WIREs RNA

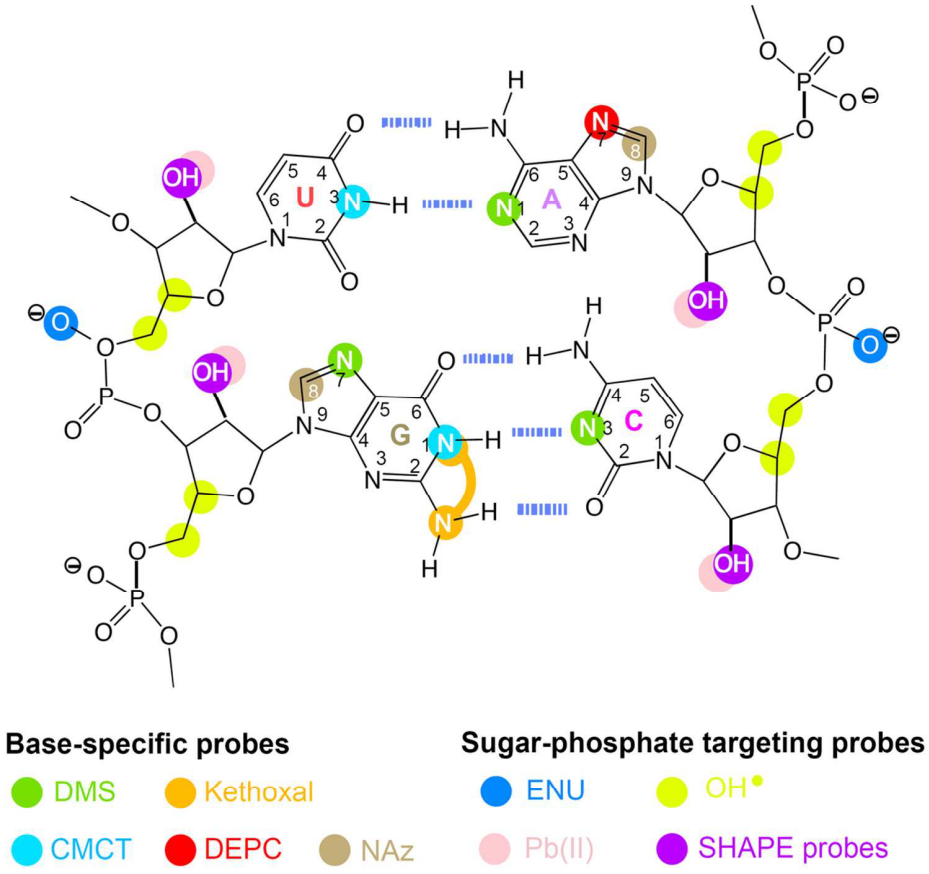
WILEY

The evolution of RNA structural probing methods: from gels to next generation sequencing

Journal:	*WIREs RNA*
Manuscript ID	RNA-851.R1
Wiley - Manuscript type:	Advanced Review
Date Submitted by the Author:	13-Sep-2018
Complete List of Authors:	Mailler, Elodie; CNRS - University of Strasbourg Paillart, Jean-Christophe; CNRS - University of Strasbourg Marquet, Roland; CNRS - University of Strasbourg Smyth, Redmond; CNRS - University of Strasbourg Vivet-Boudou, Valerie; CNRS - University of Strasbourg
Keywords:	RNA, Structure, Probing, NGS
Choose 1-3 topics to categorize your article:	RNA Structure, Dynamics and Chemistry (RAAB) < RNA Structure and Dynamics (RAAA), RNA Analyses In Vitro and In Silico (RKAB) < RNA Methods (RKAA), RNA Analyses in Cells (RKAC) < RNA Methods (RKAA)

SCHOLARONE™
Manuscripts

Review



Graphical abstract: RNA structural probing

RNA structural probing

106x107mm (300 x 300 DPI)

WIREs Wiley Interdisciplinary Reviews

Article Title: The evolution of RNA structural probing methods: from gels to next generation sequencing

By: Elodie Mailler, Jean-Christophe Paillart, Roland Marquet, Redmond P Smyth* and Valerie Vivet-Boudou*

Authors:

[**List each person's full name, ORCID iD, affiliation, email address, and any conflicts of interest.** Copy rows as necessary for additional authors. Please use an asterisk (*) to indicate the corresponding author.]

Elodie Mailler Orcid ID: 0000-0001-9766-0096 Address: Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, IBMC, 15 rue René Descartes, F-67000 Strasbourg, France Current Address: Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, 35A Convent Drive, MSC 3578, Bethesda, MD 20892-3758, USA Email address: elodie.mailler@nih.gov Conflicts of Interest: none
Jean-Christophe Paillart Orcid ID: 0000-0003-1647-8917 Address: Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, IBMC, 15 rue René Descartes, F-67000 Strasbourg, France Email address: jc.paillart@ibmc-cnrs.unistra.fr Conflicts of Interest: None
Roland Marquet Orcid ID: 0000-0002-4209-3976 Address: Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, IBMC, 15 rue René Descartes, F-67000 Strasbourg, France Email address: r.marquet@ibmc-cnrs.unistra.fr Conflicts of Interest: None
Redmond P Smyth* Orcid ID: 0000-0002-1580-0671 Address: Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, IBMC, 15 rue René Descartes, F-67000 Strasbourg, France Current Address: Helmholtz Institute for RNA-based Infection Research (HIRI), 97080 Würzburg, Germany Email address: r.smyth@ibmc-cnrs.unistra.fr Current email address: Redmond.smyth@helmholtz-hiri.de Conflicts of Interest: None
Valerie Vivet-Boudou* Orcid ID: 0000-0001-8702-1047

For Peer Review

Address: Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, IBMC, 15 rue René Descartes, F-67000 Strasbourg, France
Email address: v.vivet@ibmc-cnrs.unistra.fr
Conflicts of Interest: None

ABSTRACT
RNA molecules are important players in all domains of life and the study of the relationship between their multiple flexible states and the associated biological roles has increased in recent years. For several decades, chemical and enzymatic structural probing experiments have been used to determine RNA structure. During this time, there has been a steady improvement in probing reagents and experimental methods, and today the structural biologist community has a large range of tools at its disposal to probe the secondary structure of RNAs in vitro and in cells. Early experiments used radioactive labelling and polyacrylamide gel electrophoresis as read-out methods. This was superseded by capillary electrophoresis, and more recently by next generation sequencing. Today, powerful structural probing methods can characterize RNA structure on a genome wide scale. In this review, we will provide an overview of RNA structural probing methodologies from a historical and technical perspective.

For Peer Review

INTRODUCTION

RNA is a dynamic molecule that can fold into a variety of structures depending on alternative intramolecular base-pairings that are highly dependent on environmental factors (such as presence of biological macromolecules, chemical modifications to RNA, pH, osmolarity or temperature). RNA structure is organized hierarchically, with multiple structural motifs such as hairpins, mismatches, bulges, internal loops, pseudoknots or junctions (**Figure 1**) that together build the tertiary structure (Cruz & Westhof, 2009). Among the functions of RNA structures, the enzymatic activity of ribozymes is probably the most known. These RNAs adopt specific well-folded structures to catalyze self- or trans- RNA cleavage as well as splicing. Riboswitches are another class of RNA with function related to their structure. Upon fixation of a ligand or in response to environmental conditions, they affect gene expression by sequestering or releasing RNA motifs (Serganov & Patel, 2007). But it is now admitted that structures in mRNAs or long non-coding RNAs also play important regulatory roles. The awareness that RNA plays a key role in gene regulation (Morris & Mattick, 2014) has led to an explosion of interest in studying RNA secondary structure and its higher-order tertiary conformation (Mortimer, Kidwell, & Doudna, 2014).

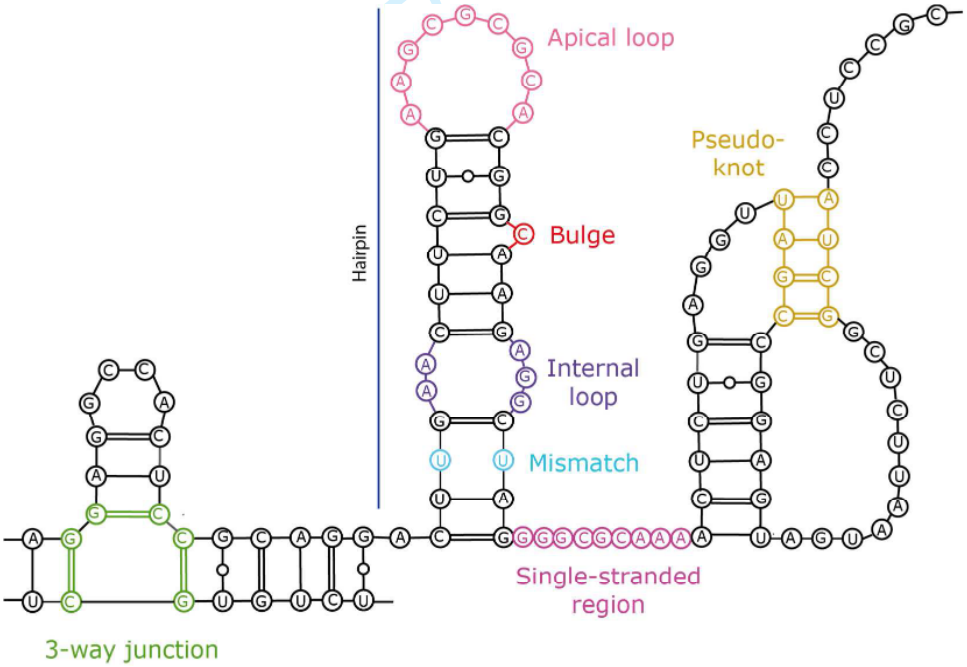


Figure 1: **Schematic representation of RNA structural motifs.** The main folding motifs are highlighted: three-way junction (green), mismatch (light blue), internal loop (purple), bulge (red), apical loop (pink), single-stranded region (dark pink) and pseudoknot (gold).

RNA secondary structure can be predicted *in silico* using phylogenetic approaches where the alignment of a large number of orthologous RNA sequences allows the prediction of base pairs and structure of large RNAs (Gutell, 1993; Juan & Wilson, 1999; Michel et al., 2000; Michel & Westhof, 1990; Schroeder, 2009; Tahi, Tran, & Boucheham, 2017). When sufficient phylogenetic data are unavailable, the thermodynamically most stable RNA structure can be *de novo* predicted using *in*

silico free-energy minimization approaches (Lorenz et al., 2011; Reuter & Mathews, 2010; Zuker, 2003). These *in silico* approaches can even explicitly model the kinetics of RNA folding during transcription (Anderson et al., 2013; Proctor & Meyer, 2013; Zhao, Zhang, & Chen, 2011). Whilst *in silico* predictions can be quite accurate for small and stable RNA domains (Miao et al., 2017, 2015), their accuracy rapidly decreases with increasing RNA length, meaning that the structures of biologically important RNA molecules, such as long non-coding RNAs (Somarowthu et al., 2015) or viral genomes (Mauger et al., 2015; Watts et al., 2009) may be poorly predicted without experimental data. Biophysical techniques such as Nuclear Magnetic Resonance (NMR) (Keane et al., 2015; Kharytonchyk et al., 2016; Lu et al., 2011), X-ray crystallography (Ennifar et al., 2001; Masquida, Sauter, & Westhof, 1999), small-angle X-ray scattering (SAXS) (Doniach & Lipfert, 2009; Fang et al., 2015; Jones et al., 2014) and more recently cryo-electron microscopy (Amunts et al., 2014; Greber et al., 2014; Hang et al., 2015; Natchiar et al., 2017; Nguyen et al., 2015) can be used to study RNA structures. Some of these biophysical methods can reach atomic (or near atomic) resolution but come with significant drawbacks. NMR can only be applied to the study of rather small RNA motifs (up to 100 nts) and is highly restricted by buffer composition, so it cannot be performed under physiological conditions. X-ray crystallography can determine at atomic level the structure of RNAs, but has difficulty resolving the structure of non-compact RNAs or unstructured and flexible RNA regions. Cryo-electron microscopy is a promising and rapidly evolving structural approach. It has seen some success with large RNA domains, but is still limited to large structures and, most often, moderate resolution. SAXS can be used with relatively large molecules, but only provides low information on global structure. Furthermore, the aforementioned techniques cannot assess the influence of the cellular environment onto the RNA structure (Tyrrell et al., 2013). **Table 1** summarizes the main advantages and drawbacks of these techniques.

	Advantage	**Drawback**
Phylogeny	Large RNAs Highly reliable	Large number of orthologous sequences needed
***In Silico* de novo prediction**	No need of extensive sequence information	Accurate only for small and stable RNA domains Only 70% of known base-pairs predicted in the lowest free-energy structure
NMR	Atomic resolution	Only for small RNA motifs Restrictions on the buffer composition Influence of environment cannot be assessed
X-ray	Atomic resolution	Difficult to resolve non-compact, unstructured and flexible regions Influence of environment cannot be assessed
Cryo-EM	Large RNAs	Still moderate resolution Limited to large structures Influence of environment cannot be assessed
SAXS	Large RNAs	Low information on global structure Influence of environment cannot be assessed

Table 1: **Main advantages and drawbacks of *in silico* and biophysical techniques used to study RNA structures.**

RNA structural probing is an experimental approach that can be used to resolve the secondary structure of a wide variety of RNAs with high accuracy and in different environments (Burlacu et al., 2017; Chea & Jones, 2018; Incarnato et al., 2017; Jayaraman & Kenyon, 2018; Rausch,

For Peer Review

Sztuba-Solinska, & Grice, 2017; Watters et al., 2018). The basis of this class of techniques is the use of enzymatic or chemical structural probes that react differently with single-stranded or double-stranded RNAs. The first generation of probing methods was developed 35-40 years ago (Ehresmann et al., 1987) but remained restricted to *in vitro* and *in vivo* analysis of short regions of RNA, as it used gel electrophoresis to resolve the sites interrogated by the structural probes. A first improvement was the replacement of gel electrophoresis with capillary electrophoresis using fluorescently labeled nucleic acids (Wilkinson et al., 2008). This allowed the analysis of longer RNA regions and facilitated the recovery of quantitative data. A few years ago, further progress was achieved through the coupling of probing methods and high-throughput sequencing technology. These "high-throughput" approaches are powerful enough to study RNA structure genome-wide and to characterize thousands of RNA structures in a single experiment (Ding et al., 2014; Rouskin et al., 2014; Underwood et al., 2010; Wan et al., 2012). Since RNA structure modulates RNA processing, localization, translation, and decay, these next generation tools allow biologists to address new questions concerning RNA structure / function relationships that were unthinkable 10 years ago.

In this review, we will first address the classical experimental methodologies for RNA structure determination along with a brief presentation of the different structural probes. We will then discuss technical aspects of how experimental data is obtained. Finally, we will discuss sources of potential bias, as well as challenges associated with the computational analysis of RNA structural probing data.

RNA PROBING REAGENTS

Enzymatic and chemical probes are used to assess RNA structure. Nucleotide accessibility or reactivity to these probes indicates their propensity to be in a paired or unpaired state and can be determined using a variety of readout methods. When designing an RNA structural probing experiment, an important consideration is the choice of the probe(s), as each class of probe has a different reactivity towards RNA. The target positions of enzymatic and chemical probes are illustrated in **Figure 2** and discussed in the following section.

For Peer Review

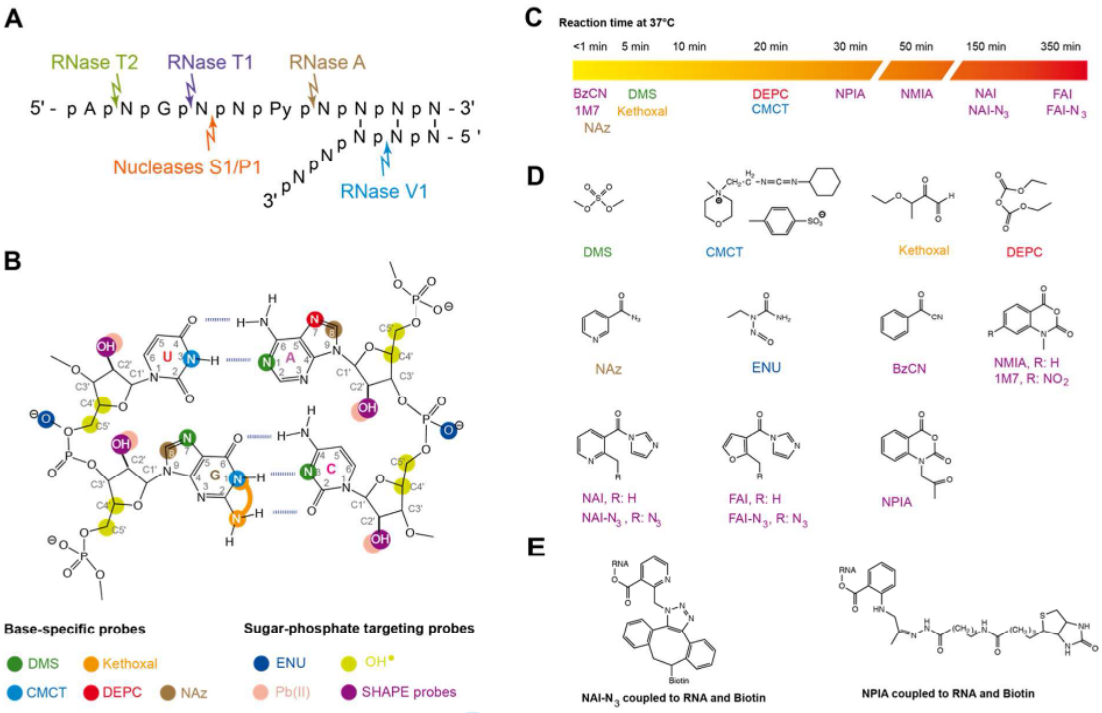


Figure 2: **RNA probing reagents. A.** Common enzymatic probes with their cleavage site on single-stranded (left) or double-stranded (right) RNA. The p indicates the phosphate group (terminal or linkage), A is for Adenosine, G for Guanosine, Py for Pyrimidine nucleosides and N for any nucleoside. The arrow indicates the cleavage site and the nature of the obtained fragment (3' or 5' phosphate). **B.** Target position of chemical probes. Hydroxyl radicals are generally generated by Fenton reagents (H_2O_2, Fe(EDTA) and sodium ascorbate) and Pb(II) is from lead acetate. The hydrogen bonds are drawn to illustrate the Watson-Crick face of the bases but do not indicate that the nucleotides are paired. **C.** Reaction time scale for the main chemical probes reported in the literature. These values are dependent on probe concentration and reaction temperature. For SHAPE probes, the reaction time corresponds to 5x the half-life. **D.** Chemical structure of the probes. **E.** Structure of the chemical probes of the SHAPE family designed to perform a streptavidin selection once they have been coupled to RNA and biotin.

Enzymatic probes

Several nucleases are commonly used to probe RNA in a wide range of physiological buffers with different structural specificities. RNase V1 cleaves double-stranded or structured regions within RNAs without base specificity and generates fragments with a 5'-phosphate. Nucleases S1 and P1 are zinc-dependent endonucleases that cut RNA in single-stranded regions regardless of the identity of the base moiety, leading to 5'-phosphate RNA fragments. RNase T1 and T2 cleave specifically the bond adjacent to the 3'-phosphate of unpaired guanosine (T1) or preferentially the bond adjacent to unpaired adenosine (T2) with formation of a 3'-phosphate RNA fragment. RNase A cleavage also results in the formation of a 3'-phosphate fragment but this enzyme is specific for pyrimidine residues (**Figure 2 A**). Other more exotic nucleases such as *Neurospora crassa* or RNases U2 and Cl3 have also been exploited as structural probes but are less commonly used (Ehresmann et al., 1987).

Due to their large size, enzymatic probes are sensitive to steric hindrance. Another limitation of the enzymatic mapping approach is that enzymes are restricted to *in vitro* applications because of their membrane–impermeable nature (Ehresmann et al., 1987).

Chemical probes

A strength of the chemical probes is their diversity. They have been developed to interrogate the Hoogsteen or Watson-Crick face of the base, as well as the sugar-phosphate backbone. The choice of a specific probe is also driven by the reaction time scale and its ability to enter living cells and quite often several chemical probes are combined to obtain an accurate RNA structure (Cordero et al., 2012; Incarnato et al., 2014; Mandiyan & Boublik, 1990; Paillart et al., 2004). A significant advantage of chemical probes is their small size, meaning that they are less sensitive to steric hindrance than enzymatic probes (**Figure 2 B-D**).

Base-specific chemical probes

Base-specific chemical probes interrogate the Hoogsteen and/or the Watson-Crick faces of the base but have the drawback of having nucleotide specificities. For example, dimethylsulfate (DMS), which methylates N1-A, N3-C and N7-G at neutral pH, is largely used to identify unpaired adenosine and cytosine nucleotides *in vitro* and in a wide variety of organisms (Dutca, Gallagher, & Baserga, 2011; Kwok et al., 2013; Paillart et al., 2004). However, as the introduction of the methyl group is located on the Hoogsteen face of G, an additional step is needed to detect this modification (Peattie & Gilbert, 1980), and nowadays DMS is less used for guanine probing, with the noticeable exception of the detection of G-quartets (Guo & Bartel, 2016). Base specific reagents also include 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metha-p-toluene sulfonate (CMCT), which reacts primarily with N3-U and N1-G of unpaired nucleotides under slightly basic conditions, 2-keto-3-ethoxy-butyraldehyde (kethoxal) forms an additional ring between the primary amine located on C2 and the N1 of unpaired guanines. The modification reaction is reversible but under slightly acidic conditions or in the presence of borate ions the adduct is stable. At neutral pH, diethylpyrocarbonate (DEPC) reacts specifically with N7-A and after treatment with aniline, it allows the detection of adenosine implicated in tertiary interactions (Brunel et al., 1991; Mandiyan & Boublik, 1990). The N7 position of A residues involves in helices are usually not reactive to DEPC, due to stacking, except at the helix ends (Ehresmann et al., 1987). More recently a new chemical compound, nicotinoyl azide (NAz) was developed to address the solvent accessibility of RNA inside and outside the cells (Feng et al., 2018). After light activation, this molecule reacts with purine nucleobases in a method called LASER (Light Activated Structural Examination of RNA) and allows the measurement of structural states of RNA as well as the identification of RNA-protein interactions.

Non-base-specific chemical probes

Other chemicals interrogate the sugar-phosphate backbone, with the advantage that they are not specific to an individual nucleotide. For example, ethylnitrosourea (ENU) is an alkylating reagent specific of the oxygen atoms of phosphate groups engaged neither in tertiary interactions nor in cation coordination (Schlegl et al., 1997; Vlassov, Giege, & Ebel, 1980). The alkylation results in the

For Peer Review

formation of an unstable phosphate tri-ester, which leads to RNA cleavage under mild alkaline treatment. Hydroxyl radicals also cleave the RNA backbone after abstraction of an hydrogen atom from the C4' and/or C5' ribose position(s), providing nucleotide-level information on solvent accessibility of the nucleic acid backbone and tertiary fold of the RNA molecule (Isel et al., 1999; Tullius & Greenbaum, 2005). Until recently, hydroxyl radical probing was limited to *in vitro* applications, since radicals were generated by Fenton reagents: H_2O_2, Fe(II)-EDTA and sodium ascorbate or peroxonitrous acid (Götte et al., 1996). But this probe has been adapted to effective footprinting of RNA-protein complexes *in vivo* whereby a synchrotron X-ray beam generates hydroxyl radicals inside the cell with a 100 ms exposure (Adilakshmi, Lease, & Woodson, 2006; Adilakshmi, Soper, & Woodson, 2009). Whilst not strictly speaking a chemical probing technique, in-line probing can be used to analyse RNA structure by taking advantage of the natural ability of each nucleotide to form a 2'-3'-cyclic phosphate product leading to the cleavage of the ribose-phosphate chain. This event is mediated by the attack of the sugar 2'-OH on the adjacent phosphate di-ester (Regulski & Breaker, 2008) and occurs more frequently when the two groups are "in-line", that is when the nucleotides are flexible. However, in-line probing is not adapted to study fast RNA structure transitions since the technique takes several hours to perform cleavages in mild conditions. Therefore, metal-induced cleavages were developed to cleave flexible regions of RNA using the same mechanism, but at the minute scale (Forconi & Herschlag, 2009). The most widely used metal probe is the lead (II) ion which activates the 2'-OH of nucleotides located in single-stranded regions such as bulges and loops but also in paired regions destabilized by non-canonical interactions or distortions (Brunel et al., 1991; Ciesiołka et al., 1998).

SHAPE (Selective 2'-Hydroxyl Acylation analysed by Primer Extension) reagents are a family of RNA structural probes that under slightly basic conditions react with the ribose 2'-OH of flexible (normally unpaired) nucleotides (McGinnis et al., 2012; Merino et al., 2005; Mortimer & Weeks, 2007). Consequently, SHAPE probes are able to interrogate the local dynamic of the four different nucleotides at the same time (Deigan et al., 2009). Moreover, the increasing number of reagents available allows the study of various biological processes on a variety of timescales (Spitale et al., 2014). Indeed, 1-methyl-7-nitroisatoic anhydride (1M7) (Gherghe et al., 2008) and benzoyl cyanide (BzCN) (Mortimer & Weeks, 2007) quickly react with RNA and are thus well suited for the study of dynamic RNAs. On the other hand, N-methylisatoic anhydride (NMIA) (Gherghe et al., 2008), 2-methylnicotinic acid imidazolide (NAI) and 2-methyl-3-furoic acid imidazolide (FAI) (Spitale et al., 2013) are less sensitive to 2'-OH attack and react in a time window compatible with *in vivo* study of RNA structures (Spitale et al., 2013; Tyrrell et al., 2013) (**Figure 2 C-D**). The reagent half-life is an important experimental parameter: fast reagents are less sensitive to specific ion and buffer choices, and are thought to better reflect steady state RNA-structures (Mortimer & Weeks, 2007; Smola, Calabrese, & Weeks, 2015). However, fast reagents may not be suitable for RNA structural probing in cells, with one group reporting that 1M7 is unsuitable for in cell RNA structural probing due to its reaction rate and poor aqueous solubility (Lee et al., 2017).

CONVENTIONAL REACTIVITY READOUT METHODS

Sites of chemical adduct formation or strand scission can be detected in several ways **(Figure 3)**. When the probing methods involve enzymatic or chemical strand scission, RNAs end-labelled at their 5' or 3' extremity can be used. In this case, the population of cleaved RNA fragments is sized by

denaturing polyacrylamide gel electrophoresis allowing the identification of the position of cleavages on the RNA molecule by comparison with a sequencing ladder. However, this method can only be used to analyse short RNAs (<300 nucleotides) near their extremities (Ehresmann et al., 1987) (**Figure 3 A**). For probing the structure of large RNAs, primer extension by reverse transcriptase is a preferred alternative (Qu, Michot, & Bachellerie, 1983). Moreover, some chemical modifications can only be detected by primer extension, as cleavage at the modified site is not possible (Baudin et al., 1989); this is for instance the case for the SHAPE family of probes. Usually the modified RNA is hybridized with a 5'-terminal [^{32}P]-labelled oligodeoxyribonucleotide complementary to a chosen sequence in the RNA. Since the reverse transcription reaction is blocked when the enzyme encounters an RNA-chemical adduct or a strand scission, a population of truncated radiolabelled cDNAs whose 3'-end corresponds to the nucleotide before the site of modification or to the nucleotide of the cleavage site is generated. This population is traditionally resolved by denaturing polyacrylamide gel electrophoresis and detection is done by either autoradiography or phosphorimaging **(Figure 3 B)**. Reactivities are mapped back to the RNA nucleotide sequence by reference to dideoxyribonucleotide sequencing reactions performed in parallel. However, cDNA synthesis is not only blocked by chemical adducts or cleavage sites but also occurs at sites with a stable RNA structure, natural modification (such as m^2G or m^6A) or due to RNA degradation. Thus, a negative control reaction must be conducted in parallel to the experiment. Band intensity of nucleotide fragments resolved on gel can be quantified by the semi-automated footprinting analysis (SAFA) program (Das et al., 2005). Data from the probed samples are then normalized to control data to estimate the reactivity to the structural probe. However, experimental readouts using radioactively labelled RT primers are labor-intensive and hazardous, and the resolution of polyacrylamide gels limits useful analysis to 150 nucleotides.

For Peer Review