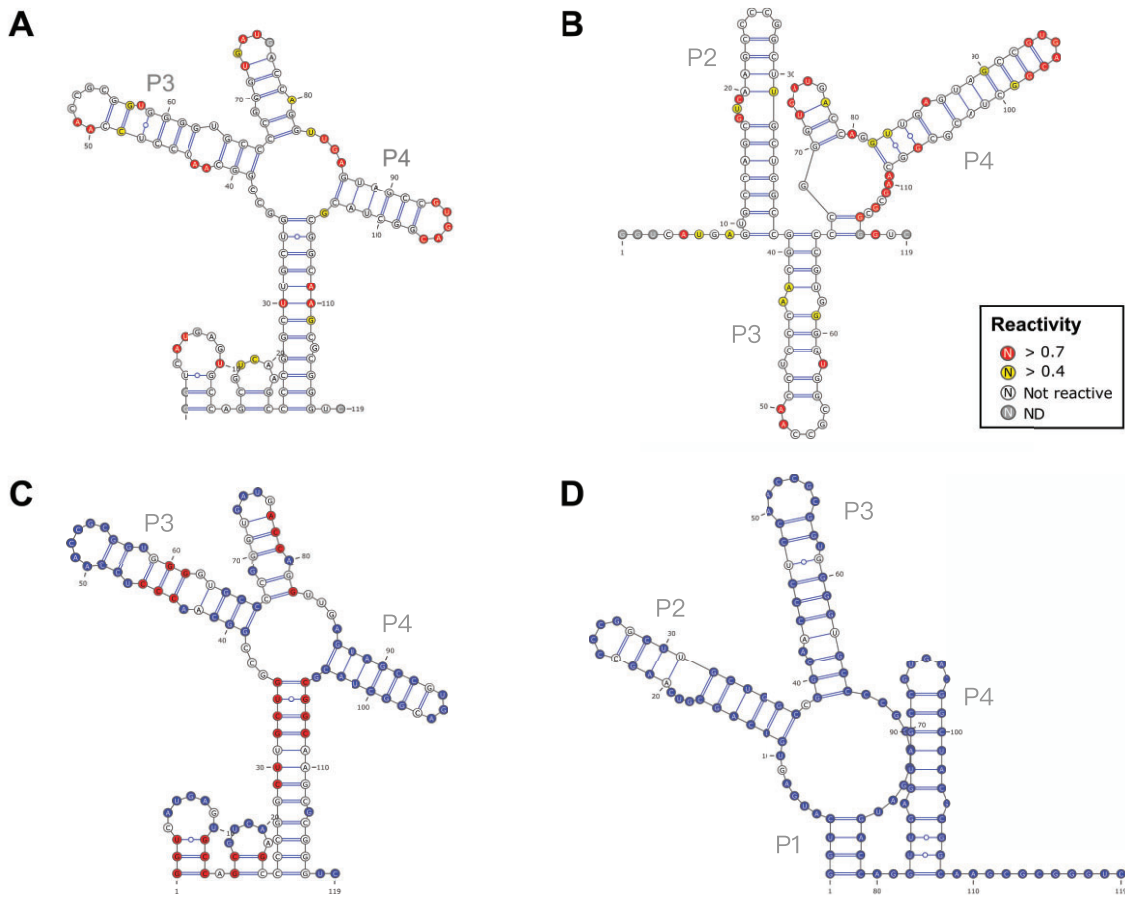
To illustrate the usefulness of the footprint analysis tool using the IPANEMAP Suite pipeline, we performed SHAPE probing on the *B. subtilis* yitJ SAM-I riboswitch in the presence and absence of its ligand, the SAM (Fig. 4A). SHAPE probing has been widely used for aptamer characterization, including for

the identification of nucleotides involved in ligand recognition and in tertiary structure formation [41, 44]. To compare the two probing profiles, IPANEMAP Suite automatically generates a SHAPE differential histogram, highlighting nucleotides with significant reactivity differences (Fig. 4B). These differences are also automatically displayed on the RNA secondary structure model (Fig. 4C).

As previously described for the SAM-I aptamer [42], the IPANEMAP Suite footprint analysis readily highlights significant reactivity decreases upon SAM recognition, allowing for the rapid identification of nucleotides involved in direct and indirect ligand contacts (Fig. 4D), as well as in the ligand-induced structural switch (Fig. 4B, D, E, and F). By automating the plotting process, the IPANEMAP Suite footprint tool eliminates the need for time-consuming manual efforts. More importantly, IPANEMAP Suite ensures the consistency and the data treatment objectivity essential to enable reliable conclusions to be drawn from such experiments.

## Conclusion

IPANEMAP Suite is an integrated workflow that organizes the user journey from the raw data to the representation of the secondary structure, while facilitating even advanced modeling strategies. It does not necessitate any specific expertise either in RNA structure or in computer science or bioinformatics. All along the pipeline, files representing intermediate results are stored in identified folders. Initialization of the project requires unambiguously naming the raw data files, and this denomination is kept throughout the pro-



**Figure 3.** Structure model refinement by using MSA. **(A)** *Mycobacterium* SAM-IV secondary structure model derived using 1M7 reactivity constraints. **(B)** The secondary structure model was predicted using DMS reactivity constraints. **(C)** Refined model generated using the combination of 1M7 and DMS reactivity constraints; the color annotation represents the result of the comparison with the crystal structure (PDB ID: 6WLR). The nucleotides in blue are predicted as in the crystal structure, those in red are different but are in double strands in both structures, while nucleotides in white have different statuses in the two structures. **(D)** Further refined model integrating the information from 1M7 and DMS probing with phylogenetic data; the color annotation represents the result of the comparison with the crystal structure (PDB ID: 6WLR). The respective accuracies (F1 scores; see Fig. 2) of the predictions in panels (A), (B), (C), and (D) were calculated as 0.405, 0.719, 0.319, and 0.923, respectively, the latter quantifying a high similarity to the crystal structure (PDB ID: 6WLR).

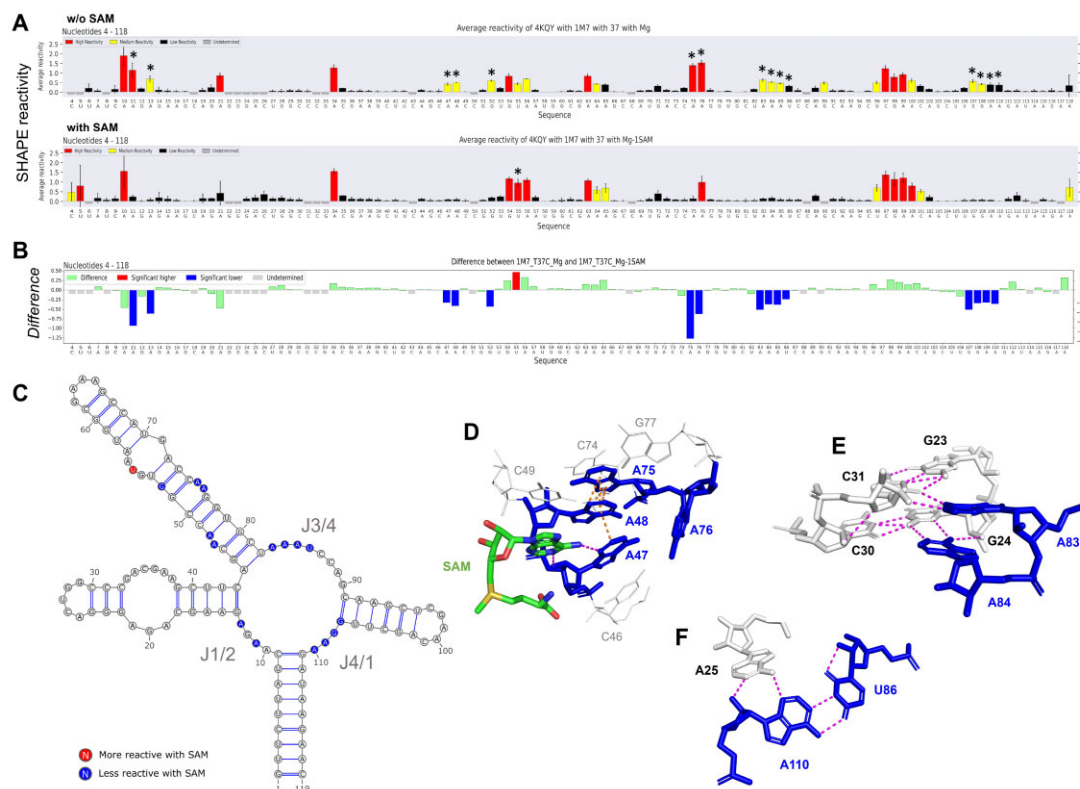
cess. IPANEMAP Suite allows you to seamlessly work on long RNA by automatically concatenating results obtained with different elongations covering the same RNA. These features avoid many file manipulations and thus prevent potential errors. It features a new version of IPANEMAP that can be informed by various structure probing datasets but also integrates the covariation information contained in an MSA. As shown above, the inclusion of multiple datasets and/or alignments may be critical to obtaining an accurate model. The footprint option facilitates comparing the reactivities obtained on the same sequence in different contexts. This enables, for instance, easily pinpointing a protein binding site on your favorite RNA. Although such analysis is routinely performed in labs studying RNA structure, this has not been implemented in any automated workflow to date. In addition to ease of use, IPANEMAP Suite provides the indispensable systematic routine to the footprinting approach. Indeed, drawing conclusions from such approaches requires a strict consistency in both the experimental and the analytical workflow. IPANEMAP Suite allows for a systematic and objective treatment of the raw data file, while offering flexibility through parameters that can be project-wide configured by the experimentalist. Finally, the comparison tool allows for an effortless

visual comparison of two structure models even for medium-sized to long RNAs.

### Comparison with other workflows

Two other analytical workflows propose a continuous pipeline from the raw reactivity files obtained from RNA structure probing experiments to a secondary structure model. RNAthor features a slightly more sophisticated algorithm for the removal of the outlier values and offers a statistical analysis to evaluate results consistency throughout replicates [25]. RNAsprobe [24] allows for modeling with various RNA prediction software, and notably with ShapeKnots, including the interesting possibility to predict pseudoknots [18]. In addition to the secondary structure model colored with reactivity, RNAsprobe may also color a 3D structure model provided by the user according to nucleotide reactivities. While IPANEMAP Suite supports a project-based workflow, which enforces the traceability from raw data to the model, the competing tools are only available as web servers and do not integrate the entire workflow starting from the raw data. In addition, IPANEMAP Suite offers advanced functionality, such as the possibility to carry out secondary structure model-





**Figure 4.** SHAPE analysis of the structural switch of *B. subtilis* yitJ SAM-I aptamer in the presence of *S*-adenosyl-methionine (SAM). **(A)** Normalized SHAPE reactivity profiles in the absence (top) and the presence (bottom) of the ligand (SAM). Asterisks indicate nucleotides with significant differential reactivity. **(B)** Histogram of the differential SHAPE reactivities ( $\Delta R$ ) calculated by subtracting the SAM-bound state profile from that of the unbound state. Positions exhibiting significant differential reactivities (see the “Materials and methods” section) are colored in blue for nucleotides less reactive in the presence of the ligand and in red for those more reactive in the presence of the ligand. **(C)** Secondary structure model colored with significant differences. Single-stranded junctions (J1/2, J3/4, and J4/1) are known to be less reactive upon SAM binding [41]. **(D)** SAM interaction with A47 induces pi-stacking (orange dotted lines) between nucleotides A47–A48–A76, which enhance the stability of those nucleotides. Base triple interactions formed after SAM binding **(E)** A83 and A84 with base pairs G23:C31 and G24:C30, respectively, and **(F)** A25 with base pair U86:A110 [42]. Hydrogen bonds are represented by magenta dotted lines. (See [Supplementary data](#) and [Supplementary Fig. S5B](#) for configuration of the footprint analysis.)

**Table 1.** Modeling accuracy of benchmark RNAs using 1M7 and DMS datasets, and phylogenetic data as F1 scores [40]

PDB	DMS	1M7	1M7 and DMS	1M7 and DMS and Phylogeny
3Q1Q	0.667	0.527	0.677	0.735
7OQC	0.450	0.468	0.466	0.731
6DNR	0.815	0.786	0.815	0.863
3L0U	0.955	0.977	0.955	0.977

*Thermatoga maritima* ribonuclease P (347 nucleotides, PDB ID: 3Q1Q), *Saccharomyces cerevisiae* U1 snRNA (568 nucleotides, PDB ID: 7OQC), *Syntrophothermus lipocalidus* PRPP riboswitch (107 nucleotides, PDB ID: 6DNR), and *E. coli* tRNA<sup>Phe</sup> (76 nucleotides, PDB ID: 3L0U) were probed with 1M7, DMS, and MSA obtained from Rfam [45]. RNA structures were modeled using 1M7, DMS, 1M7, and DMS datasets or 1M7 and DMS in conjunction with phylogenetic data using IPANEMAP Suite.

ing driven by different datasets at once, or the footprinting analysis.

Over the last few years, many high-throughput experimental [29, 46–49] and computing pipelines [21, 50, 51] have been developed. They offer many possibilities to analyze high-throughput structure probing, and notably, Superfold can also take into account different sets of probing data for secondary structure probing, although it can only be implemented with three different SHAPE reagents [21]. IPANEMAP Suite can take into account data coming from both traditional and high-

throughput structure probing, eliminate the risk of error when handling files, and allow organizing data and modeling in a consistent and project-oriented way. Finally, IPANEMAP Suite offers the unique possibility to model RNA secondary structure, integrating both multiple probing datasets and co-variation data to further improve prediction accuracy. Future versions of IPANEMAP Suite will integrate principled metrics to assess the significance of covariations, allowing for more robust predictions [52].

## Acknowledgements

We thank all members of Bruno Sargueil laboratory for crash testing the pipeline and manuscript proofreading.

**Author contributions:** P.H. was responsible for performing all the experiments, validating the data, and analyzing the data. P.H. and N.P. prepared the figures. P.H. and B.S. was responsible for data curation. P.H. and N.C. validated the software. F.X.L.M., N.P., and S.W. was responsible for software design and programming. S.W. and B.S. supervised the research for the computational and experimental aspects, respectively. B.S., S.W., and N.P. wrote the manuscript. Y.P. and N.C. reviewed the manuscript. Y.P., B.S., and S.W. secured the funding for the project. B.S. was responsible for the initial conceptualization, design and management of the project.



## Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

## Conflict of interest

None declared.

## Funding

This work was funded by ANR Grants (ANR-19-CE45-0023, ANR-19-CE30-0021, and ANR-21-CE45-0034) to Bruno Sargueil, Yann Ponty, and Sebastian Will. Work in Bruno Sargueil laboratory was also funded by CNRS and Université Paris Cité. Francois-Xavier Lyonnet du Moutier and Pierre Hardouin was funded by ANR-19-CE30-0021-03 and ANR-20-CE12-0026-02, respectively. Nan Pan was funded by ANR-21-CE45-0034-01.

## Data availability

All software is made available for installation as Conda packages in channels Sargueil-CiTCOM or available in channel bioconda. The source code is accessible in Git repositories:

- IPANEMAP Suite: <https://github.com/Sargueil-CiTCOM/ipasuite>
- IPANEMAP: <https://gitlab.inria.fr/amibio/ipanemap>
- VARNA: <https://github.com/yannponty/VARNA>
- ViennaRNA: <https://github.com/ViennaRNA/ViennaRNA>

The code and data have been deposited in Zenodo: <https://zenodo.org/records/14936737>. A figure showing refined models integrating the information from 1M7 and DMS probing conjointly with phylogenetic data, and a tutorial are available separately in Supplementary data.

## References

- Deigan KE, Li TW, Mathews DH *et al.* Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci USA* 2009;106:97–102. <https://doi.org/10.1073/pnas.0806929106>
- Washietl S, Hofacker IL, Stadler PF *et al.* RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction. *Nucleic Acids Res* 2012;40:4261–72. <https://doi.org/10.1093/nar/gks009>
- Lorenz R, Wolfinger MT, Tanzer A *et al.* Predicting RNA secondary structures from sequence and probing data. *Methods* 2016;103:86–98. <https://doi.org/10.1016/j.ymeth.2016.04.004>
- Brunel C, Romby P. Probing RNA structure and RNA–ligand complexes with chemical probes. *Methods Enzymol* 2000;318:3–21.
- McGinnis JL, Dunkle JA, Cate JHD *et al.* The mechanisms of RNA SHAPE chemistry. *J Am Chem Soc* 2012;134:6617–24. <https://doi.org/10.1021/ja2104075>
- Frezza E, Courban A, Allouche D *et al.* The interplay between molecular flexibility and RNA chemical probing reactivities analyzed at the nucleotide level via an extensive molecular dynamics study. *Methods* 2019;162–163:108–27. <https://doi.org/10.1016/j.ymeth.2019.05.021>
- Pinamonti G, Bottaro S, Micheletti C *et al.* Elastic network models for RNA: a comparative assessment with molecular dynamics and SHAPE experiments. *Nucleic Acids Res* 2015;43:7260–9. <https://doi.org/10.1093/nar/gkv708>
- Mlynsky V, Bussi G. Molecular dynamics simulations reveal an interplay between SHAPE reagent binding and RNA flexibility. *J Phys Chem Lett* 2018;9:313–18. <https://doi.org/10.1021/acs.jpclett.7b02921>
- Cantara WA, Hatterschide J, Wu W *et al.* RiboCAT: a new capillary electrophoresis data analysis tool for nucleic acid probing. *RNA* 2017;23:240–9. <https://doi.org/10.1261/rna.058404.116>
- Mitra S, Shcherbakova IV, Altman RB *et al.* High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res* 2008;36:e63. <https://doi.org/10.1093/nar/gkn267>
- Pang PS, Elazar M, Pham EA *et al.* Simplified RNA secondary structure mapping by automation of SHAPE data analysis. *Nucleic Acids Res* 2011;39:e151. <https://doi.org/10.1093/nar/gkr773>
- Karabiber F, McGinnis JL, Favorov OV *et al.* QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA* 2013;19:63–73. <https://doi.org/10.1261/rna.036327.112>
- Kim H, Cordero P, Das R *et al.* HiTRACE-Web: an online tool for robust analysis of high-throughput capillary electrophoresis. *Nucleic Acids Res* 2013;41:W492–8. <https://doi.org/10.1093/nar/gkt501>
- Lorenz R, Bernhart SH, Höner zu Siederdisen C *et al.* ViennaRNA package 2.0. *Algorithm Mol Biol* 2011;6:26. <https://doi.org/10.1186/1748-7188-6-26>
- Lorenz R, Luntzer D, Hofacker IL *et al.* SHAPE directed RNA folding. *Bioinformatics* 2015;32:145–7. <https://doi.org/10.1093/bioinformatics/btv523>
- Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 2010;11:129. <https://doi.org/10.1186/1471-2105-11-129>
- Spasic A, Assmann SM, Bevilacqua PC *et al.* Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic Acids Res* 2017;46:314–23. <https://doi.org/10.1093/nar/gkx1057>
- Hajdin CE, Bellaousov S, Huggins W *et al.* Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc Natl Acad Sci USA* 2013;110:5498–503. <https://doi.org/10.1073/pnas.1219988110>
- Deforges J, de Breyne S, Ameer M *et al.* Two ribosome recruitment sites direct multiple translation events within HIV1 Gag open reading frame. *Nucleic Acids Res* 2017;45:7382–400. <https://doi.org/10.1093/nar/gkx303>
- De Bisschop G, Allouche D, Frezza E *et al.* Progress toward SHAPE constrained computational prediction of tertiary interactions in RNA structure. *NonCoding RNA* 2021;7:71. <https://doi.org/10.3390/ncrna7040071>
- Smola MJ, Rice GM, Busan S *et al.* Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protoc* 2015;10:1643–69. <https://doi.org/10.1038/nprot.2015.103>
- Busan S, Weidmann CA, Sengupta A *et al.* Guidelines for SHAPE reagent choice and detection strategy for RNA structure probing studies. *Biochemistry* 2019;58:2655–64. <https://doi.org/10.1021/acs.biochem.8b01218>
- Saaidi A, Allouche D, Regnier M *et al.* IPANEMAP: integrative probing analysis of nucleic acids empowered by multiple accessibility profiles. *Nucleic Acids Res* 2020;48:8276–89. <https://doi.org/10.1093/nar/gkaa607>
- Wirecki TK, Merdas K, Bernat A *et al.* RNAProbe: a web server for normalization and analysis of RNA structure probing data. *Nucleic Acids Res* 2020;48:W292–9. <https://doi.org/10.1093/nar/gkaa396>
- Gumna J, Zok T, Figurski K *et al.* RNAthor—fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis. *PLoS One* 2020;15:e0239287. <https://doi.org/10.1371/journal.pone.0239287>

26. Siegfried NA, Busan S, Rice GM *et al.* RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 2014;11:959–65. <https://doi.org/10.1038/nmeth.3029>
27. Mortimer SA, Trapnell C, Aviran S *et al.* SHAPE-Seq: high-throughput RNA structure analysis. *Curr Protoc Chem Biol* 2012;4:275–97. <https://doi.org/10.1002/9780470559277.ch120019>
28. Loughrey D, Watters KE, Settle AH *et al.* SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res* 2014;42:e165. <https://doi.org/10.1093/nar/gku909>
29. Zubradt M, Gupta P, Persad S *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo*. *Nat Methods* 2017;14:75–82. <https://doi.org/10.1038/nmeth.4057>
30. von Ahsen U, Noller HF. Footprinting the sites of interaction of antibiotics with catalytic group I intron RNA. *Science* 1993;260:1500–3. <https://doi.org/10.1126/science.8502993>
31. Smola MJ, Calabrese JM, Weeks KM. Detection of RNA–protein interactions in living cells with SHAPE. *Biochemistry* 2015;54:6867–75. <https://doi.org/10.1021/acs.biochem.5b00977>
32. Coronel-Tellez RH, Pospiech M, Barrault M *et al.* sRNA-controlled iron sparing response in staphylococci. *Nucleic Acids Res* 2022;50:8529–46.
33. Allouche D, De Bisschop G, Saaidi A *et al.* RNA Secondary Structure Modeling Following the IPANEMAP Workflow. New York, NY: Springer US, 2024, 85–104.
34. Low JT, Weeks KM. SHAPE-directed RNA secondary structure prediction. *Methods* 2010;52:150–8. <https://doi.org/10.1016/j.ymeth.2010.06.007>
35. Ding Y, Lawrence CE. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res* 2003;31:7280. <https://doi.org/10.1093/nar/gkg938>
36. Bernhart SH, Hofacker IL, Will S *et al.* RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 2008;9:474. <https://doi.org/10.1186/1471-2105-9-474>
37. De Bisschop G, Sargueil B. *RNA Footprinting Using Small Chemical Reagents*. New York, NY: Springer US, 2021, 13–23.
38. Angulo J, Ulryck N, Deforges J *et al.* LOOP IIId of the HCV IRES is essential for the structural rearrangement of the 40S-HCV IRES complex. *Nucleic Acids Res* 2015;44:1309–25. <https://doi.org/10.1093/nar/gkv1325>
39. Matthews DH, Disney MD, Childs JL *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA* 2004;101:7287–92. <https://doi.org/10.1073/pnas.0401799101>
40. Mathews DH. How to benchmark RNA secondary structure prediction accuracy. *Methods* 2019;162–163:60. <https://doi.org/10.1016/j.ymeth.2019.04.003>
41. Hennelly SP, Novikova IV, Sanbonmatsu KY. The expression platform and the aptamer: cooperativity between Mg<sup>2+</sup> and ligand in the SAM-I riboswitch. *Nucleic Acids Res* 2012;41:1922–35. <https://doi.org/10.1093/nar/gks978>
42. Lu C, Ding F, Chowdhury A *et al.* SAM recognition and conformational switching mechanism in the *Bacillus subtilis* [S] box/SAM-I riboswitch. *J Mol Biol* 2010;404:803–18. <https://doi.org/10.1016/j.jmb.2010.09.059>
43. Kappel K, Zhang K, Su Z *et al.* Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat Methods* 2020;17:699–707. <https://doi.org/10.1038/s41592-020-0878-9>
44. Winkler WC, Nahvi A, Sudarsan N *et al.* An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Mol Biol* 2003;10:701–7. <https://doi.org/10.1038/nsb967>
45. Ontiveros-Palacios N, Cooke E, Nawrocki EP *et al.* Rfam 1.5: RNA families database in 2025. *Nucleic Acids Res* 2024;53:D258–67. <https://doi.org/10.1093/nar/gkae1023>
46. Lucks JB, Mortimer SA, Trapnell C *et al.* Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc Natl Acad Sci USA* 2011;108:11063–8. <https://doi.org/10.1073/pnas.1106501108>
47. Watters KE, Yu AM, Strobel EJ *et al.* Characterizing RNA structures *in vitro* and *in vivo* with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods* 2016;103:34–48. <https://doi.org/10.1016/j.ymeth.2016.04.002>
48. Flynn RA, Zhang QC, Spitale RC *et al.* Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nat Protoc* 2016;11:273–90. <https://doi.org/10.1038/nprot.2016.011>
49. Ziv O, Gabrylska MM, Lun ATL *et al.* COMRADES determines *in vivo* RNA structures and interactions. *Nat Methods* 2018;15:785–8. <https://doi.org/10.1038/s41592-018-0121-0>
50. Cheng CY, Chou F-C, Das R. Chapter 2: Modeling complex RNA tertiary folds with Rosetta. In: Chen SJ, Burke-Aguero DH (eds.), *Computational Methods for Understanding Riboswitches, Methods in Enzymology*, Vol. 553. Academic Press, 2015, 35–64.
51. Incarnato D, Morandi E, Simon LM *et al.* RNA Framework: an all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications. *Nucleic Acids Res* 2018;46:e97. <https://doi.org/10.1093/nar/gky486>
52. Rivas E. Evolutionary conservation of RNA sequence and structure. *WIREs RNA* 2021;12:e1649. <https://doi.org/10.1002/wrna.1649>