



## RNA secondary structure modelling following the IPANEMAP workflow

Delphine M Allouche, Grégoire de Bisschop, Afaf Saaïdi, Yann Ponty, Sargueil  
Bruno

### ► To cite this version:

Delphine M Allouche, Grégoire de Bisschop, Afaf Saaïdi, Yann Ponty, Sargueil Bruno. RNA secondary structure modelling following the IPANEMAP workflow. RNA Folding - Methods and Protocols, Springer Nature, In press, Methods in Molecular Biology. hal-02324783

**HAL Id: hal-02324783**

**<https://inria.hal.science/hal-02324783>**

Submitted on 22 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **RNA secondary structure modelling following the IPANEMAP workflow**

Delphine Allouche<sup>1,2</sup>, Grégoire De Bisschop<sup>1</sup>, Afaf Saaidi<sup>3</sup>,  
Yann Ponty<sup>3,□</sup> and Sargueil Bruno<sup>1,□</sup>

1 - CiTCOM, Cibles Thérapeutiques et conception de médicaments, CNRS, Université de PARIS, 4 avenue de l'observatoire 75 006 Paris, FRANCE

2 - Present Address : UMR8261, CNRS, Université de PARIS, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie 75005 Paris, FRANCE

3 - CNRS UMR 7161, LIX, Ecole Polytechnique, 1 rue Estienne d'Orves, 91120 Palaiseau, FRANCE

□ To whom correspondence should be addressed: [bruno.sargueil@parisdescartes.fr](mailto:bruno.sargueil@parisdescartes.fr); [yann.ponty@lix.polytechnique.fr](mailto:yann.ponty@lix.polytechnique.fr)

RNA secondary structure modelling has been a challenge since the early days of molecular biology. Although algorithms for RNA structure modelling are more and more efficient and accurate, they significantly benefit from the integration of experimental structure probing data. RNA structure probing consists in submitting an RNA to enzymes or small molecules that specifically react with individual nucleotides according to their pairing status. Most enzymes used are single strand specific RNases (RNases T1, U2, nuclease S1 ...) with the notable exception of the double strand specific RNase V1. Although they are low molecular weight proteins, they are too bulky to access some nucleotides of a folded RNA. Small molecules can essentially reach any nucleotide and most of them are also single-strand specific although psoralen has recently been successfully used a double strand probe (Lu et al., 2016). For the longest time, RNA probing experiments remained tedious and rather qualitative than quantitative.

RNA structure probing recently reached the medium, and then high, throughput. Pioneered and mostly developed within the Weeks lab, the SHAPE technology uses small molecules that react with flexible ribose, thus essentially reporting single-stranded nucleotides with some subtleties (Frezza et al., 2019; Steen et al., 2012). A medium throughput version of the SHAPE protocol was first developed based on capillary electrophoresis, later to be transformed into a high throughput method using next generation sequencing. The same workflows can be applied to more traditional probes such as DiMethyl Sulfate (DMS) and N-Cyclohexyl-N'-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate (CMCT) that reveal unpaired A,C and G,U respectively.

It appeared that different probes provide complementary information that further improves RNA structure prediction. We therefore developed IPANEMAP, an experimental and computational workflow that models RNA secondary structure from different sets of RNA structure probing performed with different probes, and/or in different conditions and/or on mutants (Saaidi et al. Submitted). This workflow relies on medium or high throughput structure probing, and combines statistical sampling, clustering (Ding and Lawrence, 2003) and pseudo-potentials (Deigan et al, 2009). The method was shown to produce more accurate and stable predictions than other workflows developed to date, even when a single reactivity profile is available, while the availability of multiple reactivities was shown to increase robustness and, to a lesser extent, accuracy of the modeling (Saaidi et al. Submitted).

Below, we detail a whole IPANEMAP workflow, starting with experimental probing with DMS and/or CMCT and/or SHAPE reagent. Such probing can be carried out in various relevant conditions (varying temperature, Mg<sup>2+</sup> concentration, introducing point mutations in the RNA to be modeled etc) (Saaidi et al. Submitted). Two versions of the experimental procedure (medium throughput and high throughput) are proposed, DMS and CMCT probing were adapted from Ehresmann et al. and Brunel et al. while the SHAPE probing is described in K. Weeks team publications (Karabiber et al., 2013; Low and Weeks, 2010a; Mortimer and Weeks, 2007; Smola et al., 2015a; Wilkinson et al., 2006, 2008). We then detail instructions for executing the IPANEMAP algorithm to obtain the RNA secondary structure model.

## Medium throughput RNA structure probing with DMS, CMCT, or a SHAPE reagent

### BUFFERS AND SOLUTIONS

- Specific for DMS:
  - DMS (diluted in ethanol 1:12) CAS # 77-78-1
  - DMS buffer (10X) : 400 mM HEPES pH7.5, 1 M KCl, 10 mM MgCl<sub>2</sub>
  - DMS stop reaction: 400 mM Tris pH7.5
  - Sodium hydroxide
- Specific for CMCT:
  - CMCT (42 mg/mL) CAS # 2491-17-0
  - CMCT buffer (10X) : 500 mM potassium borate pH8, 1 M KCl, 10 mM MgCl<sub>2</sub>
  - 10 % acetic acid
- Specific for SHAPE
  - Shape reagent 10X in anhydrous DMSO : 40 mM 1M7 (CAS # 4693-01-0) or 40 mM NMIA (CAS # 10328-92-4) or 400 mM BzCN (CAS # 613-90-1)
- Common to the three types of probing:
  - Ammonium Acetate 5M
  - 70 % and 100 % cold ethanol
  - Dideoxynucleotide 10 mM, either ddTTP, ddATP or ddCTP
  - MMLV Reverse Transcriptase RNase H (-) 200 u/μL (e.g Promega M1705)
  - MMLV RT buffer 5X (e.g Promega M5313)
  - Deoxynucleotide mix dATP, dTTP, dCTP and dITP, 10 mM each
  - Two differently labelled fluorescent primers at 2 μM in water, e.g. D2-PA and D4-PA (Sigma-Aldrich) if running the sequences on a CEQ8000 capillary sequencer.  
*Note: Primer sequence should be chosen at least 50 to 75 nucleotides 3' (downstream) to the region of interest*
  - RT Stop buffer 5X: 1.2 M sodium acetate pH5, 40 mM EDTA, 4 mg/mL glycogen.
  - Sample loading solution (deionised formamide)
  - Glycogen 20 mg/mL
  - Yeast tRNA 10 mg/ml

### MATERIAL

- Usual disposable labware
- Capillary electrophoresis (ABI or Ceq8000 AB Sciex)
- Thermocycler
- Dry bath
- Refrigerated centrifuge
- Fume hood

### METHODS

#### 1) RNA refolding

RNA purification protocol from cells or from *in vitro* transcription often include a denaturation step. In addition, RNA may also go through freeze and thaw cycles which also alter the native folding. In such cases, a denaturation-renaturation step, helps to homogenize RNA folding. In case a non denaturing protocol has been used to prepare the RNA this step may be skipped. Numerous refolding protocols have been proposed, their relative relevance may depend on the RNA considered (Chillón et al., 2015; Herschlag, 1995; Jaeger et al., 1993; Uhlenbeck, 1995) we propose below a “classical” protocol:

- a) Prepare two tubes with 6 pmol of RNA in 36 μl of water – Incubate for 5 minutes at 85°C.
- b) Add 4 μl of pre-warmed appropriate buffer (DMS, CMCT or SHAPE) and let cool down for 10 min at room temperature. Incubate for an additional 15 min at the probing temperature (30-37°C).

*Note: Reaction Buffers can be adapted to each situation, notably monovalent and divalent cations nature and concentrations may be changed to mimic at the best a physiological situation. MgCl<sub>2</sub> can even be left out to destabilize tertiary structure. However, do not use Tris buffer with DMS because reacts with amine groups – Use potassium borate pH8 with CMCT and increase the buffer (any) concentration when using BzCN (SHAPE reagent) to compensate for the reagent acidification*

## 2) RNA modifications:

### a) With DMS:

- Add 4 µl of DMS (or 4 µl ethanol for the mock reaction) – Incubate 5 minutes at 37°C.
- Stop the reaction by addition of 400 mM of Tris pH7.5 (final) and immediately place on ice.

Or

### b) With CMCT:

- Add 10 µl of CMCT (or 10 µl of H<sub>2</sub>O for the mock reaction) - incubate for 10 minutes at 37°C.
- Stop the reaction by adding 1 µl of tRNA (10 µg/µl). Put on ice.

Or

### c) With SHAPE Reagents:

- Add 5 µl of SHAPE reagent 10X (or 5 µl DMSO for the mock reaction).

*Note: In order to rapidly mix the solution, be careful to add the RNA solution onto the SHAPE reagent rather than the opposite*

- Incubate for five reagent half-lives at 37°C (30 minutes for NMIA, 2 minutes for 1M7 or few seconds for BzCN) after which the shape compound is hydrolysed (no inactivation step required).

*Note: The solution will turn yellowish upon reaction with 1M7 or cloudy with BzCN.*

*Note 1: All these reagents are genotoxic and/or skin and/or eye irritant. They should be handled with care under a fume hood wearing nitrile gloves and safety glasses. Waste tips and dispose soiled materials in a saturated sodium hydroxide solution, or a 10% acetic acid solution for DMS and CMCT respectively. SHAPE reagents get hydrolysed relatively rapidly in water.*

*Note 2: As the modification of one nucleotide may influence the rest of the structure, multiple modifications per RNA molecule should be avoided. As a rule of thumb we consider that such conditions are met when the amount of the full length product is at least 70% of that of that mock reaction. The modification conditions described above are pretty standard and work well for RNA between 200 and 500 nucleotides. They can be adjusted by modifying the incubation time and reagent concentration to get more or less modification (NB: concentration or time should be increased for shorter RNAs). This applies for DMS and CMCT, but not for SHAPE reagents which do not produce more than one modification per RNA under the above described conditions.*

## 3) Ethanol precipitation

- Add 1 µl of glycogen (20 mg/ml) or tRNA (10 mg/ml), 10% of ammonium acetate 5 M, 400 µl of 100% cold ethanol and incubate for at least 30 minutes at -20°C.
- Centrifuge for 30 minutes at 16 000 g at 4°C. Discard the supernatant.
- Wash with 400 µl of 70% cold ethanol.
- Centrifuge for 5 minutes at 16 000 g at 4°C. Discard the supernatant - repeat this step.
- Air dry the pellet and resuspend it in 10 µl of water.

**STOP POINT:** probed RNA can be stored at -20 °C for at least 6 months

## 4) Reverse transcription

- Transfer the resuspended RNAs to PCR tubes and prepare two PCR tubes with 6 pmol of RNA in 10 µl of water for the sequencing reactions.
- Add 1 µl of DMSO to the samples (modified RNA, mock reaction and sequencing reactions), denature for 3 minutes at 95°C then flash cool on ice.
- Add 3 µl of 2 µM D2-labelled primer to the sequencing reaction and 3 µl of D4-labelled primer to the modified / mock RNA. Incubate for 5 minutes at 65°C then anneal for 10 minutes at 35°C and store on ice.
- Add 4 µl of 5X RT buffer, 1 µl of the 10 mM deoxynucleotides mix and 1 µl of MMLV (H-) reverse transcriptase. For the sequencing reactions, also add 1 µl of 10 mM dideoxynucleotide. Incubate for 2 minutes at 35°C, 30 minutes at 42°C and 5 minutes at 55°C then store on ice.

*Note: For the sequencing reaction pick a nucleotide which is well represented all along the RNA sequence*

- Add 5 µl of 5X RT Stop buffer.

- f) Combine one elongation from the modified RNA or the mock RNA with one sequencing reaction
- g) Add 250% (v/v) of 100% ethanol, incubate for at least 30 minutes at -20°C and precipitate the DNA as in 3.b-d).
- h) Air dry the pellet and resuspend in 40 µl of sample loading solution.

*Note: Resuspension will take time at room temperature. It can be accelerated by incubating the samples for 10 minutes at 50°C.*

**STOP POINT:** Elongated cDNA can be stored at -20 °C

## 5) Sequencing by capillary electrophoresis

- a) Load side-by-side 20 µl of the elongated cDNA corresponding to the modified RNA and to the mock RNA.
- b) Run the capillary electrophoresis.
- c) Export the raw data as text files.

## 6) Data analysis with QuSHAPE

Different software such as QuSHAPE (Karabiber et al., 2013), RiboCAT (Cantara et al., 2017), CAFA (Mitra et al., 2008), FAST (Pang et al., 2011) and HiTRACE (Kim et al., 2013; Yoon et al., 2011) have been developed to analyse electropherogram.

We mostly use QuSHAPE (Karabiber et al., 2013), please note that the experimental set-up might be slightly differ when using another analysis software.

QuSHAPE can treat any type of chemical probing experiment after capillary electrophoresis separation of the RT products. Output files from ABI sequencer can be directly fed to QuSHAPE while those from CEQ8000 (AB Sciex) contains a header that needs to be removed beforehand (open the file with Excel and delete the header). For a description of QuSHAPE and its use please see (Karabiber et al., 2013) and the “help” of the software.

As a result, QuSHAPE exports a tab delimited text file that contains the areas of the (+) ('RX') and (-) ('BG') reactions and their difference ("RX-BG") for each nucleotide, row-by-row. QuSHAPE also offers the possibility to normalize the results, however we have developed our own script for normalization including some specific treatment for DMS and CMCT, it also allows for the averaging of the reactivities issued from replicates (see next chapter). If using directly QuSHAPE output format the result file as described in “chapter XXX”.

## 7) Normalization and averaging of the replicates

QuSHAPE data analysis yields raw reactivity values ("RX-BG") corresponding to the differences of the cognate area for the (+) and the (-) reactions for each nucleotide. Values corresponding to intrinsic reverse transcriptase stops need to be eliminated, the data are then normalized to allow the comparison between different experiments. To this end, two different approaches have been proposed in (Low and Weeks, 2010a), and should be used depending on the RNA length.

- a) Quantile normalization (RNA shorter than 300 bps)

The 2% of the most reactive peaks of all reactivities (modified and unmodified condition) are removed, they empirically correspond to intrinsic stops (outlier). Then, the next 8 % of the most reactive peaks are averaged ( $M_8$ ) and finally all reactivities are normalized by the average ( $M_8$ ).

**Or**

- b) Parametric outlier removal (RNAs beyond 300 bps)

In this approach, 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 highest remaining reactivities (outliers removed) are averaged. All reactivities (outliers included) are normalized by this average value.

Typically, the probing of an RNA in a given condition is performed in triplicate before further analysis.

To ease the normalization process as described above (Low and Weeks, 2010a), we have developed a python script (version 2.7) that allows the normalization, and yields for each nucleotide the average value of the replicates, and the standard deviation (The “Macro\_CE” archive is available at [https://github.com/afafbioinfo/Macro\\_CE](https://github.com/afafbioinfo/Macro_CE) ).

The “SHAPE\_Treat” folder contains:

- **The Qu-shape-output folder:** in which all the QuSHAPE.txt output files to be analysed should be put

*Note: The file title should be as follow: RNA\_type-of-probe\_replicate-number.txt (ex: TtLSU\_DMS\_2.txt)*

- **The CE\_reactivity\_profile folder:** contains the resulting files (RNA\_type\_of\_probe\_number\_of\_triplicat.txt). This file contains three columns: nucleotide number; nucleotide type and normalized reactivity.
- **The Reactivity folder:** contains normalized reactivities in a "RNA\_probing\_reagent.shape" file. Such file contains two columns: the first is the nucleotide number while the second is the corresponding SHAPE mean normalized value.
- **Macro\_Qushape.py:** python script
- **Macro\_Qushape.config:** This file that can be edited with any text editor, it contains all the parameters to be modified (nucleotides, lowervalue, threshold and Desactiv\_threshold) for the normalization process.

The script can be customized for i) Parameters considered for the normalization and ii) Parameters considered for the final calculation of the reactivity

i) Parameters for normalization:

- **Nucleotides:** SHAPE reagents hit any type of nucleotides, but reagents such as DMS and CMCT hit only two types of nucleotides. We therefore define the outlier on all nucleotides, but normalise the values considering the targeted nucleotides only. To configure this, enter the nature of the nucleotides targeted by the chemical agent (A,C,G,U for SHAPE reagents, A,C for DMS; G,U for CMCT).

*Note: Do not use any space after commas.*

- **Lowervalue:** Reactivities are not supposed to be negative, notwithstanding, it happens that even after the scale factor correction some values are still negative. The negative threshold for which the reactivity will be set to zero can be configured. Default value = -0.1

ii) Parameters considered in the final analysis:

The averaged reactivity for each nucleotide is calculated from the replicates.

Rules for the analysis:

- If at least two reactivities are equal to -10 (outlier), the average is set to -10.
- If there is only one -10 in 3 sets of experiments, the mean is calculated with the two remaining reactivities.
- **Threshold:** This parameter has been added to characterize nucleotides which reactivity is variable and ensure the exclusive use of solid data. It is a maximal difference to the mean accepted for one of the replicate. Default value= 0.2.  
if the difference<0.2, for all the replicates, the average is calculated, in contrast if the difference >0.2, in one of the replicates, the reactivity for the nucleotide is set to -20.
- **Desactiv\_threshold:** It is the highest value for which the "threshold" correction will be applied (Default value= 0.6 ).

Once the parameters set, run the script by typing the following command line from the Qu-shape-output folder:  
> **python2.7 Macro\_Qushape.py**

The script generates two output files: one in the "CE\_reactivity\_profile" folder that lists the normalized values for each reaction whereas in the "Reactivity" folder are found files with the normalized average reactivity for each nucleotide. This file has to be renamed to be fed into IPANEMAP (see below).



## High throughput SHAPE probing: the SHAPE-MaP protocol (Smola et al., 2015b)

Different protocols have been established for high-throughput SHAPE probing, the *SHAPEseq* on premature stops to detect the modified nucleotides (ref lucks), while the SHAPE-MaP rely on the property of the reverse transcriptase to make errors when copying a modified nucleotide in presence of Manganese ions (ref smola). Although coupled to NGS, such techniques do not allow to probe a full transcriptome because each nucleotide should be read 5-20 000 times to obtain a significant reactivity. To overcome this limitation *icSHAPE* selects for modified RNA fragments before the reverse transcription steps, (spitale flynn).

Below is described the SHAPE-MaP protocol that we routinely use, note that it can be readily adapted to DMS (zubradt), or to any other probe.

### **BUFFERS AND SOLUTIONS**

- RNA folding buffer (3X): 300 mM HEPES pH 8; 300 mM KOAc and 30 mM MgCl<sub>2</sub>.
- RNA denaturing buffer (10X): 500 mM HEPES and 40 mM EDTA.
- SHAPE-MaP pre-buffer (10X): 500 mM Tris-HCl pH8, 750 mM KCl and 100 mM DTT.
- Anhydrous DMSO.
- SHAPE-reagent : 1M7 (CAS # 4693-01-0) 100 mM or NMIA (CAS# 10328-92-4) 100 mM or BzCN mM (CAS # 613-90-1) dissolved at 100 mM in anhydrous DMSO.
- MnCl<sub>2</sub> 30 mM.
- Specific primers 2 µM or Random 9-mer primers 200 ng/µl (NEB S1254)
- dNTP stock 5 mM.
- SuperScript III (Invitrogen)
- 70 %, 80 % and 100 % cold ethanol
- NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB E7770)
- Multiplex Oligo (NEB E7335/E7500/E7710/E7730)
- G50 sephadex (GE Healthcare)

### **MATERIALS**

- Eppendorf 1.5 mL
- PCR tubes (0.2 mL)
- Mini column (Biorad)
- Ampure XP Beads (Beckman)
- Micro-volume spectrophotometer
- Agilent Bioanalyser - Chip and Reagents for DNA analysis (usually available on sequencing platforms)
- Thermocycler

#### **1. RNA modifications with SHAPE reagents**

##### ***Modification condition and negative control***

- a) In PCR tubes add 10 pmol of RNA in 12 µl of water. Incubate for 3 minutes at 95°C, then immediately put on ice for 3 minutes.

**Note:** In this protocol it is particularly important to use DNA-free RNA

- b) Flash spin before adding 6 µl of RNA folding buffer (3X) and incubate for 20 minutes at 37°C.

**Note:** Other denaturation/renaturation protocol can be applied according to the RNA to be modified or any other specificity of the system studied. The Probing buffer can also be adapted (see above)

- c) In separate PCR-tubes add 1 µl of the SHAPE-probe at 100 mM (modification reaction) or 1 µl of DMSO (negative control).
- d) Add 9 µl of RNA folded directly on the SHAPE-probe and mix well by pipetting. Repeat this step for the negative control.
- e) Incubate for 6 minutes at 37°C for 1M7, and about 30 min for NMIA (approximately 5 times the half-life). Store the PCR-tubes on ice.

##### ***Denaturing condition***

- a) In a PCR-tube add 5 pmol of RNA in 3 µl of water. Then, add 1 µl of RNA denaturing buffer (10X) and 50 % of formamide (v/v).

- b) Incubate for 1 minute at 95°C.

**Note:** For RNA with strong secondary structure may be incubated up to 5 minutes at 95°C.

- c) In another tube add 1 µl of SHAPE-probe at 100 mM. Make a short spin, then, quickly add the denatured RNA, mix well by pipetting and incubate for another minute at 95°C. Place the PCR-tube on ice.

### **RNA purification**

Bring the final volume of each sample to 50 µl with water and place the tubes on ice time while preparing the G50 Sephadex mini-column.

- Place a mini column in a 1.5 mL Eppendorf and add 800 µl of G50 slurry. Spin for 30 seconds at 1000g. Discard the “eluate”.
- Complete with G50 slurry and centrifuge for 30 seconds at 1000g. Discard the eluate.
- Repeat the centrifugation step twice. The first for 30 seconds, the second for 1 minute.
- Place the mini-column in a new 1.5 ml Eppendorf. Add RNA on the sephadex, spin for 2 minutes at 1000g. Recover the eluate

**STOP POINT:** RNA probed and denatured can be stored at -20 °C for at least 6 months.

## **2. cDNA synthesis by reverse transcription and purification**

Prepare 5X SHAPE-MaP pre-buffer by combining an equal volume of SHAPE-MaP pre-buffer 10X and of dNTP stock at 5 mM. Then add an equal volume of 30 mM of MnCl<sub>2</sub> to the SHAPE-MaP pre-buffer to obtain 2.5X SHAPE-MaP buffer.

**Note:** It is important to prepare this buffer extemporaneously because, manganese oxidizes over time.

The reverse transcription can be performed with specific or random primers. We recommend specific primers for RNA up to 300-400 nt long.

### **Reverse transcription with specific primers (RNA up to 400 nt)**

- In separate PCR tubes add 10 µl (corresponding to roughly 1 pmol) of the modified/unmodified/denatured RNA. Incubate for 3 minutes at 95°C then place on ice for at least 1 minute.
- Spin briefly, add 1 µl of 2 µM specific primers, incubate for 5 minutes at 65°C and cool down on ice.
- Add 8 µl of 2.5X SHAPE-MaP buffer, mix well by pipetting and incubate for 2 minutes at 50°C.
- Place the tube on ice and add 1 µl of SuperScript III reverse transcriptase. Flick the tube and spin briefly before incubating for 3 hours at 50°C.
- Inactivate the SuperScript III by heating samples at 70°C for 15 minutes. Place the samples on ice.

Or

### **Reverse transcription with random primers**

- In separate PCR tubes add 10 µl modified/unmodified/denatured RNA. Incubate for 3 minutes at 95°C then place on ice for at least 1 minute.
- Spin briefly and add 200 ng/µl of random nonamer primers, incubate for 5 minutes at 65°C and cool down on ice.

**Note:** Nonamers ensure a better sequence coverage than shorter or longer randomers.

- Add 8 µl of 2.5X SHAPE-MaP buffer, mix well by pipetting and incubate for 2 minutes at 25°C.
- Place the tube on ice and add 1 µl of SuperScript III reverse transcriptase. Flick the tube and spin briefly before incubating for 10 minutes at 25°C then for 3 hours at 50°C.

**Note:** It is important to use a reverse transcriptase devoid of RNase H activity to allow for the DNA second strand synthesis.

- Inactivate the SuperScript III by heating samples at 70°C for 15 minutes. Then, place samples on ice.

### **cDNA purification**

Proceed as in “RNA purification” chapter “Medium throughput RNA structure probing with DMS, CMCT, or a SHAPE reagent” part 3, but do not bring the final volume to 50 µl.

**STOP POINT:** cDNA can be stored at -20 °C for at least one year.

The steps described below are designed for Illumina sequencing. If using another type of NG sequencing the library preparation must be adapted. Also note that the ShapeMapper script will also have to be adapted to accept output files other than Illumina’s. At least 3 different barcodes (libraries) need to be use, one for each condition: modified (BC1)/unmodified (BC2) and denatured (BC3). In each library, several RNA can be mixed if



their sequence can be unambiguously distinguished. For RNA with identical or close sequences, use different barcodes for each condition.

### **3. Illumina library preparation and quality control**

#### **Second strand synthesis (orange cap, E7770)**

Prepare the samples on ice and follow the manufacturer instructions.

*Note: To improve the yield of DNA second strand synthesis, incubate for 2 hours instead of 1 hour as indicated in manufacturer's instructions.*

#### **dsDNA purification with Ampure XP beads**

a) Leave the beads on the bench for at least 30 minutes until they reach room temperature.

*Note: This step significantly increased the purification yield.*

b) In a 1.5 ml Eppendorf tube add 144 µl of the beads slurry on the “second strand synthesis” reaction, mix well by pipetting and incubate for 5 minutes at room temperature.

c) Place the 1.5 ml Eppendorf on a magnetic rack and wait for 2 more minutes. Discard the liquid.

d) Wash the beads with 200 µl of cold ethanol (80%).

*Note: Rotate the tube a quarter turn every 30 seconds to mix the beads and improve the wash.*

e) Discard the liquid and repeat the step. Spin briefly to discard all the liquid.

f) Air dry the beads for 5-10 minutes.

*Note: Do not over dry the beads. The beads become mat when they are dry.*

g) Add 53 µl of 0.1X TE, mix well and incubate for 2 minutes on magnetic rack.

h) Remove 50 µl of purified DNA and place it in PCR tubes.

**STOP POINT:** dsDNA can be stored at -20 °C for at least one year.

#### **End-It Repair and Adaptors Ligation (green and red cap, E7770)**

Samples are prepared on ice following the manufacturer instructions.

#### **Ligated DNA purification with Ampure XP beads**

Proceed as for “dsDNA purification” but add 85 µl of beads and elute with 15 µl of 0.1XTE.

**STOP POINT:** dsDNA can be stored at -20 °C for at least one year.

#### **Library enrichment (bleu cap from E7770 and barcode from E7335/E7500/E7710/E7730)**

Follow the manufacturer's indications.

*Note: Typically, 8-10 PCR cycles are enough and avoid the formation of adaptor dimers.*

#### **Library purification with Ampure XP beads**

Proceed as for “dsDNA purification” but add only 45 µl of beads and elute with 25 µl of 0.1X TE.

Quantify the samples by OD<sub>260</sub>.

**STOP POINT:** dsDNA library can be stored at -20 °C for at least one year.

#### **Library verification with a Bioanalyzer (Agilent DNA 1000 Chip)**

Incubate the gel dye-mix at least 30 minutes at room temperature before using.

Follow manufacturer's indications.

*Notes: For a clear image load 5 ng sample in a final volume of 6 µl per well .*

*If in the elution profile you observed a thin peak at H 85 bp it is due to the presence of primers. It is suitable to perform another purification with Ampure XP beads to improve the sequencing quality.*

*If in your elution profile you observed a thin peak at H 120 bp, it is due to the presence of adaptor-dimer. To reduce this peaks dilute your sample ten times with 0.1X TE, verify the library quality on a new chip and use the diluted sample for sequencing.*

### **4. Illumina sequencing**

*Notes: Separate projects or a project including multiple RNA can be loaded on the same Illumina sequencing.*

*If loading separate projects with RNA of the same or similar sequences be careful to use different bar codes for each project (see above). As for the number of projects that can be loaded on the same run keep in mind that each nucleotide should be covered about 20 000 times (the significance threshold being 5000 reads).*

We have also adapted the ShapeMapper script for Ion Torrent sequencing (available at <https://github.com/afafbioinfo/ReadsMap>), but probably due to a higher error rate of this sequencing technique we found the results obtained less accurate.

### 5. Data analysis is carried out using ShapeMapper2 as described in (Busan and Weeks, 2018)

ShapeMapper2 generates four different output files:

- Condition.**map**: contains a table composed of four columns: the first is the sequence position, the second contains SHAPE reactivities with the standard errors in the third column for each nucleotide (fourth column).
- Condition.**shape**: is a tab delimited text file with the position of the nucleotide and the associate SHAPE reactivities. This file can be used as input for IPANEMAP software.

**Notes:** A SHAPE reactivity equal to -999 corresponding to a nucleotide which isn't mapped 5000 times.

- Condition\_**histogram.pdf**: is a graphical summary of the sequencing. Those histograms represents the mutation rate for each condition, the read depth and the reactivities analysis. In a successful experiment the mutation rate in the modified condition is higher than the untreated condition, the majority of nucleotides have read above 5 000 and finally reactivities are positive with a small standard errors.
- Condition\_**profiles.pdf**: contains three graphs. The first graph represent SHAPE reactivity value for each nucleotide. The second graph correspond to the mutation rate for each condition (modified, unmodified and denatured) and for each nucleotide. The final graph shows the coverage for each condition. These graphs allow the experiment validation or are helpful for troubleshooting the problems.

### Predicting structural models from one or several probing experiments using IPANEMAP

Once reactivity profiles have been produced in one or several conditions, IPANEMAP can jointly integrate them and produce one or several secondary structure models that are well-supported by the experimental data.

SHAPE reactivities can be interpreted as follows:

- **R**= -20: Nucleotides with variable values (when multiple values available).
- **R**= -10: Removed outliers.

**Note:** Negative values are not interpreted by IPANEMAP

- Low value  $< \mathbf{R} < 0.4$ : Nucleotides with low reactivity.
- $0.4 < \mathbf{R} < 0.7$ : Nucleotides with medium reactivity.
- $0.7 < \mathbf{R}$ : Nucleotides with high reactivity.

### Installing IPANEMAP.

IPANEMAP can indifferently be installed on recent versions of Mac, Windows or Linux operating systems.

**Dependencies.** IPANEMAP consists of a set of Python 2.7+ scripts, and requires the prior installation, and accessibility from the command line, of the following dependencies:

- 1) Install a recent version of Python 2.7, available from <https://www.python.org/downloads/release/python-2716/>
- 2) Install the scikit-learn machine learning library, version 0.2 for Python 2.7+ at <https://scikit-learn.org/stable/install.html>
- 3) Install the ViennaRNA package, version posterior to 2.0, downloadable from the TBI Vienna at <https://www.tbi.univie.ac.at/RNA/#download>
- 4) Install a recent version of Java: <https://www.java.com/en/download/manual.jsp>

**Note:** Depending on your operating system, the above-mentioned software may be installed through some package manager, or using an environment manager such as Conda.

**Downloading IPANEMAP.** The software can then be downloaded from:

<https://github.com/afafbioinfo/IPANEMAP>

Download a ZIP archive of the most recent version of the software and unzip it in a directory of your choice.

## Setting up the analysis

Throughout the following, we assume that the name of the analyzed RNA is 'myRNA', and that the conditions are named 'Cond1', 'Cond2'...

1. Prepare a FASTA file named 'myRNA.fa' whose first line contains the FASTA header (eg '> myRNA') and the second line consists of the RNA sequence (eg 'CCCAAAGGG');
2. Prepare reactivity files 'myRNACond1.txt', 'myRNACond2.txt'... (for instance 'HIVgRNADMS.txt', 'HIVgRNANMIA.txt ...') Files should be formatted to contain one line per nucleotide, each consisting of: i) the nucleotide number; and ii) the reactivity score, both separated by space/tabulations;
3. Copy/move reactivity files in 'Constraints\_Probing' subfolder of the IPANEMAP folder.

*Note: All names/paths can be changed to any other name (alphanumeric characters only), but such modifications should be reflected in the file names. The README.md file also describes how to set up many additional options.*

## Running the analysis and retrieving the results

- 1) Open a terminal and move to the IPANEMAP installation folder
- 2) Execute the command:  

```
python2.7 IPANEMAP.py --RNA myRNA.fa --cond Cond1 Cond2... > Best.txt
```

*Note: Under some operating systems, "python2.7" may be safely replaced by "python" in the above command*

*Note: Many additional options, including the possibility to specify user-defined hard constraints are available and documented in the 'README.md' file found at the root of the installation.*

**Dominant conformations.** The main results are now located in the file 'Best.txt' featuring one or several dominant secondary structure models, along with additional information. For instance:

Structure	dG	#SupportingConditions
BoltzmannProbability		
(((...)))...(((...)))	-4.3 1	0.546

indicates a single dominant conformer, associated with a secondary structure in dot-bracket notation. The Turner free-energy, number of conditions supporting the conformer, and the accumulated pseudo-Boltzmann probability across conditions, are also reported for each conformer

**Complete conformers list.** The complete list of the main conformations across the pseudo-Boltzmann ensemble can be found in a file 'output/Centroids.fa'. The file is simply a list of sequence/structures, and includes conformers that are dominated, ie either less stable or less supported by experiments than the ones reported in the primary output.

**Visual representation.** Using default options, IPANEMAP produces files 'output/img/Optimal-1.svg', 'output/img/Optimal-2.svg'... and 'output/img/Centroid-1.svg'... Those files are VARNA-rendered drawings of the conformers present in the two both above-described lists. Individual nucleotides are color-coded with the reactivity of the first condition ('Cond1').

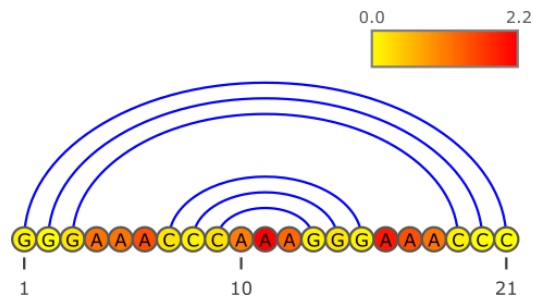


Fig. 1: Typical VARNa-rendered output of IPANEMAP, color-coded using the reactivity of the first condition

**Note:** SVG files can be visualized within any modern browser, and can be converted to a large list of formats using software like Inkscape or Adobe Illustrator.

IPANEMAP workflow was extensively tested using different benchmark RNA and multiple probing conditions (Saaidi et al. Submitted). It appeared from that study that IPANEMAP provide more accurate and stable results than other existing workflow when using only one set of probing conditions. It is safer to use two conditions because if one of them is artifactual the second will correct the problem. However it is important to note that the prediction is never improved compared to the results obtained with the better of the two conditions. Modeling accuracy is improved when using at least three different conditions (Saaidi et al. Submitted).

#### ACKNOWLEDGMENTS:

Research in B.S laboratory is funded by the CNRS, the University of Paris and a grant from “la Fondation pour la Recherche Médicale” (FRM DBI20141423337) awarded to BS and YP laboratories. AS and DA were recipient of a PhD fellowship from FRM (FRM DBI20141423337), GdB was recipient of a fellowship from the French Ministry for Education and Research (MESR).