

Probing Techniques of Secondary and Tertiary RNA Structure and a Case Study for RNA G-Quadruplexes



Johanna Mattay

Contents

1	Introduction	161
2	Secondary and Tertiary RNA Structure	161
2.1	Hierarchical Folding of RNA	161
2.2	Examples of Tertiary Structure Motifs	163
3	Techniques to Study RNA Secondary Structure	164
3.1	RNA Structure Prediction in Silico	164
3.2	Enzymatic Probing Techniques	164
3.3	Chemical Probing Techniques	166
4	Probing Techniques to Study the RNA G-Quadruplex Motif	173
4.1	In Silico Prediction of G4s	173
4.2	Visualisation of RNA G4s by Immunolabelling	174
4.3	Chemical Probing of RNA G4s Coupled to Sequencing	174
4.4	Immunoprecipitation of RNA G4s	176
4.5	Visualisation of RNA G4s with Fluorescent Probes	176
4.6	Disruption of G4 Structures with Antisense Oligonucleotides	177
5	Conclusion	177
	References	178

Abstract Like proteins, an RNA is only functional when it is folded into its native conformation and adopts a specific secondary and tertiary structure. Hence, the analysis of RNA structure is essential to understand the cellular roles of distinct RNA molecules. Technical approaches used to study RNA structure comprise bioinformatics tools, structural probing, and biophysical methods to integrate sequence and 3D structure information. In this review, I focus on structural probing techniques of RNA secondary structure. I discuss basic enzymatic and chemical probing techniques, and present novel approaches in combination with high-throughput sequencing. A focus is laid on SHAPE techniques and its various developments and applications. Finally, at the example of RNA G-quadruplexes, it is highlighted how an array of probing techniques can be combined to study a specific RNA structural motif in vitro and in vivo.

J. Mattay (✉)

Institute of Biochemistry, University of Münster, Corrensstr. 36, 48149 Münster, Germany
e-mail: johanna.mattay@gmx.de

Keywords RNA secondary structure · Chemical probing · DMS · SHAPE-seq · SHAPE-MaP · Enzymatic probing · G-quadruplex

Abbreviations

mRNA	Messenger RNA
rRNA	Ribosomal RNA
tRNA	Transfer RNA
ncRNA	Noncoding RNA
snRNA	Small nuclear RNA
snoRNA	Small nucleolar RNA
miRNA	MicroRNA
lncRNA	Long noncoding RNA
nt	Nucleotide
G4	G-quadruplex
ATP	Adenosine triphosphate
RBP	RNA-binding protein
bp	Base-pairs
ssRNA	Single-stranded RNA
dsRNA	Double-stranded RNA
DMS	Dimethyl sulfate
CMCT	1-Cyclohexyl-(2-morpholinoethyl) carbodiimide metho- <i>p</i> -toluene sulfonate
kethoxal	2-Keto-3-ethoxy-butylaldehyde
NMIA	<i>N</i> -Methylisatoic anhydride
1M7	1-Methyl-7-nitroisatoic anhydride
1M6	1-Methyl-6-nitroisatoic anhydride
NAI	2-Methylnicotinic acid imidazolide
BzCN	Benzoyl chloride
•OH	Hydroxyl radical
s ² U	2-Thio-uridine
2-deaza-A	2-Deaza-adenosine
UTR	Untranslated region
CDS	Coding sequence
RT	Reverse transcription
cDNA	Complementary DNA
EDTA	Ethylenediaminetetraacetic acid
DBCO	Dibenzocyclooctyne
NGS	Next generation sequencing
m ⁶ A	<i>N</i> ⁶ -Methyladenosine
hnRNP	Heterologous nuclear ribonucleoprotein
carboxyPDS	Carboxypyridostatin
PDS	Pyridostatin

PCR	Polymerase chain reaction
SAXS	Single-angle X-ray scattering
AFM	Atomic force microscopy
NMR	Nuclear magnetic resonance spectroscopy
cryo-EM	Cryogenic electron microscopy
ASO	Antisense oligonucleotide
LNA	Locked nucleic acid
ALS	Amyotrophic lateral sclerosis
3D	3-Dimensional
FRET	Förster resonance energy transfer
smFRET	Single-molecule FRET

1 Introduction

RNA as a central molecule in biology covers functions from posttranscriptional processing over regulation of gene expression to metabolite sensing. Apart from mRNAs, most RNAs are not translated into protein (noncoding RNAs, ncRNAs), including the abundant rRNAs and tRNAs for ribosomal function. In addition, small nuclear and nucleolar RNAs (snRNAs, snoRNAs) mediate RNA processing steps, microRNAs (miRNAs) control RNA turnover, and long noncoding RNAs (lncRNAs) are regulators of RNA function and biogenesis. To perform their diverse functions, RNAs must fold into their native structures in a cellular environment. Hydrogen bonds from base-pairing and π -stacking of the aromatic ring bases define the RNA secondary structure elements. Long-range interactions to the sugar-phosphate backbone and between distant bases are crucial for tertiary structure. Determining RNA structure is challenging: an RNA of a given nucleotide sequence can adopt multiple low-energy states, with the preferred conformation being dependent on protein binding, ionic environment, nucleobase modifications, and other cellular conditions. Thus, the analysis of RNA structure and mechanisms of RNA folding are crucial to understand the fascinating cellular functions and regulation of RNAs.

2 Secondary and Tertiary RNA Structure

2.1 Hierarchical Folding of RNA

Merely four nucleotides and a highly charged negative phosphate backbone make it challenging for RNA to fold into energetically favourable conformations. RNA folding in 3-dimensional space follows a hierarchical order (Brion and Westhof 1997; Westhof et al. 1996). First, short independently stable helices form rapidly

by Watson/Crick base-pairing. Second, these secondary structure elements undergo tertiary interactions and higher order structures.

Secondary structure elements are uniform: typically, an RNA is composed of a set of short A-form helices of max. 10 bp length, in which the majority of nucleotides are comprised (Russell 2008). The stability of each base-pair is dictated only by Watson/Crick hydrogen-bonding and stacking with the directly adjacent bases.

The primary interaction level of an RNA helix is coaxial stacking: two adjacent helices separated by a phosphodiester bridge stack end-to-end on each other to a colinear arrangement (Butcher and Pyle 2011; Walter et al. 1994). In tRNA, for example, the D-stem coaxially stacks with the anticodon stem, and the T-stem chooses the acceptor arm as the stacking partner, forming the cloverleaf structure (Fig. 1a) (Quigley and Rich 1976). The choice of stacking partners is determined by sequence since the two helix end-standing base-pairs stack via their aromatic bases. Stacking partners can be altered by mutations and non-canonical base-pairs (Sutton and Pollack 2015; Walter and Turner 1994; Yesselman et al. 2019).

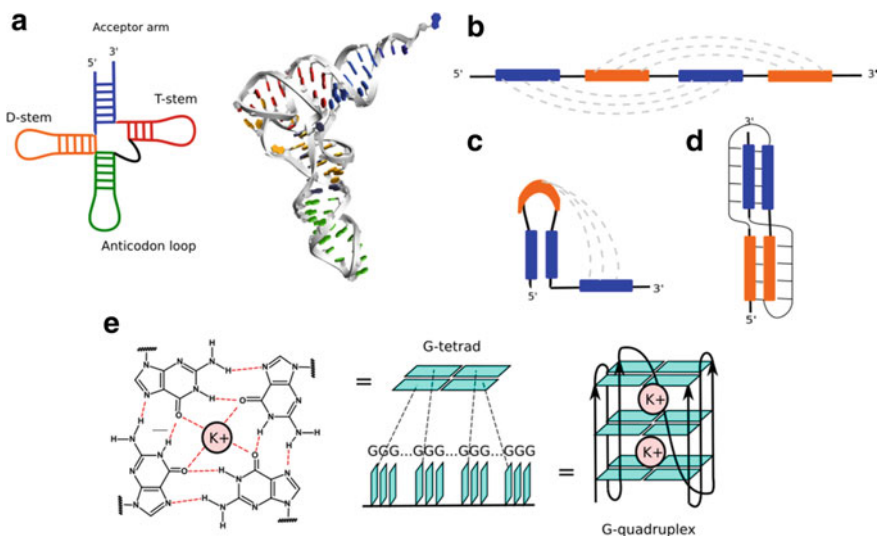


Fig. 1 Common secondary and tertiary structures of RNA. **a** Secondary structure of tRNA. The D-stem/T-stem and the acceptor arm/anticodon loop coaxially stack to form the 3D cloverleaf structure. Shown is the tertiary structure of yeast tRNA^{Phe}. PDB: 1EHZ (Shi and Moore 2000). **b** Long-range tertiary interactions between separate helices. **c** Pseudoknot structure by interaction of a hairpin loop with a ss region. **d** Coaxial stacking of helices, here in the form of two pseudoknot helices. Adapted from (Butcher and Pyle 2011). **e** G-quadruplex. Four guanines from a G-rich strand assemble through Hoogsteen base-pairing. Here, three of these planar G-tetrads stack upon each other to form a G-quadruplex. Shown here is a unimolecular G4 with parallel strand direction. G4s can also form intermolecularly from separate RNA strands or DNA/RNA hybrids. G4s can have antiparallel strand direction, comprise 2–5 stacked G-tetrads, and repeats of G4s can stack with each other to higher order structures

Tertiary interactions form between rigid secondary structures and flexible single-stranded regions (Fig. 1b). Long-range interactions are governed by non-canonical base-pairs such as Hoogsteen, C:A or A:G, electrostatic interactions to the sugar-phosphate backbone, and π -stacking of bases. 29 non-Watson/Crick interactions have been described. In a few cases, tertiary elements can form through Watson/Crick base-pairing and are thermodynamically as stable as secondary structures.

The free energy released to form a short RNA helix can reach 10 kcal/mol, and a GC-rich 10mer duplex can reach a dissociation half-life of 100 years (Turner 1989). Thus, RNA has the problem to become kinetically trapped in stable, but misfolded secondary structure intermediates. Their free energy can vary by only 0.5 kcal/mol from the native structure, as exemplified for tRNA^{Phe} (Jaeger et al. 1989). Due to the multiple loose, transient tertiary interactions, native structures are often not thermodynamically favoured over competing tertiary structures. The RNA folding problem is more serious for long RNAs, which can result in slow folding times up to the minute scale (Weeks 1997).

In vivo, there are two regulatory mechanisms thought to prevent RNAs from misfolding and kinetic traps (Incarnato and Oliviero 2017; Shcherbakova et al. 2008). First, RNA polymerase kinetics, i.e., directionality, velocity, and pausing, guide the order and speed of folding events during transcription (Heilman-Miller and Woodson 2003; Lai et al. 2013; Schroeder et al. 2002). Second, many RNA-binding proteins may act as chaperones to stabilize folding intermediates. They can bind either in a passive way, e.g., hnRNPs like the U1 protein and ribosomal proteins, or actively through ATP hydrolysis, as seen for DEAD-box helicases (Herschlag 1995; Russell 2008; Weeks 1997).

2.2 *Examples of Tertiary Structure Motifs*

Coaxial stacking is the basis of several tertiary motifs, e.g., kissing loops and pseudoknots. Kissing loops form when the loops of two helices base-pair with each other. The L-shape of tRNA results from a kissing loop between the D-stem and the T-stem (Fig. 1a) (Quigley and Rich 1976). In pseudoknots, a loop region of an RNA helix forms Watson/Crick interactions with a single-stranded region outside of this helix (Fig. 1c) (Russell 2008). The A-minor motif is a triple helical structure in which an A interacts via Hoogsteen base-pairing with both nucleotides of a GC base-pair (Butcher and Pyle 2011). It is a building block for tetraloop interactions and kink turns (Keating et al. 2008; Klein et al. 2001). Ribose zippers glue together other motifs by 2' OH hydrogen-bonding between backbone RNA strands (Tamura and Holbrook 2002).

G-quadruplexes (G4s) are stable tertiary structures that assemble from stretches of guanine repeats (Fig. 1e). Four Gs in a four-stranded arrangement assemble to a tetrad through Hoogsteen base-pairing. Two or more of these planar G-tetrads then stack upon each other to a G-quadruplex, which is stabilized in the centre by a K⁺ ion. RNA G4s (G4s) are found in the UTRs of mRNA, in 5' introns of pre-mRNA, in

ncRNAs such as the telomer-associated lncRNA TERRA, and in expansion segments of rRNA (Collie et al. 2010; Kharel et al. 2020; Mestre-Fos et al. 2019, 2020). G4s are most commonly known to act as transcriptional roadblocks in R-loops. However, they cover diverse functions such as the modulation of translation and splicing and the involvement in liquid–liquid phase separation.

3 Techniques to Study RNA Secondary Structure

3.1 RNA Structure Prediction *in Silico*

The nearest-neighbour model finds those base-pairings in an RNA sequence that undergo minimal free energy change (ΔG^0) upon folding (Mathews 2004; Xia et al. 1998). The thermodynamically most stable structure is determined based on hydrogen bonding energies of the base-pair and stacking with the adjacent bases. A second method of structure prediction relies on phylogenetic alignment of orthologous sequences and analysis of covariation sites (Russell 2013). A further development, the maximum expected accuracy, relies only on highly probable (>99%) single- and double-stranded regions and these high-confidence base-pairs are used to assemble the most accurate structure (Lu et al. 2009).

3.2 Enzymatic Probing Techniques

Early mapping techniques exploited endoribonucleases for sequence-specific cleavage of RNAs. The enzymes cut at a specific nucleotide (e.g., RNase T1, RNase A) or are nonselective (RNase I, nuclease S1) (Fig. 1d). While most enzymes prefer single-stranded RNA (ssRNA), RNase V1 targets double-stranded RNA (dsRNA) (Fig. 1d) (Ziehler and Engelke 2001). A drawback is the low resolution: some sites cannot be accessed by the sterically demanding enzymes. Therefore, a combination e.g. of RNase T1 (G), RNase A (C, U), and RNase V1 (dsRNA) gives a detailed secondary structure footprint. Due to the nature of enzyme catalysis, enzymatic probing is not suitable for quantifying the extent of cleavage and thus cannot quantify probed sites. Enzymatic probing coupled to high-throughput sequencing relies on the same principles as those for chemical probing (see 0) (Table 1).

Table 1 Probing reagents and associated high-throughput sequencing techniques for chemical and enzymatic probing of RNAs

Method	Reagent	Application	Specificities	Reference
<i>Chemical probing reagents</i>				
Chemical probing	DMS	In vitro	Unpaired A (N ¹), C (N ³), G (N ⁷)	Waduge et al. (2019)
Chemical probing	CMCT	In vitro/in vivo	Unpaired G (N ¹), U (N ³)	Harris et al. (1995)
Chemical probing	kethoxal	In vitro/In vivo	Unpaired G (N ¹), 2-NH ₂)	Harris et al. (1995)
SHAPE	NMIA	In vitro	Ribose 2' OH (flexible nt)	Merino et al. (2005)
SHAPE	1M7	In vitro/In vivo	Ribose 2' OH (flexible nt)	Mortimer and Weeks (2007)
SHAPE	NAI	In vivo	Ribose 2' OH (flexible nt)	Spitale et al. (2013)
SHAPE	BzCN	In vitro	Ribose 2' OH (flexible nt)	Mortimer and Weeks (2008)
Chemical probing	• OH	In vivo	Solvent-exposed nt	Costa and Monachello (2014)
<i>Sequencing-based chemical probing techniques</i>				
SHAPE-Seq	1M7	In vitro/In vivo	IVT RNA with barcode seq	Lucks et al. (2011)
ChemModSeq	1M7, NAI, DMS	Ex vivo	Probability of RT drop-off rate for each nt Assembly of complexes	Hector et al. (2014)
SHAPE-MaP	1M7	Ex vivo	Introduction of noncomplementary nucleotides by RT at 2' O-adduct sites De novo RNA motif discovery	Siegfried et al. (2014)
MAP-Seq	DMS, CMCT, 1M7	In vitro	IVT RNA with barcode seq	Seetin et al. (2014)
RING-MaP	DMS	In vitro	IVT RNA Mutation frequencies in single transcript to calculate correlation coefficients Through-space interactions Multiple conformations for single RNA	Homan et al. (2014)
Structure-Seq	DMS	In vivo	RNA structure ensembles related to protein function 3nt periodicity in highly translated CDS	Ding et al. (2014)

(continued)

Table 1 (continued)

Method	Reagent	Application	Specificities	Reference
DMS-Seq	DMS	In vivo	Higher number of mRNAs unfolded in vivo than in vitro	Rouskin et al. (2014)
CIRS-Seq	DMS, CMCT	Ex vivo/ deproteinized	mRNA 5'/3' UTRs and lncRNAs highly structured RBP binding site prediction	Incarnato et al. (2014)
icSHAPE	NAI-N ₃	In vivo	Biotin-clickable handle for affinity purification m ⁶ A site prediction	Spitale et al. (2015)
HRF-Seq	• OH	Ex vivo	Determines ribose accessible surface area	Kielpinski and Vinther (2014)
MOHCA-Seq	• OH	In vitro	IVT RNA Mutate-and-map structural modeling	Cheng et al. (2015)
<i>Sequencing-based enzymatic probing techniques</i>				
PARS	RNase V1, nuclease S1	In vitro, ex vivo	Prediction of ss or ds conformation of base Structured CDS in mRNAs Structured AUG sites correlate with low translation	Kertesz et al. (2010)
PARTE	RNase V1	In vitro	RNA folding stability based on melting temp of dsRNA regions ncRNAs have distinct folding energy	Wan et al. (2012)
Frag-Seq	Nuclease P1	In vitro	Novel structures in ncRNAs Cutting score for each site	Underwood et al. (2010)

3.3 Chemical Probing Techniques

3.3.1 Base-Specific Chemical Probing

Chemical probing can assess any RNA region (Incarnato and Oliviero 2017). Nucleotides not engaged in base-pairing or tertiary interactions react with small electrophilic probes and are probed proportional to their accessibility (Chillón and Marcia 2020). It allows quantitative analysis because the number of modification products is directly proportional to the reactivity of the nucleotide. Dimethyl sulfate (DMS) methylates adenosine in N^1 and cytidine in N^3 position as well as guanosine in N^7 (Fig. 2a) (Wells et al. 2000). As the Watson/Crick interface is altered by methylation in A and C, but not in G, DMS probing identifies unpaired A and C in primer extension assays. Complementary, 1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate (CMCT) acylates guanosine in N^1 and uridine in N^3 position (Fig. 2b). A combination of DMS and CMCT is often used to assess the flexibility of all nucleotides in an RNA.

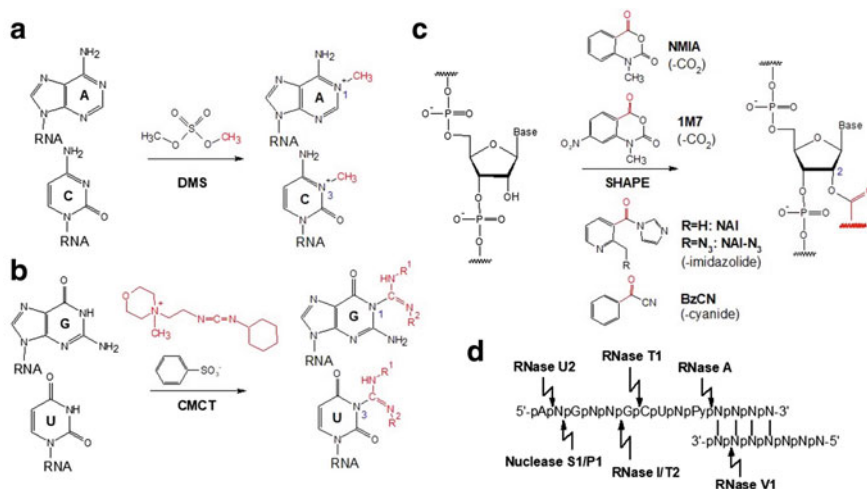


Fig. 2 Structural probing reagents. **a-c** Chemical probing. All reagents probe for non-base-paired nucleotides in ssRNA. **a** DMS methylates adenine and cytosine bases. **b** CMCT reacts with guanine and uracil bases. **c** SHAPE reagents for acylation of flexible 2' OH groups of the ribose backbone: NMIA, 1M7, NAI, BzCN. **d** Enzymatic probing. RNases and nucleases have different cleavage selectivity either at 3' or 5' of the phosphodiester bond, respectively. With the exception of RNase V1 specific for dsRNA, all enzymes cut ssRNA

Chemical probes are used on in vitro-transcribed (IVT) RNA or on selected cell-extracted and purified targets. The cell-permeable DMS allows treatment in vivo in numerous organisms including bacteria, yeast, and mammalian cells (Wells et al. 2000). A structural map of the mouse lncRNA Xist was obtained by treating cells with DMS (Fang et al. 2015a). The in vivo use of CMCT and 2-keto-3-ethoxybutyraldehyde (kethoxal) usually requires prior cell permeabilization (Harris et al. 1995). Kethoxal probes for G by ring formation between the N^1 and the 2-amino group. Of note, novel glyoxal and kethoxal derivatives can enter cells without permeabilization (see 0).

3.3.2 High-Throughput Readout of Chemical Probing

During reverse transcription (RT) of the probed RNA, the introduced modification blocks DNA polymerase from read-through and extension of the cDNA strand, and the enzyme drops off one nucleotide before the reacted site. For enzymatic probing, the cleaved RNA fragments directly result in cDNAs truncated at the site of cleavage due to polymerase run-off. To rule out background termination of RT that can occur on untreated RNA due to secondary structure or natural base modifications (Ziehler and Engelke 2001), controls omitting the probe are compulsory.

In its traditional form, RT was performed with 5'- 32 P-labelled primers, cDNA fragments were separated on sequencing gels and compared to ddNTP-sequencing

standards, and band intensities were quantified by autoradiography (Das et al. 2005). The use of capillary electrophoresis immensely accelerated automation. Fluorescent peaks of the probed substrates are quantified by priming the cDNAs with a fluorescent marker (Mitra et al. 2008; Vasa et al. 2008). Today, next-generation sequencing (NGS) technology allows genome-wide analysis of any RNA as opposed to a handful of *in vitro* targets for traditional readouts.

DMS Probing Coupled to Sequencing

After the pioneering work by Lucks et al. for NGS-based chemical probing (SHAPE-Seq, see 0) (Lucks et al. 2011), the MAP-Seq approach (multiplexed accessibility probing) implemented high-throughput sequencing for DMS probing (Seetin et al. 2014). Alternatively, RING-MaP (mutational profiling) relies on the incorporation of noncomplementary nucleotides during RT. The RT conditions and polymerase are chosen in a way that DNA polymerase reads through the DMS-adducts instead of stopping (see SHAPE-MaP 0). Mutations inserted at the sites of adduct formation are recorded simultaneously on one transcript and are used to analyse interdependencies of DMS-reactive sites and to calculate correlation coefficients (Homan et al. 2014). Thus, transient nucleotide interactions through space can be determined as well as RNA interaction groups (RINGS), which make up the multiple conformations a single RNA can adopt in solution (Table 1).

While the above techniques are limited to IVT RNA or a few purified single targets, Structure-Seq was the first application for genome-wide DMS probing *in vivo* (Ding et al. 2014). Applied to total RNA, it is used to identify RNA structural ensembles that can be associated with general protein functions. Structural characteristics of mRNAs were determined such as a 3 nt-periodicity in codons of highly translated mRNAs, and alternative polyadenylation sites based on high or low structured regions were discerned. The method DMS-Seq probes native RNA structure directly in DMS-treated yeast cells (Rouskin et al. 2014). It revealed a lower number of structured mRNAs in dividing cells. DMS-Seq data from ATP-depleted cells implied that mRNA structuring is restricted by ATP-dependent helicase unwinding steps. To keep mRNAs predominantly unfolded *in vivo* might be advantageous for the cell to provide a uniform structure to mRNAs for ribosome accession and translation. Nevertheless, hundreds of structured domains were also found in mRNAs.

A method that combines DMS and CMCT treatment of isolated RNA deproteinized with Proteinase K (*ex vivo*) has proven useful to infer an unexpected high structuring for lncRNAs. Equally, the 5' and 3' UTRs of mRNAs were found to be highly structured, while low structuring and thus good accessibility was found at ribosome binding sites and stop codons. CIRS-Seq (chemical inference of RNA structure) could also verify the 3 nt-periodicity of mRNA codons and the binding site of the RNA-binding protein Lin28a (Incarnato et al. 2014). Structure-seq2 provides an improved library preparation protocol to improve overall sequencing read coverage and quality using hairpin adapters for decreased ligation bias. Biotinylated dCTPs

during RT are used to replace gel purification steps and remove unwanted ligation by-products (Ritchey et al. 2017).

Glyoxal derivatives have been developed for in vivo probing of RNA targets as alternative to DMS. The molecules successfully entered rice, *Bacillus subtilis*, and *Escherichia coli* cells to modify solvent-exposed G, C, and A residues (Mitchell et al. 2018). Keth-seq employs an azide-modified kethoxal, N₃-kethoxal, to probe G bases on a transcriptome-wide scale (Weng et al. 2020). It entered mouse embryonic stem cells in 1 min and successfully probed their RNA secondary structures.

3.3.3 Non-Base-Specific Chemical Probing

Principles of SHAPE

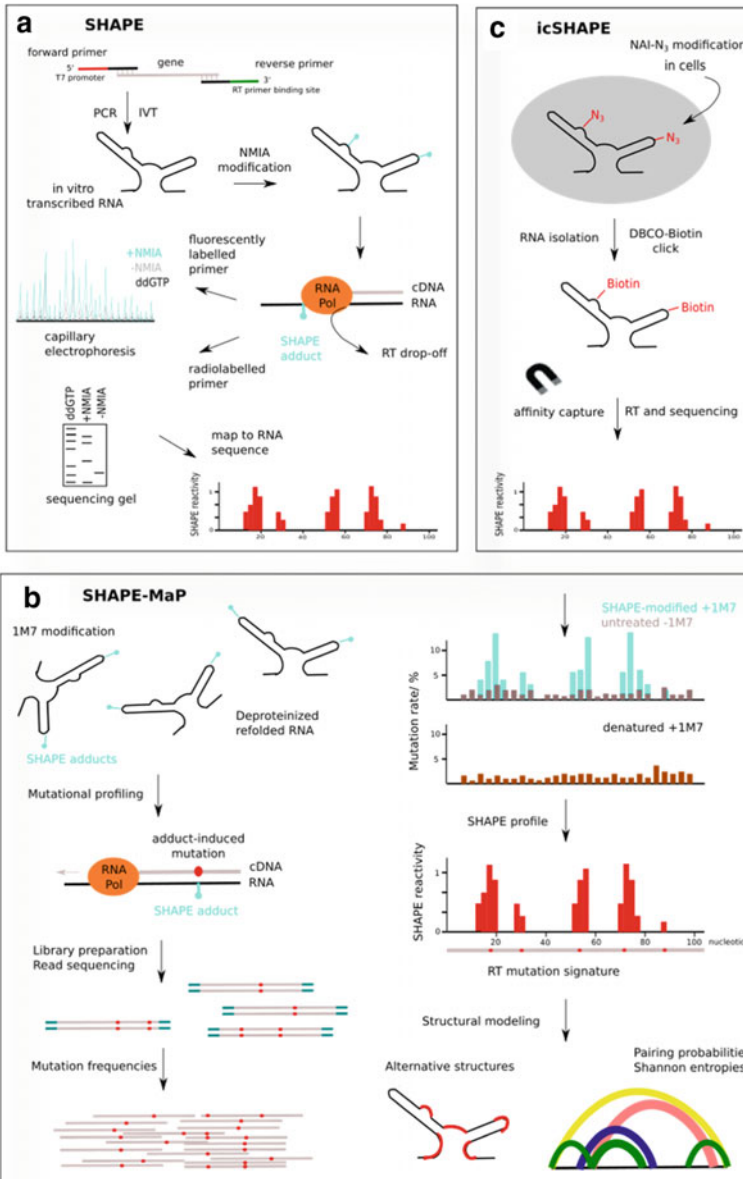
The reactivity of the ribose backbone can be exploited if it is not engaged in secondary or tertiary interactions, e.g., duplexes, Hoogsteen base-pairing or RNA triple helices. The 2' OH group is reacted with an electrophilic probe, *N*-methylisatoic anhydride (NMIA), to form a SHAPE adduct (Fig. 2c). SHAPE, or selective 2'-hydroxyl acylation analysed by primer extension, probes any nucleotide of an RNA. During primer extension, the 2' O-acylation induces the polymerase to fall off one nucleotide before the modified one (Fig. 3a). SHAPE readout is quantitative: the reactivity score (usually 0–1) of each site is directly proportional to local flexibility, i.e., the more SHAPE adducts are formed, the less constrained and more flexible is the nucleotide. Importantly, the reactivity map only corresponds to nucleotide flexibility but not to solvent accessibility (Gherghe et al. 2008; Merino et al. 2005).

Mechanistically, SHAPE reactivity of the 2' OH group is increased by rare ribose C3' or C2' endo conformations and by electronegative and proximal substituents that serve as base catalysts for 2' OH deprotonation and stabilise the tetrahedral transition state (McGinnis et al. 2012). Hence, reactivity is also strongly influenced by RNA modifications. For instance, the substituents in 2-thio-uridine (s²U) and 2-deaza-adenosine (2-deaza-A) decrease nucleophilicity because of electronegativity effects and increased distance to the 2' OH.

SHAPE probing data including folding constraints can be implemented into algorithms to predict RNA secondary structure, e.g., in RNAstructure (Mathews et al. 2004; Rice et al. 2014). Even pseudoknots whose stability factors are poorly understood due to transient tertiary and protein interactions can be predicted (Hajdin et al. 2013).

SHAPE Reagents for in Vitro and in Vivo Use

The most common SHAPE reagent, 1-methyl-7-nitroisatoic anhydride (1M7) (Fig. 2c), is a more electrophilic derivative of NMIA with shorter reaction times (70 s versus 20 min for NMIA) (Mortimer and Weeks 2007). 1M7 can be combined with derivatives of slightly different SHAPE reactivity. For instance, the human lncRNA



MEG3 was probed with NMIA, 1M7 and its regioisomer 1-methyl-6-nitroisatoic anhydride (1M6), and its secondary structure map was confirmed by DMS probing (Uroda et al. 2019). Combinatorial incorporation of SHAPE data from NMIA and 1M6—which detect noncanonical (NMIA) and tertiary (1M6) interactions based on differential kinetics and stacking interactions, respectively—can accurately predict secondary structures of RNAs that are difficult to model (Rice et al. 2014). The

◀**Fig. 3** Structural mapping techniques coupled to massively parallel sequencing analysis. **a** SHAPE: The target RNA is in vitro transcribed from a PCR template. 2' OH acylation with NMIA causes reverse transcriptase to stop at the modified site. The radiolabelled cDNA fragments are quantified and sequenced by gel electrophoresis, or, if a fluorophore-labelled RT primer is used, by capillary electrophoresis. Finally, SHAPE reactivities are plotted back onto the RNA sequence. **b** SHAPE-MaP: 1M7-modified RNA induces mutations that are inserted by a DNA polymerase at sites of 2' O-adducts during RT. Mutation frequencies from sequencing reads are converted to SHAPE reactivities, plotted on an RNA secondary structure map or used for prediction of tertiary structure elements. Adapted from (Siegfried et al. 2014). **c** icSHAPE: RNA is modified in vivo with cell-permeable NAI-N₃. Isolated RNA is then treated in vitro with DBCO-Biotin in a copper-free azide-alkyne click reaction. Biotinylated transcripts are enriched on streptavidin beads for RT and NGS analysis

shotgun (3S) approach has been used for mouse RepA, a repeat element of the lncRNA Xist responsible for X chromosome-silencing in females. Herein, the generated fragments are probed individually by 1M7 SHAPE and their reactivity profiles are compared to that of the full-length transcript (Liu et al. 2017b; Novikova et al. 2013).

1M7 is commonly used for in vivo application in bacteria and eukaryotes, e.g. for 16S rRNA in *E. coli* and MEG3 in human fibroblast cells (McGinnis and Weeks 2014; Tyrrell et al. 2013; Uroda et al. 2019). However, the first SHAPE reagent that was developed exclusively for in vivo use was 2-methylnicotinic acid imidazolide (NAI) (Fig. 2c). NAI circumvents the pitfalls of NMIA of low solubility, cross-reactivity, and short half-life (Spitale et al. 2013). Comparison of in vitro and in vivo SHAPE profiles of 5S rRNA revealed functionally important nucleotides that differ in reactivity due to tertiary or protein interactions in cellular ribosomes.

Another SHAPE reagent, benzoyl cyanide (BzCN), was developed to probe RNA folding dynamics on a timescale of 1–2 s (Fig. 2c). As an example, RNase P forms several tertiary motifs including a tetraloop-receptor motif and a T-loop from A-minor interactions. When RNase P was probed with BzCN in 5 s-intervals during in vitro folding, the kinetics of the folding intermediates and a hierarchical folding pathway could be derived (Mortimer and Weeks 2008, 2009).

In Vivo SHAPE Probing Coupled to Sequencing

Shape-Seq. SHAPE-Seq combines 1M7 probing of an IVT RNA with deep sequencing of the aborted cDNA fragments (Lucks et al. 2011). During RT, a 4 nt-barcode unique for each RNA species is introduced. For a mixture of mutant transcripts, here of RNase P, subtle conformational variations can be analysed. The method is limited as each RNA species of interest has to be generated from a 3'-extended PCR template to contain an RT primer binding site and the barcode template sequence. ChemModSeq combines NGS of random hexamer-primed cDNAs and a novel algorithm for calculating RT drop-off rates and their probabilities to be caused by SHAPE adducts for each nucleotide position (Hector et al. 2014). It is suited

to study RNA conformational dynamics during assembly of complexes and could elucidate structural intermediates of yeast 40S and 80S ribosome biogenesis. Thus, it overcomes obstacles typically encountered in cryo-EM of heterogeneous and unstable particle purification.

SHAPE-MaP. The widely used method SHAPE-MaP is based on the incorporation of noncomplementary nucleotides at the sites of 2' O-modification (Siegfried et al. 2014). In this mutational mapping (MaP), the rate of SHAPE adduct formation is directly converted to mutation frequencies by read counting (Fig. 3b). To obtain a SHAPE-MaP profile with relative SHAPE reactivity for each position, data from the untreated (−1M7) sample is subtracted from data from the treated (+1M7) sample after normalisation to a 1M7-treated denatured RNA control. If the RNA was probed e.g. in presence and absence of a ligand, the conformational changes during ligand coordination can be profiled by calculating the SHAPE difference of + ligand versus −ligand conditions. In addition, calculation of pairing probabilities and Shannon entropies can refine alternatively structured domains or regions with multiple conformations in equilibrium, and even discover RNA motifs de novo. Based on high Shannon entropies, three pseudoknots were predicted in the HIV-1 genome in regions hitherto unknown to contain defined RNA motifs (Siegfried et al. 2014).

Alternative SHAPE protocols. The in vivo click SHAPE method (icSHAPE) uses an azide-containing NAI reagent (NAI−N₃) (Fig. 3c) to click a biotin moiety to the modified nucleotides. The biotin handle is used for affinity capture and enrichment, followed by RT and NGS (Fig. 3c) (Spitale et al. 2015). icSHAPE sequencing data can be used to predict N⁶-methyladenosine (m⁶A) modification sites more accurately than based on the DRACH sequence motif only. In icSHAPE data from cells expressing m⁶A methyltransferase METTL3, the methylated m⁶A sites show higher SHAPE reactivity compared to cells depleted for METTL3. This is because m⁶A disrupts base-pairing in duplex helices and leads to more unstructured regions.

Recently, SmartSHAPE was developed to probe low abundant RNA specimen from primary or immune cells to decrease the input amount of RNA from 1 μg to 1 ng (10⁵ cells) (Piao et al. 2022). As improvements to the original icSHAPE protocol, RNaseI digestion of artifact truncated RNAs improved true positive RT stop signals and on-bead library preparation further increased RNA yield. By profiling the RNA structure landscape of two intestinal macrophage cell lines in mice, it was demonstrated that RNA structural changes directly regulate immune responses.

SHAPE coupled to direct RNA sequencing. Nanopore sequencing has advanced the detection of natural RNA modifications including ribose 2' O-methyl (Nm) and pseudouridine (ψ) by measuring differences in current signal and dwell time between modified RNA and unmodified control of the same sequence. Methods combining SHAPE and long-read direct RNA sequencing have demonstrated the applicability of Nanopore sequencing to detect chemical modifications introduced exogenously. This was demonstrated for modification by the SHAPE reagent 1-acetylimidazole (AcIm) which forms small 2' O-acetyl adducts (NanoSHAPE) (Stephenson et al. 2022), and in NAI-N₃-probed human RNA to phase combinations of structures between isoforms (Aw et al. 2021). Novel model-free algorithms further allow the identification of

similar and conserved RNA structures in different organisms by direct comparison of their SHAPE reactivity profiles (Morandi et al. 2022). This can be helpful in the context of finding druggable and unique RNA targets.

Hydroxyl Radical Probing

RNA is treated with an amount of hydroxyl radicals ($\bullet\text{OH}$) that is equivalent to provoke one cleavage event per molecule on average. The extent of backbone cleavage is then proportional to the solvent-accessible surface of each nucleotide (Mitra et al. 2008; Vasa et al. 2008). $\bullet\text{OH}$ radicals are generated in situ with Fenton reagents such as H_2O_2 and Fe(II)-EDTA or by synchrotron X-ray beams (Götte et al. 1996; Sclavi et al. 1997). In combination with NGS, HRF-Seq (hydroxyl radical footprinting) and MOHCA-Seq (multiplexed $\bullet\text{OH}$ cleavage analysis with paired-end sequencing) allow high-throughput analysis of RNA on a genome-wide scale (Table 1) (Cheng et al. 2015; Kielbinski and Vinther 2014). HRF-Seq of tumour suppressor MEG3 in combination with SHAPE revealed two pseudoknot regions that interact to form a kissing loop motif. This conformational change results in activation of the p53 pathway and cell cycle arrest (Uroda et al. 2019).

4 Probing Techniques to Study the RNA G-Quadruplex Motif

Probing reagents can also be designed to recognize a specific structural motif. Small molecule ligands, antisense oligonucleotides, and antibodies can be applied to modify, isolate, or visualize the structural motif and in certain cases to stabilize or disrupt the secondary structure. In the following, the RNA G-quadruplex structure serves as a model to present how different approaches and probing techniques can be combined to comprehensively study a distinct motif and its biology in cells.

4.1 *In Silico Prediction of G4s*

Prediction of DNA G4 structures from G-rich consensus sequences ($\text{G}_{3+N_{1-7}}\text{G}_{3+N_{1-7}}\text{G}_{3+N_{1-7}}\text{G}_{3+}$) has been performed computationally (Puig Lombardi and Londoño-Vallejo 2020). The presence of 700,000 DNA G4s that were found in the human genome by G4 probing coupled to high-throughput sequencing (Chambers et al. 2015) as compared to only 375,000 predicted loci (Huppert and Balasubramanian 2005) has yet again demonstrated the limitations of in silico prediction. The high false-negative rate is mainly due to non-G sequence variations in the consensus sequence and to regulatory factors in vivo that govern the equilibrium between folded and unfolded states.

4.2 Visualisation of RNA G4s by Immunolabelling

While G4s are well-known secondary structures in DNA, the first evidence for G4 formation in RNA was given by visualization with a G4-specific antibody, BG4 (Biffi et al. 2014). In fixed human cells, incubation with a FLAG-tagged BG4 revealed fluorescent BG4 foci in the cytoplasm, which were indicative of RNA G4 structures. By an increase in cytoplasmic foci, but not in nuclear signal, it was also shown that the RNA G4-specific probe carboxypyridostatin (carboxyPDS, see 0) could exclusively detect cytoplasmic RNA G4s when it was applied to living cells prior to fixation.

4.3 Chemical Probing of RNA G4s Coupled to Sequencing

G4-seq. Shortly after, G4-seq was the first method to map RNA G4s on a transcriptome-wide scale. The method makes use of reverse transcriptase stalling induced by fully folded G4 structures (Fig. 4a). To identify RT read drops, isolated RNA is treated under G4-favourable conditions (K^+ or stabilising ligand, e.g., pyridostatin (PDS), BRACO-19) to allow for G4 folding. This sample is compared to a normalization control obtained under G4-unfavourable conditions (Li^+) (Kwok et al. 2016; Yang et al. 2018). 3300 to 11,000 G4 sites were detected in human mRNAs under physiological K^+ conditions or with PDS, respectively. The majority were found in the 5' and 3' UTRs and were enriched in polyadenylation signals. This is consistent with a role in transcriptional and translational regulation and mRNA processing. Since the technique requires total or polyA-enriched RNA, the binding of proteins or other endogenous ligands is not taken into account. This opens the debate as to whether the identified G4 sites are actually formed physiologically.

G4-DMS-seq. To overcome this problem, G4-DMS-seq adds a DMS treatment step prior to the G4-seq protocol. Since the N^7 positions of the guanines are hidden when constrained in a G-tetrad, stable G4 structures are protected from DMS methylation, and will produce RT stops (Fig. 4c). A reduction in RT stops upon DMS treatment in vivo (+DMS) and a similarity to the in vitro/ $-K^+$ conditions led the authors to conclude that G4s are mainly present in their unfolded state in mammalian cells (Guo and Bartel 2016).

G4-SHAPE. For the same purpose of capturing G4s in their physiologically folded state, the authors developed G4-SHAPE. Herein, NAI was shown to preferentially react with the exposed 2' OH group of a loop-adjacent G in a stable G4 structure (Fig. 4d). From low SHAPE reactivity profiles, the authors concluded that G4s adopt a globally unfolded state in vivo (Guo and Bartel 2016). However, multiple fluorescent imaging studies have now verified the dynamic folding and unfolding of RNA G4 structures in cells (see 0). Of note, SHAPE probing of G4 candidates and RNAfold analysis have proven that alternative stable secondary structures compete with G4 folding. In the same work it was suggested that G4 formation also affects long-range tertiary folding (Kwok et al. 2016).

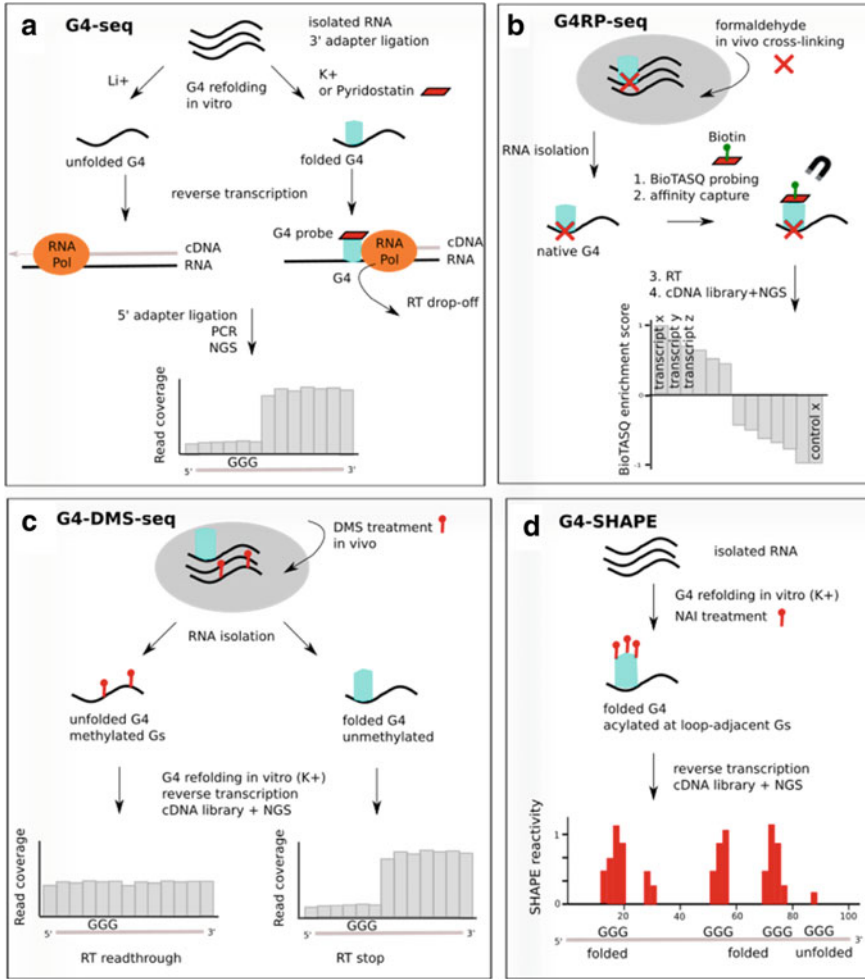


Fig. 4 Probing and visualization techniques for RNA G-quadruplex structures. **a** In G4-seq, total RNA is treated in vitro under G4-inducing conditions or with a stabilizing ligand. G4 sites are identified by a drop-off in RT reads at the site of G4 folding when compared to the untreated control under G4-disfavorable conditions. **b** G4RP-seq applies cross-linking in cells to freeze transiently folded G4s. Incubation in vitro with a biotinylated BioTASQ ligand and affinity capture with streptavidin beads is used to enrich and identify G4-containing transcripts by RT-qPCR or sequencing. **c** G4-DMS-seq makes use of specific methylation of G residues in unfolded G4 structures. RNA from DMS-treated cells is allowed to refold to G4s in vitro only if the G residues were protected from DMS methylation in folded G4 structures. This allows to probe for G4 folding in vivo after RT stop and NGS analysis. **d** G4-SHAPE probes the flexible 2' OH of loop-adjacent G residues in a G4 structure. High SHAPE reactivities indicate folded G4s, whereas low reactivities indicate unfolded G4s in vivo

4.4 Immunoprecipitation of RNA G4s

G4RP-seq. Techniques involving antibodies such as BG4 are equivalent to ChIP-seq experiments, which are commonly used to capture DNA G4 structures in the native chromatin state. iCLIP protocols coupled to RNA sequencing can be used to elucidate binding regions of G4-binding proteins, and thus indirectly assess potential G4 sites (Kharel et al. 2020). More directly, the method G4RP-seq similarly works with crosslinking and affinity purification and with a novel G4-specific probe. The group synthesized a biotinylated probe, BioTASQ, that selectively binds, pulls down, and enriches G4-containing transcripts on streptavidin beads. but transient formation of G4 structures (**Error! Reference source not found.**B) (Yang et al. 2018). Formaldehyde was used to covalently freeze transient G4s in vivo and to minimize ligand-induced stabilization of G4s in the in vitro probing steps with BioTASQ.

In an alternative approach, the intrinsic peroxidase activity of a G4-hemin complex can be exploited in a reaction with H₂O₂ and a biotin substrate to self-biotinylate the G4. Here, the biotinylated DNA G4 was then used for affinity pulldown, purification, and PCR (Einarson and Sen 2017). The self-biotinylation of G4-hemin might also be applied to RNA G4s for RT stop analysis and NGS protocols.

4.5 Visualisation of RNA G4s with Fluorescent Probes

Small molecules that are used for G4 probing can also be applied for G4 visualisation in living cells. RNA G4s can be detected in vivo with turn-on fluorophore probes such as QUMA-1 and Naphtho-TASQ (N-TASQ) (Chen et al. 2018; Laguerre et al. 2015). Treatment with both QUMA-1 and N-TASQ does not require cell fixation and permeabilization as compared to antibody-based fluorescent detection (see 0). QUMA-1 is used for real-time imaging of dynamic folding and unfolding of G4s by tracking the mobility, appearance/disappearance, and merging of fluorescent foci over time. In this way, even the assembly to higher-order G4 structures and the dynamic unfolding of G4s by the helicase DHX36 could be visualized.

To screen for new ligands that are selective for endogenous RNA G4s as opposed to DNA G4s, a click-chemistry approach was developed by Di Antonio et al. (Di Antonio et al. 2012). An alkynylated pyridostatin was incubated with a library of azides containing variable functional groups. In the presence of the G4-forming telomeric-repeat RNA TERRA, adducts between PDS and azide would form only if they were successfully interacting with and stabilized by the G4 structure. By mass spectrometry quantification and competition assays with the DNA G4-forming telomere H-Telo, a carboxy-terminal PDS derivative, carboxyPDS, was validated as a novel RNA-selective small molecule probe.

4.6 *Disruption of G4 Structures with Antisense Oligonucleotides*

G4s are a unique example of a structural motif as individual G4s can be distinguished by sequence identity. Antisense oligonucleotide (ASO) probes have the advantage that they selectively target individual G4s due to sequence-specific base-pairing. In a recent study, DNA probes that disrupt genomic G4s were designed and applied to G4-forming DNA promoters to relieve the secondary structure. The precise positioning of chemically locked nucleosides (LNAs) improved the G4 sequence-binding affinities of the ASOs. The LNA probes led to disruption of DNA G4 structures in a reporter gene promoter. By this, gene expression was activated by facilitating RNA polymerase read-through (Chowdhury et al. 2022).

This example shows that in parallel to examining the sites, quantity, and dynamics of secondary structures, the interference with motif-specific probes is equally important to expand the data on G4s on their biological functions. Here, disruption of individual G4s alleviates polymerase stalling at promoters and allows to study the effect on gene expression. In general, stabilising or destabilising probes can be used as chemical biology tools for switching on or off a motif selectively and to explore its function in cells.

5 Conclusion

The development of next-generation sequencing techniques has paved the way to a high-throughput readout of chemical probing data. Since a decade, SHAPE-seq and DMS-seq have served as models for several variations of probing techniques. Of high importance are the mutational mapping (MaP)-approaches to study several flexible nucleotides simultaneously on one transcript.

Third-generation sequencing such as Oxford Nanopore technology is rapidly improving. It already allows the mapping of RNA base modifications, e.g., of m⁶A and ψ , based on current and dwell time (Leger et al. 2021). Recently, a novel SHAPE-MaP reagent, 1-acetylimidazole, has been demonstrated to generate RNA adducts which can be used for structural mapping in single-molecule sequencing (Stephenson et al. 2022). It will be exciting to apply this technique to lncRNAs and to study different mRNA isoforms. Importantly, NanoSHAPE opens an unprecedented advance to analyse modifications and structural mapping in parallel. This is central to RNA research because base modifications impose an immense impact on RNA secondary structure. For instance, a single m⁶A in the lncRNA MALAT1 disrupts a duplex hairpin structure, thereby exposing the single-stranded U-tract for access to an m⁶A reader protein hnRNP-C, with downstream effects on mRNA processing (Liu et al. 2015). m⁶A can also alter the RNA structure to facilitate the binding of low-complexity RBPs (Liu et al. 2017a). On a transcriptome-wide level, it will be of interest to develop deconvolution techniques for conformational ensembles of

e.g. m⁶A-containing transcripts, to derive preferred conformations for m⁶A-modified mRNAs from structural probing data.

A full picture of 3D RNA tertiary structure can be obtained by applying biophysical low-resolution techniques. Solution structures are studied with single-angle X-ray scattering (SAXS) and with advanced atomic force microscopy (AFM) approaches (Ding 2023; Fang et al. 2015b; Lee et al. 2023). While traditional structural analysis by X-ray crystallography or NMR spectroscopy provide a high-resolution tertiary structure, those are limited to short transcripts or isolated domains, and have only been applied successfully to a few RNA targets (Chillón and Marcia 2020). Electron microscopy, i.e., cryo-EM or negative staining EM, is of increasing importance to study RNA structural dynamics and conformational ensembles. Careful sample preparation of full-length transcripts can provide detailed structures of single RNAs or in complex with their cognate RBP (Bonilla and Kieft 2022; Ma et al. 2022).

A combination of secondary structure probing and biophysical techniques for tertiary structure and dynamics in solution is best suited to gain a comprehensive picture of an RNA target. Importantly, the integration of different experimental data into bioinformatic prediction tools is constantly advancing to obtain more accurate RNA structure models (Li et al. 2020). If no solution or crystal samples are available, Förster resonance energy transfer (FRET) can be applied to study distances and long-range interactions in fluorescently tagged RNA molecules. Single-molecule FRET (smFRET) probes folding dynamics of an immobilized RNA molecule and can deconvolute conformational changes in RNP assembly processes such as ribosome biogenesis and transcription (Duss et al. 2018; Feng et al. 2021).

Drug development is poorly established for cellular RNA targets, mainly owing to their conformational diversity and dynamics. In future, it will be of high biomedical interest to screen for small molecules that target specific disease-associated RNAs, as exemplified for the lncRNA Xist (Aguilar et al. 2022). The targeting of a stable secondary structure motif, such as an individual G4 of specific sequence, can guide the way to target an individual disease-related transcript. The LNA-modified DNA probes provide an important basis to use antisense oligonucleotides as tools to interfere with stable DNA structures. It will be exciting to see how LNA probes can be designed to selectively bind G4 structures formed in RNA. Further development of G4-disrupting molecules as opposed to G4-stabilising probes will be crucial, since studies have reported the uncontrolled accumulation of DNA G4s associated with diseases like ALS (Simone et al. 2018).

References

- Aguilar R, Spencer KB, Kesner B et al (2022) Targeting Xist with compounds that disrupt RNA structure and X inactivation. *Nature* 604:160–166
- Aw JGA, Lim SW, Wang JX et al (2021) Determination of isoform-specific RNA structure with nanopore long reads. *Nat Biotechnol* 39:336–346

- Biffi G, Di Antonio M, Tannahill D et al (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat Chem* 6:75–80
- Bonilla SL, Kieft JS (2022) The promise of cryo-EM to explore RNA structural dynamics. *J Mol Biol* 434:167802
- Brion P, Westhof E (1997) Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct* 26:113–137
- Butcher SE, Pyle AM (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc Chem Res* 44:1302–1311
- Chambers VS, Marsico G, Boutell JM et al (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol* 33:877–881
- Chen X-C, Chen S-B, Dai J et al (2018) Tracking the dynamic folding and unfolding of RNA G-quadruplexes in live cells. *Angew Chem Int Ed Engl* 57:4702–4706
- Cheng CY, Chou F-C, Kladwang W et al (2015) Consistent global structures of complex RNA states through multidimensional chemical mapping. *Elife* 4:e07600
- Chillón I, Marcia M (2020) The molecular structure of long non-coding RNAs: emerging patterns and functional implications. *Crit Rev Biochem Mol Biol* 55:662–690
- Chowdhury S, Wang J, Nuccio SP et al (2022) Short LNA-modified oligonucleotide probes as efficient disruptors of DNA G-quadruplexes. *Nucleic Acids Res* 50:7247–7259
- Collie GW, Haider SM, Neidle S et al (2010) A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res* 38:5569–5580
- Costa M, Monachello D (2014) Probing RNA folding by hydroxyl radical footprinting. *Methods Mol Biol* 1086:119–142
- Das R, Laederach A, Pearlman SM et al (2005) SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* 11:344–354
- Di Antonio M, Biffi G, Mariani A et al (2012) Selective RNA versus DNA G-quadruplex targeting by in situ click chemistry. *Angew Chem Int Ed Engl* 51:11073–11078
- Ding J (2023) High-resolution atomic force microscopy imaging of RNA molecules in solution. *Methods Mol Biol* 2568:133–145
- Ding Y, Tang Y, Kwok CK et al (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505:696–700
- Duss O, Stepanyuk GA, Grot A et al (2018) Real-time assembly of ribonucleoprotein complexes on nascent RNA transcripts. *Nat Commun* 9:5087
- Einarson OJ, Sen D (2017) Self-biotinylation of DNA G-quadruplexes via intrinsic peroxidase activity. *Nucleic Acids Res* 45:9813–9822
- Fang R, Moss WN, Rutenberg-Schoenberg M et al (2015a) Probing Xist RNA structure in cells using targeted Structure-seq. *PLoS Genet* 11:e1005668
- Fang X, Stagno JR, Bhandari YR et al (2015b) Small-angle X-ray scattering: a bridge between RNA secondary structures and three-dimensional topological structures. *Curr Opin Struct Biol* 30:147–160
- Feng XA, Poyton MF, Ha T (2021) Multicolor single-molecule FRET for DNA and RNA processes. *Curr Opin Struct Biol* 70:26–33
- Gherghe CM, Shajani Z, Wilkinson KA et al (2008) Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S_2) in RNA. *J Am Chem Soc* 130:12244–12245
- Götte M, Marquet R, Isel C et al (1996) Probing the higher order structure of RNA with peroxonitrous acid. *FEBS Lett* 390:226–228
- Guo JU, Bartel DP (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* 353:aaf5371
- Hajdin CE, Bellaousov S, Huggins W et al (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc Natl Acad Sci USA* 110:5498–5503
- Harris KA, Crothers DM, Ullu E (1995) In vivo structural analysis of spliced leader RNAs in *Trypanosoma brucei* and *Leptomonas collosoma*: a flexible structure that is independent of cap4 methylations. *RNA* 1:351–362

- Hector RD, Burlacu E, Aitken S et al (2014) Snapshots of pre-rRNA structural flexibility reveal eukaryotic 40S assembly dynamics at nucleotide resolution. *Nucleic Acids Res* 42:12138–12154
- Heilman-Miller SL, Woodson SA (2003) Effect of transcription on folding of the *Tetrahymena* ribozyme. *RNA* 9:722–733
- Herschlag D (1995) RNA chaperones and the RNA folding problem. *J Biol Chem* 270:20871–20874
- Homan PJ, Favorov OV, Lavender CA et al (2014) Single-molecule correlated chemical probing of RNA. *Proc Natl Acad Sci U S A* 111:13858–13863
- Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* 33:2908–2916
- Incarnato D, Oliviero S (2017) The RNA epistructurome: uncovering RNA function by studying structure and post-transcriptional modifications. *Trends Biotechnol* 35:318–333
- Incarnato D, Neri F, Anselmi F et al (2014) Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol* 15:491
- Jaeger JA, Turner DH, Zuker M (1989) Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci USA* 86:7706–7710
- Keating KS, Toor N, Pyle AM (2008) The GANC tetraloop: a novel motif in the group IIC intron structure. *J Mol Biol* 383:475–481
- Kertesz M, Wan Y, Mazor E et al (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467:103–107
- Kharel P, Becker G, Tsvetkov V et al (2020) Properties and biological impact of RNA G-quadruplexes: from order to turmoil and back. *Nucleic Acids Res* 48:12534–12555
- Kielbinski LJ, Vinther J (2014) Massive parallel-sequencing-based hydroxyl radical probing of RNA accessibility. *Nucleic Acids Res* 42:e70
- Klein DJ, Schmeing TM, Moore PB et al (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J* 20:4214–4221
- Kwok CK, Marsico G, Sahakyan AB et al (2016) rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods* 13:841–844
- Laguerre A, Hukezalie K, Winckler P et al (2015) Visualization of RNA-quadruplexes in live cells. *J Am Chem Soc* 137:8521–8525
- Lai D, Proctor JR, Meyer IM (2013) On the importance of cotranscriptional RNA structure formation. *RNA* 19:1461–1473
- Lee Y-T, Fan L, Ding J et al (2023) Combining biophysical methods for structure-function analyses of RNA in solution. *Methods Mol Biol* 2568:165–177
- Leger A, Amaral PP, Pandolfini L et al (2021) RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nat Commun* 12:7198
- Li B, Cao Y, Westhof E et al (2020) Advances in RNA 3D structure modeling using experimental data. *Front Genet* 11:574485
- Liu N, Dai Q, Zheng G et al (2015) *N*⁶-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* 518:560–564
- Liu N, Zhou KI, Parisien M et al (2017a) *N*⁶-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res* 45:6051–6063
- Liu F, Somarowthu S, Pyle AM (2017b) Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat Chem Biol* 13:282–289
- Lu ZJ, Gloor JW, Mathews DH (2009) Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA* 15:1805–1813
- Lucks JB, Mortimer SA, Trapnell C et al (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc Natl Acad Sci U S A* 108:11063–11068
- Ma H, Jia X, Zhang K et al (2022) Cryo-EM advances in RNA structure determination. *Signal Transduct Target Ther* 7:58
- Mathews DH (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA* 10:1178–1190

- Mathews DH, Disney MD, Childs JL et al (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA* 101:7287–7292
- McGinnis JL, Weeks KM (2014) Ribosome RNA assembly intermediates visualized in living cells. *Biochemistry* 53:3237–3247
- McGinnis JL, Dunkle JA, Cate JHD et al (2012) The mechanisms of RNA SHAPE chemistry. *J Am Chem Soc* 134:6617–6624
- Merino EJ, Wilkinson KA, Coughlan JL et al (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 127:4223–4231
- Mestre-Fos S, Penev PI, Suttapitugsakul S et al (2019) G-quadruplexes in human ribosomal RNA. *J Mol Biol* 431:1940–1955
- Mestre-Fos S, Ito C, Moore CM et al (2020) Human ribosomal G-quadruplexes regulate heme bioavailability. *J Biol Chem* 295:14855–14865
- Mitchell D, Ritchey LE, Park H et al (2018) Glyoxals as in vivo RNA structural probes of guanine base-pairing. *RNA* 24:114–124
- Mitra S, Shcherbakova IV, Altman RB et al (2008) High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res* 36:e63
- Morandi E, van Hemert MJ, Incarnato D (2022) SHAPE-guided RNA structure homology search and motif discovery. *Nat Commun* 13:1722
- Mortimer SA, Weeks KM (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* 129:4144–4145
- Mortimer SA, Weeks KM (2008) Time-resolved RNA SHAPE chemistry. *J Am Chem Soc* 130:16178–16180
- Mortimer SA, Weeks KM (2009) Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat Protoc* 4:1413–1421
- Novikova IV, Dharap A, Hennelly SP et al (2013) 3S: shotgun secondary structure determination of long non-coding RNAs. *Methods* 63:170–177
- Piao M, Li P, Zeng X et al (2022) An ultra low-input method for global RNA structure probing uncovers Regnase-1-mediated regulation in macrophages. *Fundam Res* 2:2–13
- Puig Lombardi E, Londoño-Vallejo A (2020) A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res* 48:1–15
- Quigley GJ, Rich A (1976) Structural domains of transfer RNA molecules. *Science* 194:796–806
- Rice GM, Leonard CW, Weeks KM (2014) RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA* 20:846–854
- Ritchey LE, Su Z, Tang Y et al (2017) Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure in vivo. *Nucleic Acids Res* 45:e135
- Rouskin S, Zubradt M, Washietl S et al (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505:701–705
- Russell R (2008) RNA misfolding and the action of chaperones. *Front Biosci* 13:1–20
- Russell R (2013) *Biophysics of RNA folding*, 1st edn. *Biophysics for the Life Sciences*, vol 3. Springer New York, Imprint Springer, New York
- Schroeder R, Grossberger R, Pichler A et al (2002) RNA folding in vivo. *Curr Opin Struct Biol* 12:296–300
- Scavi B, Woodson S, Sullivan M et al (1997) Time-resolved synchrotron X-ray footprinting, a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding. *J Mol Biol* 266:144–159
- Seetin MG, Kladwang W, Bida JP et al (2014) Massively parallel RNA chemical mapping with a reduced bias MAP-seq protocol. *Methods Mol Biol* 1086:95–117
- Shcherbakova I, Mitra S, Laederach A et al (2008) Energy barriers, pathways, and dynamics during folding of large, multidomain RNAs. *Curr Opin Chem Biol* 12:655–666

- Shi H, Moore PB (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* 6:1091–1105
- Siegfried NA, Busan S, Rice GM et al (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 11:959–965
- Simone R, Balendra R, Moens TG et al (2018) G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO Mol Med* 10:22–31
- Spitale RC, Crisalli P, Flynn RA et al (2013) RNA SHAPE analysis in living cells. *Nat Chem Biol* 9:18–20
- Spitale RC, Flynn RA, Zhang QC et al (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* 519:486–490
- Stephenson W, Razaghi R, Busan S et al (2022) Direct detection of RNA modifications and structure using single-molecule nanopore sequencing. *Cell Genom* 2:100097
- Sutton JL, Pollack L (2015) Tuning RNA flexibility with helix length and junction sequence. *Biophys J* 109:2644–2653
- Tamura M, Holbrook SR (2002) Sequence and structural conservation in RNA ribose zippers. *J Mol Biol* 320:455–474
- Turner DH (1989) Thermodynamics and kinetics of base-pairing and of DNA and RNA self-assembly and helix coil transition. *Nucleic Acids*
- Tyrrell J, McGinnis JL, Weeks KM et al (2013) The cellular environment stabilizes adenine riboswitch RNA structure. *Biochemistry* 52:8777–8785
- Underwood JG, Uzilov AV, Katzman S et al (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat Methods* 7:995–1001
- Uroda T, Anastasakou E, Rossi A et al (2019) Conserved pseudoknots in lncRNA MEG3 are essential for stimulation of the p53 pathway. *Mol Cell* 75:982–995.e9
- Vasa SM, Guex N, Wilkinson KA et al (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* 14:1979–1990
- Waduge P, Sakakibara Y, Chow CS (2019) Chemical probing for examining the structure of modified RNAs and ligand binding to RNA. *Methods* 156:110–120
- Walter AE, Turner DH (1994) Sequence dependence of stability for coaxial stacking of RNA helices with Watson-Crick base paired interfaces. *Biochemistry* 33:12715–12719
- Walter AE, Turner DH, Kim J et al (1994) Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc Natl Acad Sci USA* 91:9218–9222
- Wan Y, Qu K, Ouyang Z et al (2012) Genome-wide measurement of RNA folding energies. *Mol Cell* 48:169–181
- Weeks KM (1997) Protein-facilitated RNA folding. *Curr Opin Struct Biol* 7:336–342
- Wells SE, Hughes JM, Haller Igel A et al (2000) Use of dimethyl sulfate to probe RNA structure in vivo. In: *RNA-ligand interactions part B*, vol 318. Elsevier, pp 479–493
- Weng X, Gong J, Chen Y et al (2020) Keth-seq for transcriptome-wide RNA structure mapping. *Nat Chem Biol* 16:489–492
- Westhof E, Masquida B, Jaeger L (1996) RNA tectonics: towards RNA design. *Fold Des* 1:R78–R88
- Xia T, SantaLucia J, Burkard ME et al (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37:14719–14735
- Yang SY, Lejault P, Chevrier S et al (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat Commun* 9:4730
- Yesselman JD, Denny SK, Bisaria N et al (2019) Sequence-dependent RNA helix conformational preferences predictably impact tertiary structure formation. *Proc Natl Acad Sci USA* 116:16847–16855
- Ziehler WA, Engelke DR (2001) Probing RNA structure with chemical reagents and enzymes. *Curr Protoc Nucleic Acid Chem* Chapter 6:Unit 6.1