



REVIEW

RNA Regulations and Functions Decoded by Transcriptome-wide RNA Structure Probing



Meiling Piao^a, Lei Sun^b, Qiangfeng Cliff Zhang^{*,c}

MOE Key Laboratory of Bioinformatics, Beijing Advanced Innovation Center for Structural Biology, Center for Synthetic and Systems Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

Received 27 January 2017; revised 9 May 2017; accepted 27 May 2017
 Available online 12 October 2017

Handled by Jinbiao Ma

KEYWORDS

RNA structure probing;
 RNA structure;
 RNA secondary structure;
 Structure–function relationship;
 RNA regulation

Abstract RNA folds into intricate structures that are crucial for its functions and regulations. To date, a multitude of approaches for probing structures of the whole transcriptome, *i.e.*, **RNA structures**, have been developed. Applications of these approaches to different cell lines and tissues have generated a rich resource for the study of **RNA structure–function relationships** at a systems biology level. In this review, we first introduce the designs of these methods and their applications to study different RNA structures. We emphasize their technological differences especially their unique advantages and caveats. We then summarize the structural insights in RNA functions and regulations obtained from the studies of RNA structures. And finally, we propose potential directions for future improvements and studies.

Introduction

RNA is a molecule with diverse functions. In addition to transferring genetic information from DNA to protein, RNA can catalyze specific biochemical reactions, similar to the action of a protein enzyme. These RNA enzymes, *i.e.*, ribozymes, are vital to life by participating in a variety of basic biological processes, including RNA splicing, translation, and also tRNA

biosynthesis [1]. Some RNAs, known as riboswitches, can also regulate gene expression by altering their own conformations in response to changes in the cellular environment or binding of ligands [2–4]. These, plus their ability to encode genetic information like many RNA viruses, have stimulated an interesting “RNA world” hypothesis, speculating that RNA may have been precursors to all life on Earth [5,6].

One of the most significant findings in genomics in the last two decades is the discovery of pervasive transcription and the large number of non-coding RNAs (ncRNAs) in human transcriptome [7]. RNAs that have no or little coding potential and are longer than 200 nucleotides are collectively defined as long ncRNAs (lncRNAs) [8]. Many lncRNAs are found to carry out different types of functions, including regulating chromatin states and consequently gene expression, sponging small RNAs (sRNAs) and proteins for fast cellular regulations, or

* Corresponding author.

E-mail: qc Zhang@tsinghua.edu.cn (Zhang QC).

^a ORCID: 0000-0003-3586-1743.

^b ORCID: 0000-0003-1400-2157.

^c ORCID: 0000-0002-4913-0338.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2017.05.002>

1672-0229 © 2017 The Authors. Production and hosting by Elsevier B.V. on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

being scaffolds to bring together other RNAs and proteins to facilitate their cross-talking [8]. However, with the broad definition of lncRNAs and their big variations in sequence and expression, it is very challenging to understand what their functions are and how they are regulated.

Fortunately, there has been a well-established general rule that sequence determines structure determines function, especially for proteins [9]. RNA can also fold into intricate shapes by local and long-range pairing of nucleotides. Well-known examples include the aforementioned ribozymes, riboswitches, and some lncRNAs. Studies have shown that, like protein structures, RNA structures are critical for their correct functioning and aberrancy of RNA structures could possibly lead to human disease [10,11]. It is thus a valid approach to study RNA functions and regulations from the perspective of RNA structures.

However, our current knowledge on RNA structures is very limited. Traditionally, structures of macromolecules are resolved with methods including X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and more recently cryo-electron microscopy (cryo-EM). Unfortunately, RNA molecules are usually much more flexible, dynamic, and thus structurally heterogeneous [12]. Therefore, it is usually very difficult to apply these techniques to obtain RNA structures. For example, X-ray crystallography requires that target structures form highly-ordered crystals. But in most cases, without the help of binding proteins, RNA molecules can adopt a big number of alternative structures, which makes crystallization challenging or even impossible. The large number of different conformations may also be well beyond the capability of new technologies like cryo-EM. Moreover, NMR spectroscopy is strongly limited to sRNAs and cannot be applied to study many other functional RNA molecules [13]. In addition, none of these methods can be used to study RNAs under cellular or physiological conditions.

Thus, for many years, our understanding on RNA structures and their functional relevance relies primarily on computational predictions. These predictions typically use thermodynamic calculation to obtain secondary structure models with lowest free energy [14–16], or sequence co-variation analysis to determine base-pairings that have been maintained through evolution [17,18]. However, these approaches usually cannot take into account *trans*-acting factors like proteins, other RNAs, and small ligands, as well as other physiological conditions. As a consequence, they can only generate *in silico* secondary structural predictions of a given RNA alone. In addition, computational predictions do not work very well for big RNAs with complex structural elements like pseudoknots, kissing loops, or long-range interactions [19,20].

Fortunately, recently we have witnessed the fast development of a new type of approaches, resurging from RNA structure probing analysis with chemicals and enzymes developed as early as 1970s [21]. It has been long known that a wide variety of chemicals and enzymes can react differently with RNA nucleotides in different structure context (Table 1). Many of these reactions leave footprints on the modified RNA molecules, which can be read out with gel filtration, or sequencing nowadays. Indeed, when combined with deep sequencing, these methods have the potential to reveal structures of the

whole transcriptome, *i.e.*, RNA structurome, in a single experiment.

In this review, we will introduce the designs of these methods and their applications to study RNA structuromes in different species. We will emphasize their technological differences, especially their unique advantages and caveats. We will then summarize the structural insights into RNA functions and regulations obtained from the studies of RNA structuromes. And finally, we propose potential directions for future improvements and studies.

Methods for probing RNA structurome

Different enzymatic cleavages and chemical modifications have distinct preferences toward single-stranded RNA (ssRNA) and double-stranded RNA (dsRNA). RNA structure probing approaches are designed to utilize these structural preferences. The technological development follows a path from low-throughput to genome-wide, from *in vitro* to *in vivo*, and from one-dimension to two-dimension analyses (Table 1).

Enzymatic cleavage methods

Endonucleases were found to have structural specificity decades ago [44]. Since then, a variety of nucleases have been used to generate cleavage sites that contain structural information. For instance, RNase A and T1 cleave unpaired regions and generate products with 5'-OH and 3'-P end [45,46], whereas RNase V1 only cleaves paired regions and generates 3'-OH and 5'-P end [47].

Nuclease digestion is then followed by high-throughput sequencing to read cleavage sites (Figure 1). Parallel analysis of RNA structures (PARS) is the first study to obtain mRNA secondary structure profile at transcriptome level, by utilizing RNase S1 and RNase V1 to cut ssRNA and dsRNA regions, respectively [22]. The ratio of cleavage sites of RNase V1 to RNase S1 is calculated as a PARS score to represent the tendency to form RNA secondary structures at single-nucleotide resolution. Similarly, parallel analysis of RNA structures with temperature elevation (PARTE) applies RNase-seq of V1 at a series of elevating temperatures to calculate RNA folding energies [23], whereas fragmentation sequencing (FragSeq) relies on nuclease P1 to cleave ssRNA [26]. These methods all focus on analysis of cleavage sites as output from the experiments. Alternatively, dsRNA-Seq and ssRNA-Seq (ds/ssRNA-Seq) look for enriched RNase-insensitive ssRNA and dsRNA regions, respectively, after thorough digestion by dsRNase and ssRNase [27–29]. Protein interaction profile sequencing (PIP-seq) incorporates ds/ssRNA-Seq with crosslinking methods. For PIP-seq, RNAs are firstly crosslinked with proteins and ds/ssRNA-Seq is applied with and without proteinase treatment [31]. It is also noteworthy to mention that hydroxyl radical can react with RNA riboses exposed to the solvent and lead to RNA cleavage. Hydroxyl radical footprinting (HRF) profiles RNA solvent-accessibility by sequencing these RNA cleavage sites [48].

Using the enzymatic cleavage methods, *in vitro* RNA structuromes of multiple species have been successfully generated. In particular, deproteinized PARS is able to measure near

Table 1 Summary of high-throughput approaches to probing RNA structure

Name	Type	Reagent	<i>In vitro/in vivo</i>	Organisms studied	Features and limitations	Refs.
Enzymatic cleavage	PARS, PARTE	Nuclease S1 and RNase V1	<i>In vitro</i>	Yeast, human	F: PARTE can calculate RNA folding energies L: Nuclease cannot permeate through cell membrane, making <i>in vivo</i> study impossible; signal is rather sparse because of large size of enzymes	[22–25]
	FragSeq	Nuclease P1	<i>In vitro</i>	Mouse	F: Similar to PARS, using samples without nuclease and with or without PNK as controls; focusing on short nuclear RNA to avoid fragmentation L: <i>In vitro</i> only; limited resolution	[26]
	ds/ssRNA-seq	RNase I, RNase V	<i>In vitro</i>	Arabidopsis, <i>Drosophila</i> , <i>C. elegans</i>	F: Sequencing the remaining regions after thorough digestion with ds/ssRNA nuclease L: <i>In vitro</i> only; limited resolution	[27–30]
	PIP-seq	RNase one, RNase V1	<i>In vitro</i>	Arabidopsis	F: Revealing relationship between RBP occupancy and RNA secondary structure L: Assuming that removing RBP doesn't affect RNA structure; <i>in vitro</i> only; limited resolution	[31]
Nucleotide modification	DMS-seq	DMS	<i>In vitro and in vivo</i>	Yeast, human	F: DMS can permeate through cell membrane, able to be used in living cells L: DMS has nucleotide bias, only able to react with adenines and cytosines	[32]
	DMS-MaPseq	DMS	<i>In vitro and in vivo</i>	Yeast, human	F: Utilizing the mutation rate caused by modification as the output signal instead of RT stop; higher signal-to-noise ratio L: Requiring higher sequencing depth; nucleotide bias	[33]
	Structure-seq	DMS	<i>In vitro and in vivo</i>	Plants	F: Similar to DMS-seq, including background control via detecting RT stops without DMS treatment L: Nucleotide bias	[34]
	Mod-seq	DMS	<i>In vivo</i>	Yeast	F: Similar to DMS-seq; focusing on rRNAs L: Nucleotide bias	[35]
	CIRS-seq	DMS, CMCT	<i>In vitro</i>	Mouse	F: Allowing to probe four nucleotides by combining DMS and CMCT; avoiding effects of RBP via deproteinization	[36]
	icSHAPE	NAI-N3	<i>In vitro and in vivo</i>	Mouse, human	F: The first <i>in vivo</i> SHAPE probing study; modified fragments further enriched by biotin isolation, making the signal super clean	[37]
	SHAPE-MaP	1M7, 1M6, NMIA		Synthetic RNA	F: Utilizing the mutation rate caused by modification as the output signal instead of RT stop; higher signal-to-noise ratio L: Requiring higher sequencing depth	[38]
	SHAPE-seq 1.0/2.0	1M7, NMIA, BzCN		Synthetic RNA	F: SHAPE has no bias toward four nucleotides L: SHAPE cannot permeate through cell membrane	[39,40]

(continued on next page)

Table 1 (continued)

Name	Type	Reagent	<i>In vitro/in vivo</i>	Organisms studied	Features and limitations	Refs.
Base-pair crosslinking, proximity ligation	PARIS	AMT	<i>In vivo</i>	Mouse, human	F: Direct mapping of duplex groups; enrichment conducted by 2D gel filtration L: AMT can only crosslink uridine; efficiency of proximity ligation needs to be improved	[41]
	SPLASH	Biotinylated psoralen	<i>In vivo</i>	<i>E. coli</i> , yeast, human	F: Direct mapping of duplex group; enrichment conducted by biotin isolation L: Psoralen can only crosslink uridine; efficiency of proximity ligation needs to be improved	[42]
	LIGR-seq	AMT	<i>In vivo</i>	Human	F: Direct mapping of duplex group L: AMT can only crosslink uridine; efficiency of proximity ligation needs to be improved	[43]

Note: PARIS, parallel analysis of RNA structures; PARTE, parallel analysis of RNA structures with temperature elevation; FragSeq, fragmentation sequencing; ds/ssRNA-Seq, double-stranded RNA-Seq and single-stranded RNA-Seq; PIP-seq, protein interaction profile sequencing; DMS-MaPseq, dimethyl sulfate mutational profiling with sequencing; Mod-seq, map RNA chemical modification using high-throughput sequencing; CIRS-seq, chemical inference of RNA structures followed by massive parallel sequencing; SHAPE, selective 2' hydroxyl acylation analyzed by primer extension; icSHAPE, *in vivo* click SHAPE; SHAPE-MaP, SHAPE and mutational profiling; PARIS, psoralen analysis of RNA interactions and structures; SPLASH, sequencing of psoralen crosslinked, ligated, and selected hybrids; LIGR-seq, ligation of interacting RNA and high-throughput sequencing; DMS, dimethyl sulfide; CMCT, 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate; NAI-N3, 2-methylnicotinic acid imidazolidine-azide; IM7, 1-methyl-7-nitroisatoic anhydride; IM6, 1-methyl-6-nitroisatoic anhydride; NMIA, N-methylisatoic anhydride; BzCN, benzoyl cyanide; AMT, 4'-aminomethyl trioxsalen; PNK, polynucleotide kinase; RBP, RNA binding protein.

in vivo RNA structures [22]. The main drawback of these methods, however, is that normally nucleases are not permeable through the cell membrane, making *in vivo* probing very challenging or impossible.

Nucleotide modification methods

Some chemicals can modify RNA nucleotides in specific structure contexts. Some of these modifications can block reverse transcription (RT) preceding the modified sites and thus allow for detection by reading RT stop sites. Nucleotide modifications mainly fall into two groups: base modification and backbone modification. Dimethyl sulfide (DMS) is a base modification reagent, which is frequently used to alkylate Watson–Crick (WC) face of unpaired adenine (A) and cytosine (C), while 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate (CMCT) reacts with unpaired uracil and guanine (G) [49,50]. Different experiments have been designed based on DMS modification (Figure 1). DMS-seq probes enriched DMS modifications *in vivo*, *in vitro*, or after RNA denaturation [32]. Structure-seq also includes a background control without DMS treatment, which properly excludes natural RT stops [34]. Mod-seq is similar but assesses secondary structure of rRNA instead of enriched mRNA [35]. The recently-developed DMS-MaPseq method uses reverse transcriptase mismatch rather than truncation products in collecting RNA structure information, thus improving signal-to-noise ratio [33]. One of the main limitations of DMS, however, is that it only provides profiles of A and C. Chemical inference of RNA structures followed by massive parallel sequencing (CIRS-seq) thus combines DMS and CMCT to cover all four nucleobases in natively-folded RNA with deproteinization [36].

RNA base-pairing brings geometric constraints to ribose involved in secondary structure, and thus protects the backbone from chemical modification [51]. Selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) is a method based on backbone modification for probing RNA secondary structure [51]. SHAPE reagents, like 1-methyl-7-nitro-isatoic anhydride (IM7) and N-methylisatoic anhydride (NMIA), specifically modify the ribose of unstructured nucleotides without any bias toward one or more of the four nucleobases. Notably, complex structures besides canonical base pairing are also detectable by SHAPE reagents, allowing for probing more complicated interactions. SHAPE-seq was designed by adapting SHAPE methods in combination with high-throughput sequencing [52]. Although limited by permeability of the probing reagents (IM7 and NMIA), SHAPE-seq provides the unbiased RNA secondary structure profiling. SHAPE modification has been found to generate mutations during RT reaction after alteration of RT reaction conditions [38]. Based on this discovery, SHAPE-MaP locates the modification sites by analyzing mutation sites [38] (Figure 1). SHAPE reagents were initially used to probe RNA structure *in vitro*. But cell-permeable ones, e.g., 2-methyl-3-furoic acid imidazolidine (FAI) and 2-methylnicotinic acid imidazolidine (NAI), can be used for *in vivo* RNA structure probing [53]. A newly-developed technology, *in vivo* click SHAPE (icSHAPE) uses an optimized SHAPE compound NAI-N3 with increased permeability and incorporates biotin-streptavidin isolation to enrich modified RNA fragments [37] (Figure 1). The

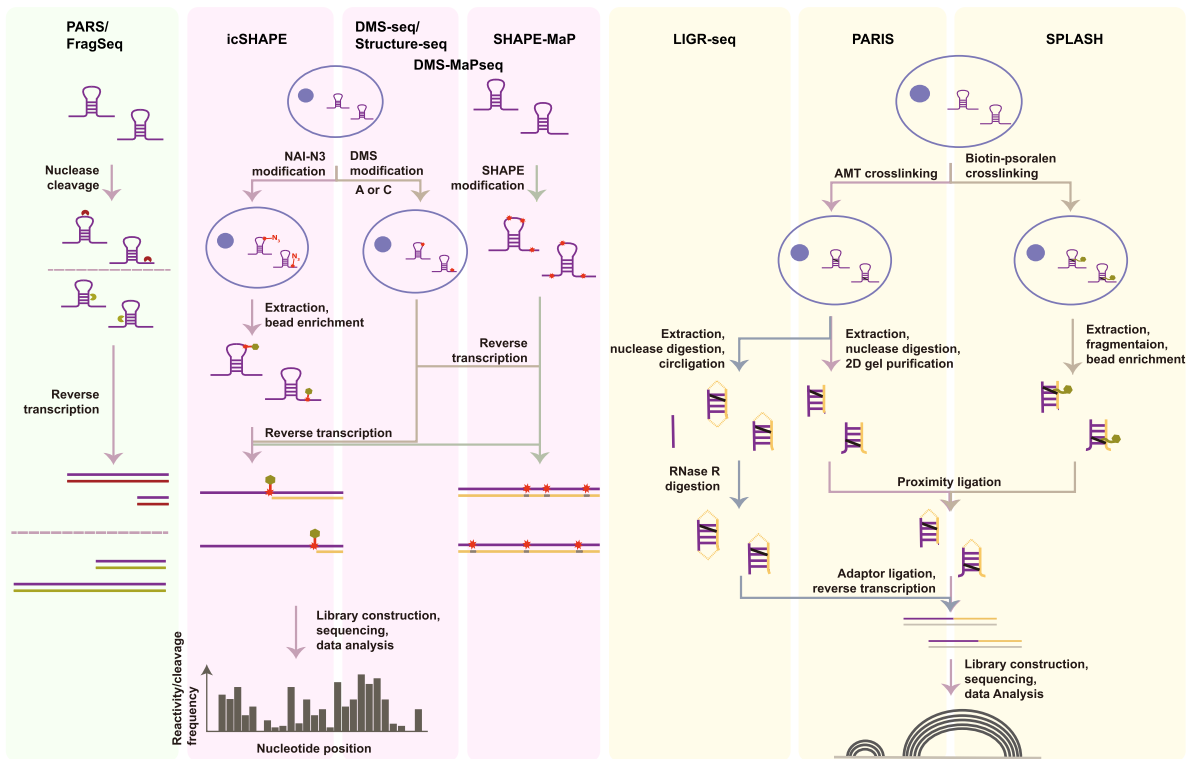


Figure 1 Experimental workflow of some representative high-throughput approaches for RNA structure probing

PARS, parallel analysis of RNA structures; FragSeq, fragmentation sequencing; SHAPE, selective 2' hydroxyl acylation analyzed by primer extension; icSHAPE, *in vivo* click SHAPE; SHAPE-MaP, SHAPE and mutational profiling; DMS-MaPseq, dimethyl sulfate mutational profiling with sequencing; LIGR-seq, ligation of interacting RNA and high-throughput sequencing; PARIS, psoralen Analysis of RNA interactions and structures; SPLASH, sequencing of psoralen crosslinked, ligated, and selected hybrids; DMS, dimethyl sulfide; NAI-N3, 2-methylnicotinic acid imidazolide-azide; AMT, 4'-aminomethyl trioxsalen.

biotin-streptavidin isolation system is introduced to the clickable azide moiety of icSHAPE reagent, thereby achieving higher sensitivity for modified RNAs.

Cross-linking methods

The two aforementioned types of methods only focus on detecting which regions of RNA are single-stranded and which are double-stranded. Information on another big part of RNA structure, the detailed intermolecular or intramolecular base-pairing pattern, however, is missing. Psoralens are well known mutagens that crosslink DNA or RNA duplexes by forming adducts with adjacent thymines (Ts) or uridines (Us) when activated by UV photon [54,55]. Three recent studies, psoralen analysis of RNA interactions and structures (PARIS), sequencing of psoralen crosslinked, ligated, and selected hybrids (SPLASH), ligation of interacting RNA and high-throughput sequencing (LIGR-seq), utilize psoralens to crosslink the duplex regions of RNA [41–43] (Figure 1). The RNA is then fragmented and retrieved after RNase and protease digestion. Ends of the crosslinked fragment duplex are then ligated via proximity ligation, followed by reverse crosslinking and library construction for sequencing. After mapping, gapped reads are collected as indication of direct base-pairing.

Apparently, proximity ligation cannot tell duplex regions from unpaired fragments. Various methods are thus used in order to enrich duplex regions to reduce background noise. PARIS involves 2D gel electrophoresis to separate duplexes from unpaired regions, which gets clean duplexes but lowers the yield at the same time [41]. SPLASH uses biotinylated psoralen as the crosslinking reagent, which makes it possible to enrich duplexes by streptavidin beads [42]. LIGR-seq employs RNase R to digest uncircularized RNA after proximity ligation by CircLigase [43]. RNase R digestion, however, is performed after ligation, since ssRNAs still can be ligated and confound the results at this stage.

Insights of RNA functions and regulations from RNA structurome

RNA structure is crucial for gene function and regulation by influencing RNA transcription, processing, localization, translation, and degradation. The canonical roles of RNA structure in many different biological processes have been reviewed elsewhere [56–61]. Thanks to the transcriptome-wide RNA structure probing, we are now able to understand how RNA structure is regulated and functions at a systems level. Here, we only briefly summarize the novel insights into functional

significance of RNA structure in various cellular processes from these systems biology studies.

Transcription

The life cycle of a RNA molecule begins with its transcription. There is accumulating evidence showing that RNA folds into structures along with its transcription, and this co-transcriptional folding plays a critical role in defining RNA functions [62–66]. Intermediate structures often form and present for a certain period of time when transcription proceeds from the 5' to the 3' end of the RNA sequence [67,68], and later concedes to globally more stable conformations [63,69].

The relationship of RNA transcription and those transient intermediate structures remains elusive for most cases. A study in 1980s on the *tryptophan* (*trp*) operon showed that the formation of transient RNA structures is influenced by the interplay of co-transcriptional RNA folding and translation [70]. The leader of this *trp* operon encodes a tryptophan-rich peptide. Its transcript can assume two alternative structural elements: the attenuator and the anti-terminator. While the attenuator structure prevents further transcription, the anti-terminator permits it. These two structure elements form co-transcriptionally and are regulated by the binding and activity of ribosomes to the leader.

We currently do not have much insight into the *in vivo* RNA folding pathways during transcription. None of the aforementioned high-throughput probing experiments has investigated the interplay between RNA structures and transcription. However, combining nascent RNA sequencing with structure probing, or using other methods to carefully isolate different stages of transcripts, we may be able to study their intricate relationships. In addition, recent progresses on *in vivo* methodology development are likely to improve our ability to interrogate the role of RNA structures in transcription, and *vice versa* [71,72].

Processing

Most nascent RNAs are subjected to further processing, including capping, splicing, and polyadenylation before they are better prepared to meet their functional roles. Among those, alternative splicing (AS) is a widespread means that vastly increases transcript and protein diversity [73,74]. RNA splicing involves many *cis*-acting sequence elements. The basic ones are important for spliceosome binding and splicing reaction, including the 5' splice site, the branch-point, and the 3' splice site. In addition, several classes of auxiliary regulatory signals that play critical roles in splicing regulation have been defined as well. These include exon splicing enhancers (ESEs) and silencers (ESSs), as well as intron splicing enhancers (ISEs) and silencers (ISSs), categorized based on their location and their effects on splicing. Splicing factors that recognize and bind to the enhancer/silencer elements are defined as activator/repressor proteins, respectively. The structural effects of RNA on splicing involve these basic and auxiliary elements, as well as their interplay with the spliceosome complex and splicing factors.

Many examples have shown that RNA structure can regulate AS, by affecting spliceosome recognition of basic *cis*-acting elements [75] or through influencing splicing factor

binding to auxiliary elements [76]. There also exists another type of RNA structure regulation on splicing, *i.e.*, forming (usually long-range) base-pairings that facilitate joining of a common exon to different alternative exons. An incredibly interesting example comes from the AS of the *Drosophila* gene encoding Down syndrome cell adhesion molecule (*Dscam*) [77]. There have been some nice reviews that summarize individual instances in all these three aspects [58,78–80]. Here we only focus on general structural observations of splicing obtained in recent high-throughput structure probing experiments.

RNA structure can regulate splicing by directly interfering with spliceosome binding sites. In the PARS study of the *in vitro* human structurome, RNA structure signals were screened across exon-exon junctions in already spliced mRNAs. It is observed that the splicing donor dinucleotides AG are more accessible compared to nearby nucleotides, whereas the acceptor nucleotide G/A tends to be more structured [24] (Figure 2). Further analysis of the structural signal of pre-mRNA splicing in *Arabidopsis* structurome using ds/ssRNA-Seq confirmed this signature on splicing donor/acceptor sites with more details in flanking intronic regions [30]. The structure score of splicing donor, however, is higher than that of splicing acceptor in *Arabidopsis* structurome, which is opposite in human structurome. In a later study of the *in vivo Arabidopsis* structurome, Ding et al. revisited RNA splicing using the Structure-seq. They found that in the region upstream of the splicing donor site, structures are less accessible for the unspliced events than for the spliced events [34]. This suggests that secondary structure at the splice donor sites may disfavor splicing. This finding is consistent with the ds/ssRNA-Seq study of *in vivo Arabidopsis* structurome for U12-type introns and constitutive introns [30].

RNA structures can also regulate splicing by directly affecting splicing factor binding. The icSHAPE study of the mouse structurome analyzed RNA structure signatures in auxiliary regulatory sites [37], focusing on the splicing factor Rbfox2 (fox-1 homolog in mouse), a member of the “feminizing locus on X” (Fox) family of RNA-binding proteins (RBPs) [37]. Spitale and colleagues compared *in vivo* and *in vitro* RNA structures at Rbfox2-binding sites and found high level of differences, suggesting a strong structural effect of the splicing factor binding *in vivo*. The structural signatures and rearrangement were later shown to be effective in identifying true Rbfox2-binding sites [37]. The structure significance in defining a true splicing factor binding sites was also exemplified in a later study on the binding of heterogeneous nuclear ribonucleoprotein C (hnRNPC) to polyU tracts [81]. By integrative data analysis of m⁶A modification, RNA structures, and RBP binding sites [82], Hafner et al. found that m⁶A in the complementary strands of U-rich hairpins weakens the hairpin secondary structure and promotes hnRNPC binding. In addition, knock-down of two genes encoding m⁶A methyltransferases, *METTL3* and *METTL14*, reduced hnRNPC binding and affected AS through disrupting hairpin structures [81].

Localization

The majority of RNAs are localized to distinct cellular domains with exquisite temporal and spatial control, providing an important mechanism for gene expression regulation [59,83,84]. RNA localization is usually controlled through a

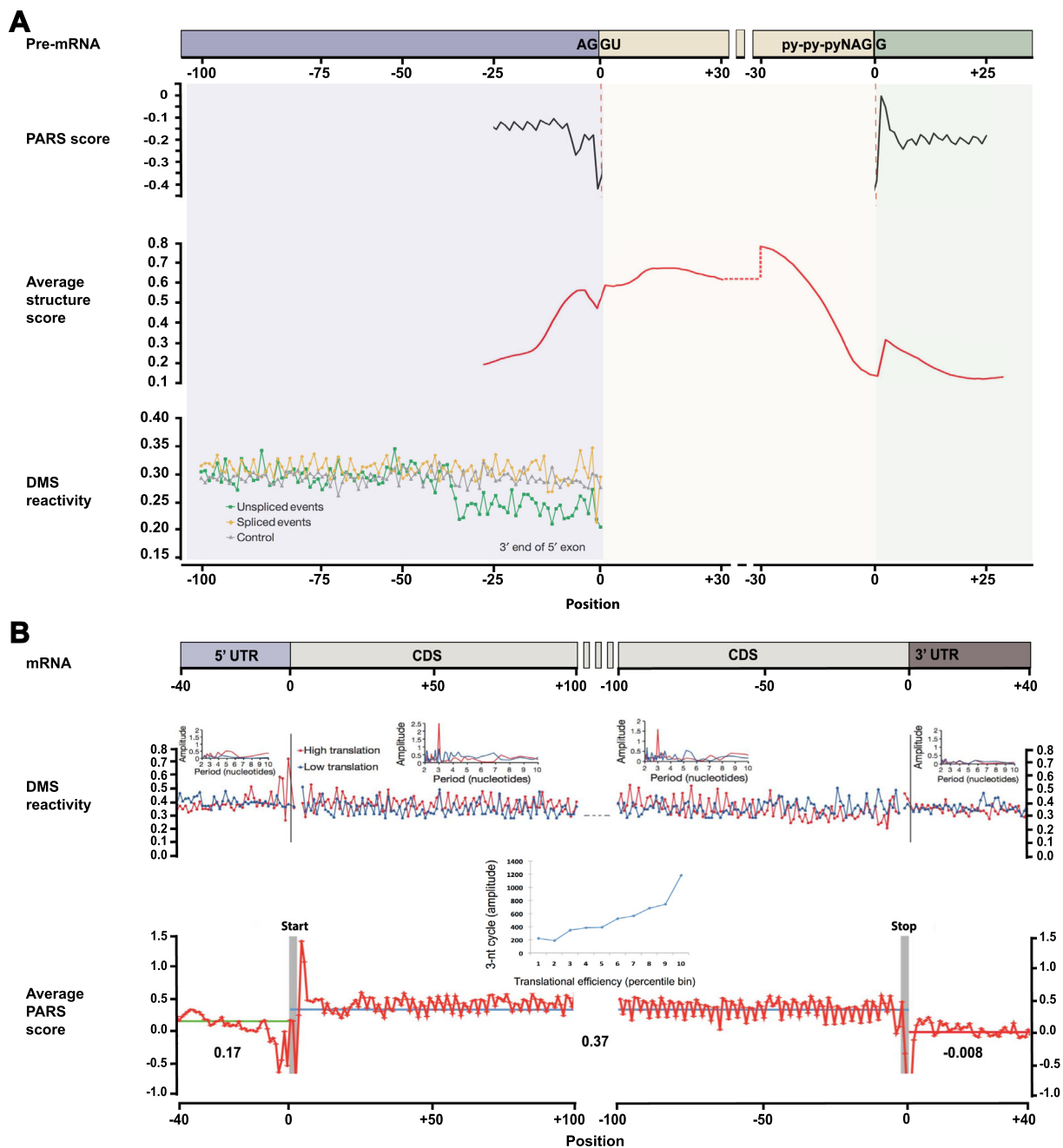


Figure 2 Structural landscapes of RNA splicing and translation

Structural landscapes of RNA splicing and translation are shown in panels **A** and **B**, respectively. Note that different structural score metrics are used in different technologies. For PARS and structure scores, the higher score means the more secondary structures, while for DMS reactivity, the higher the score means the fewer secondary structures. The figure is adapted from [25,36,39,74]. PARS: parallel analysis of RNA structures; DMS: dimethyl sulfide.

set of *cis*-acting elements present in the RNA, which encode the cellular “address” of the host transcripts. These *cis*-acting elements, primarily located within the 3' UTR, are called “localization elements” or “zipcodes” [85]. And a combination of RBPs, which often function in association with cytoskeletal motors, recognize these zipcodes to regulate RNA transport throughout the cell [86,87].

Accumulating evidences suggest that not only the sequences but also the structures of these zipcodes are critical for RNA localization. For example, a study showed that a stem-loop

structural element, BLE1, is critical for the transport of the host *bicoid* mRNA from the nurse cells into the oocyte [88]. More interestingly, sometimes the primary sequences lack conserved sequence zipcodes [89]. Therefore, efforts have been concentrated on the discovery of conserved structural motifs of zip codes, in particular stem loops [90,91].

The experimental studies of RNA structures can potentially shed light on zipcode discovery. The analysis of the *Saccharomyces cerevisiae* structurome has revealed that mRNA encoding proteins with specific sub-cellular locations or

involved in some metabolic pathways are more structured in the coding region [22]. On the contrary, mRNAs that encode subunits of the ribosome tend to have much less structure in their 5' UTR and coding sequences. More structural analyses, however, are needed to scrutinize, identify, and annotate structural motifs from this rich set of data.

RNA of some secretory proteins is exported from the nucleus by using a signal sequence coding region (SSCR) in the transcripts [92]. In the aforementioned study of the *S. cerevisiae* structurome, Kertesz et al. examined the structures of the transcripts that are predicted to encode a signal peptide. They found that the SSCRs and their proximity sequences have a lower PARS score, suggesting that specific secondary structures may assist RNA nuclear export [22].

Translation

Many RNAs are made for translation. Long before the high-throughput probing experiments, it had been observed that RNA structure plays an important role in translation regulation. For example, the temperature-sensitive structures, e.g., RNA thermometers or riboswitches, are able to affect mRNA translation [93]. These RNA structures adopt different conformations to inhibit or allow the binding of ribosomes, thus regulating expression of the encoded proteins [94]. Another example is from a study in 1980s. It was found that structures formed around the translation start site of an mRNA impede its translation initiation, a rate-limiting step that significantly influences translation efficiency [95]. This finding was later confirmed by a large-scale study calculating the mRNA folding near the ribosomal binding site and for correlation with protein abundance [96]. A later transcriptome-wide study repeated this by correlating predicted RNA structure [97] with translation efficiency from experimental polysome profiling [98]. In addition, using computational analysis, Shabalina et al. predicted an interesting distinguishing feature of the CDS region, i.e., a three-nucleotide periodicity in mRNA secondary structure [99]. This intriguing feature was later confirmed by many large-scale *in vitro* and *in vivo* RNA structure probing experiments [22,32,34,37]. The structural landscape of translation elements, including those of 5' UTR, start site, CDS, stop site, and 3' UTR, however, is much more complex as later revealed in these transcriptome-wide experimental studies. We will summarize relevant findings below.

The first whole-genome structural probing experiment was performed on HIV-1 genomic RNA using SHAPE technologies, focused on structure–translation relationships [100]. It was found that both the 5' UTR and the 3' UTR are associated with increased level of RNA secondary structures than coding region. Interestingly, there exists a distinct pattern of structures in coding region, which correlates well with protein and domain boundaries. These findings implicate a role of RNA structure in translation pausing and co-translational protein folding.

However, following studies of PARS [22,24], ds/ssRNA-seq [28,29], DMS-seq [32], and Structure-seq [34] experiments performed on different organisms revealed that the relative structural contents vary among 5' UTRs, 3' UTRs, and the coding regions (Figure 2). The studies on *Drosophila*, *Caenorhabditis elegans*, and human mRNAs agreed with the HIV-1 analysis, whereas opposite results were obtained for yeast and *Arabidop-*

sis. It is not entirely clear whether this is due to *in vitro* or *in vivo* structure features probed by different technologies or characteristics defined by species.

Nevertheless, the low structure contents around the start and stop codon, and also the three-nucleotide periodicity have been universally observed [22,24,34,36] (Figure 2). These studies further painted a finer global view of structure–translation relationships. In a study examining RNA structures in *S. cerevisiae* using PARS, Kertesz et al. observed that mRNA structure of the translation start site are negatively correlated with ribosome density throughout the transcript [22]. Notably, they also found that the three-nucleotide periodic repeat pattern is significantly correlated with translation efficiency. Given the *in vitro* nature of the aforementioned study, this observation was later revisited and confirmed in the *Arabidopsis* structurome study *in vivo* [34].

Moreover, an interesting positive correlation between the level of mRNA structure and its overall ribosome association has been revealed in another study of *Arabidopsis* structurome using ds/ssRNA-seq [29]. It is possible that mRNA structure could slow down or even stall the translocation of ribosomes and cause them to form clusters on mRNAs. Further investigation is needed to examine whether the increased ribosome association would affect protein translation and consequently its abundance.

The translation rates are not uniform along the CDS region. *In vitro* studies have suggested that the presence of RNA secondary structure promotes ribosome pausing [101]. However, complicated by multiple factors involved, including RNA structure, tRNA abundance, and codon choice, it is difficult to figure out how RNA structures *in vivo* may influence ribosome pausing, begging for more integrated quantitative studies. Interestingly, a mouse embryonic stem (ES) cell structurome study using icSHAPE technology [37] revealed a distinctive structure signature at ribosome pause sites: more structures at the exit (E) and peptidyl-tRNA (P) sites, and fewer structures at the aminoacyl-tRNA (A) site. This structural pattern was also observed *in vitro* when ribosome binding to mRNA is depleted, and in negative control sites with similar sequence context. This suggests that the RNA structure pattern of ribosome pausing sites is probably encoded in their sequences. Notably, the flanking 5' region of the negative control sites showed the lack of typical three-nucleotide periodic signal, suggesting it may play some role in ribosome pausing regulation [37].

Recently, a structurome study of yeast and human using psoralen crosslinking reported again that dense structures around the start codon could inhibit RNA translation, whereas structures of long-range 5'-to-3' interactions could promote translation [42]. Interestingly, it was found that large RNA conformational changes *in vivo* could change translation efficiency, suggesting a potential mechanism for translation regulation. In summary, all these genome-wide studies show that mRNA structures exert a significant effect on its translation at multiple levels in various organisms.

Stability and degradation

RNA is degraded in a carefully-controlled way at the end of its life cycle. RNA structures are also found to be involved in RNA stability and degradation as well. For instance, in

addition to regulating translation, the small structural elements of riboswitches can also influence RNA stability [102]. In eukaryotes, RNA degradation is mainly accounted for by the exosome complex, an exonuclease that works from the 3' to the 5' end and needs an ssRNA region of about 30-nucleotide in the 3' end of its targets [103]. This is consistent with the recent crosslinking study of yeast and human structuromes, which suggest that RNAs with structures present in the 5' end only are associated with faster degradation, whereas structures at the 3' end could inhibit exosome-mediated RNA decay [42].

It is then natural to speculate that RNA stability is positively correlated with RNA structure content in its 3' end. In a study on *in vitro* structurome of *S. cerevisiae*, the RNA folding energies were measured using PARTE, an expansion of the PARS technology by applying it at different temperatures [23]. As a result, it was found that mRNAs with low average melting temperatures (*i.e.*, less structured) decreased most rapidly in abundance following heat shock. Notably, inactivation of the exosome significantly decreased the degradation of mRNAs with unstable structures.

This finding is consistent with a study probing *E. coli* structurome *in vitro* using ds/ssRNA-seq. Del Campo et al. revealed that mRNA abundances are positively correlated with CDS secondary structures [25]. Conversely, mRNA abundances are found to be significantly negatively correlated with RNA structures in *Arabidopsis* [29]. Using Degradome sequencing, a weak but significant positive correlation was detected between mRNA structure and its turnover, suggesting that mRNAs of high level of structures are associated with increased rate of degradation. Further examination with small RNA-Seq analysis revealed a strong positive correlation between mRNA structure and sRNA abundance. This raises an interesting hypothesis that that RNA structures may be cleaved and processed into sRNAs [29].

Concluding remarks and outlook

Transcriptome-wide RNA structure maps, *i.e.*, RNA structuromes, and studies of structure–function relationships have generated many new insights into RNA biogenesis, processing, localization, translation, and degradation. To date, studies have been focusing on general principles of the most basic biological processes. In the future, more in-depth investigations are needed for specific biological events in certain cell lines, tissues, environments, or conditions including human diseases. For example, what are the roles of structure in the infection of RNA virus? Is RNA structure an important regulator in early development when transcription is silent? Can we find any RNA structure biomarkers in human disease and could they be a diagnosis target for disease development? Comparative studies are especially required to uncover structures that may be causative factors or direct effectors. The future of these applications is to infer RNA functions based on data mining and classification of structure elements in different biological contexts, thus providing a knowledgebase for functional and mechanistic studies.

In addition to high quality data generation, well-designed bioinformatics analysis is the key to these applications. Technologies have been developed including a computational framework that processes sequencing data. The data process-

ing follows normal RNA-seq analysis pipelines that include sequencing data quality control and trimming, reads alignment and abundance estimation. It also calculates a structure score to represent the preference for individual nucleotide or a sequence region to be single or double-strand. For enzymatic cleavage and nucleotide modification technologies, the structure score of a nucleotide is normally defined as the ratio of the number of RT stops or cleavage sites mapped to that position divided by the same number in a control experiment, in normal or log space [22,24,34,37,39,40]. More accurate algorithms with sophisticated statistical models are developed later for some methods. For example, Ouyang et al. used hypergeometric tests with false discovery rate adjustment to identify reliable structural states and then incorporate the information for RNA structure inference [104]. Aviran et al. introduced probabilistic framework that models polymerase drop-off and chemical modification, and uses maximum-likelihood estimation to infer structural state for every nucleotide [105]. Zou and Ouyang developed a joint Poisson-gamma mixture to model multiple RNase-seq data and combine it with hidden Markov model to infer RNA structures [106]. In a recent study, a beta-uniform mixture hidden Markov model is used to calculate a statistically interpretable score for nucleotide structure preference [107]. These new methods usually can yield higher accuracy for RNA structure inference or generate confident structure estimations at much lower sequence coverage levels.

Better structure probing technologies that combine the strengths of new chemicals and creative sequencing designs are also desired to provide more accurate and comprehensive structure information. For example, most probing methods, except for the SHAPE-MaP [38] and the recently-developed DMS-MaPseq [33], use reverse transcriptase truncation products to collect RNA structure information. However, as shown in the analysis of DMS-MaPseq, using reverse transcriptase mismatch can improve signal-to-noise ratio and can be used to probe structures of low-abundant transcripts. More importantly, the ability to report multiple structure features per sequencing read allows for probing RNAs in multiple conformations and single-molecule structure analyses based on co-occurrence of DMS modifications on one read.

Most probing methods that use enzymatic cleavage or base and sugar modifications only generate one-dimensional averaged structure information. Newly-developed methods that use UV or psoralen crosslinking can provide two-dimensional information, but their resolution and coverage are so far limited (*e.g.*, PARIS, SPLASH, and LIGR-seq [41–43]). Tools that can greatly improve our ability to obtain direct base-pairing information include: (a) crosslinkers that can connect different bases with high efficiency; (b) methods that can locate the exact sites of base-pairing; and (c) computational pipelines with higher resolution and accurate duplex confidence calculation.

It has been a long-standing question on the interplay between RNA structure and RBP binding. It is also of great interest to find out how this interplay would affect RNA probing. The PIP-seq method is able to obtain information on both protein binding and RNA structures. Silverman et al. analyzing RBP-binding sites in *Arabidopsis* and found that most of these sites were more of single-stranded fragments flanked by structured regions [31]. However, it remains a big myth whether the structure signature is a cause or a consequence.

It is possible both could be true as the interplay of protein binding and RNA structures is complex and possibly varies from one protein to another. With the rich resource of RNA structure data and RBP binding data from high-throughput CLIP experiments [108,109], it is now becoming possible to carry out a systematic study to investigate the relationship between RNA structure and RBP binding. But it should be noted that some structure probing methods could possibly generate biased information for this types of analysis. For example, RBP binding could cause steric effects that lead to inefficient cleavage or modification [110,111].

A long-term goal of the RNA structure study is to construct structure models of RNA and protein (RNP) complexes. To date, very limited number of RNA 3D experimental structures are available in the PDB database [112]. In the future, it would be necessary to integrate secondary structure probing methods with 3D methods like crystallization, NMR, and in particular cryo-EM, small angle X-ray scattering (SAXS) technologies. Although the development in cryo-EM now allows for structure determination of RNP complexes with near atomic resolution [113–116], it remains very challenging and success has been limited to a few cases. Nonetheless, cryo-EM, as well as SAXS, is very efficient in capturing overall shapes of big RNP complexes. Secondary structure probing should be able to help identify stable RNA structural domains to be fitted into the shape of the whole big RNP complexes. Finally, computational modeling will generate a high-resolution and complete picture of RNP complexes to help elucidate their regulations and functions and shed light on the mechanism and treatment of diseases related to RNA structures.

Competing interests

None.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 31671355) and the National Thousand Young Talents Program of China to QCZ.

References

- [1] Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 1982;31:147–57.
- [2] Grundy FJ, Henkin TM. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. *Mol Microbiol* 1998;30:737–49.
- [3] Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR. Genetic control by a metabolite binding mRNA. *Chem Biol* 2002;9:1043.
- [4] Mandal M, Breaker RR. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* 2004;5:451–63.
- [5] Gilbert W. Evolution of antibodies. The road not taken. *Nature* 1986;320:485–6.
- [6] Cech TR. The RNA worlds in context. *Cold Spring Harb Perspect Biol* 2012;4:a006742.
- [7] Birney E. Evolutionary genomics: come fly with us. *Nature* 2007;450:184–5.
- [8] Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* 2016;17:47–62.
- [9] Anfinsen CB. Principles that govern the folding of protein chains. *Science* 1973;181:223–30.
- [10] Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol* 2011;21:354–61.
- [11] Halvorsen M, Martin JS, Broadaway S, Laederach A. Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet* 2010;6:e1001074.
- [12] Schroeder R, Barta A, Semrad K. Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol* 2004;5:908–19.
- [13] Varani G, Aboulela F, Allain FHT. NMR investigation of RNA structure. *Prog Nucl Magn Reson Spectrosc* 1996;29:51–127.
- [14] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;31:3406–15.
- [15] Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res* 2003;31:3429–31.
- [16] Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 2010;11:129.
- [17] Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments. *Bioinformatics* 2009;25:1335–7.
- [18] Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 2008;9:474.
- [19] Puton T, Kozłowski LP, Rother KM, Bujnicki JM. CompaRNA: a server for continuous benchmarking of automated methods for RNA secondary structure prediction. *Nucleic Acids Res* 2013;41:4307–23.
- [20] Mathews DH, Turner DH. Prediction of RNA secondary structure by free energy minimization. *Curr Opin Struct Biol* 2006;16:270–8.
- [21] Silverman IM, Berkowitz ND, Gosai SJ, Gregory BD. Genome-wide approaches for RNA structure probing. *Adv Exp Med Biol* 2016;907:29–59.
- [22] Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, et al. Genome-wide measurement of RNA secondary structure in yeast. *Nature* 2010;467:103–7.
- [23] Wan Y, Qu K, Ouyang Z, Kertesz M, Li J, Tibshirani R, et al. Genome-wide measurement of RNA folding energies. *Mol Cell* 2012;48:169–81.
- [24] Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, et al. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 2014;505:706–9.
- [25] Del Campo C, Bartholomaeus A, Fedyunin I, Ignatova Z. Secondary structure across the bacterial transcriptome reveals versatile roles in mRNA regulation and function. *PLoS Genet* 2015;11:e1005613.
- [26] Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, Mathews DH, et al. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat Methods* 2010;7:995–1001.
- [27] Zheng Q, Ryvkin P, Li F, Dragomir I, Valladares O, Yang J, et al. Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. *PLoS Genet* 2010;6:e1001141.
- [28] Li F, Zheng Q, Ryvkin P, Dragomir I, Desai Y, Aiyer S, et al. Global analysis of RNA secondary structure in two metazoans. *Cell Rep* 2012;1:69–82.
- [29] Li F, Zheng Q, Vandivier LE, Willmann MR, Chen Y, Gregory BD. Regulatory impact of RNA secondary structure across the Arabidopsis transcriptome. *Plant Cell* 2012;24:4346–59.
- [30] Gosai SJ, Foley SW, Wang D, Silverman IM, Selamoglu N, Nelson AD, et al. Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the *Arabidopsis* nucleus. *Mol Cell* 2015;57:376–88.

- [31] Silverman IM, Li F, Alexander A, Goff L, Trapnell C, Rinn JL, et al. RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. *Genome Biol* 2014;15:R3.
- [32] Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* 2014;505:701–5.
- [33] Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo*. *Nat Methods* 2017;14:75–82.
- [34] Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 2014;505:696–700.
- [35] Talkish J, May G, Lin Y, Woolford Jr JL, McManus CJ. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* 2014;20:713–20.
- [36] Incarnato D, Neri F, Anselmi F, Oliviero S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol* 2014;15:491.
- [37] Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, et al. Structural imprints *in vivo* decode RNA regulatory mechanisms. *Nature* 2015;519:486–90.
- [38] Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 2014;11:959–65.
- [39] Mortimer SA, Trapnell C, Aviran S, Pachter L, Lucks JB. SHAPE-Seq: high-throughput RNA structure analysis. *Curr Protoc Chem Biol* 2012;4:275–97.
- [40] Loughrey D, Watters KE, Settle AH, Lucks JB. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res* 2014;42:e165.
- [41] Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 2016;165:1267–79.
- [42] Aw JG, Shen Y, Wilm A, Sun M, Lim XN, Boon KL, et al. *In vivo* mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol Cell* 2016;62:603–17.
- [43] Sharma E, Sterne-Weiler T, O’Hanlon D, Blencowe BJ. Global mapping of human RNA–RNA interactions. *Mol Cell* 2016;62:618–26.
- [44] Chang SH, RajBhandary UL. Studies on polynucleotides. LXXXI. Yeast phenylalanine transfer ribonucleic acid: partial digestion with pancreatic ribonuclease. *J Biol Chem* 1968;243:592–7.
- [45] Loverix S, Steyaert J. Deciphering the mechanism of RNase T1. *Methods Enzymol* 2001;341:305–23.
- [46] Raines RT, Ribonuclease A. *Chem Rev* 1998;98:1045–66.
- [47] Lowman HB, Draper DE. On the recognition of helical RNA by cobra venom V1 nuclease. *J Biol Chem* 1986;261:5396–403.
- [48] Kielpinski LJ, Vinther J. Massive parallel-sequencing-based hydroxyl radical probing of RNA accessibility. *Nucleic Acids Res* 2014;42:e70.
- [49] Lawley PD, Brookes P. Further studies on the alkylation of nucleic acids and their constituent nucleotides. *Biochem J* 1963;89:127–38.
- [50] Metz DH, Brown GL. The investigation of nucleic acid secondary structure by means of chemical modification with a carbodiimide reagent. I. The reaction between *N*-cyclohexyl-*N'*-beta-(4-methylmorpholinium)ethylcarbodiimide and model nucleotides. *Biochemistry* 1969;8:2312–28.
- [51] Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 2005;127:4223–31.
- [52] Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, et al. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc Natl Acad Sci U S A* 2011;108:11063–8.
- [53] Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. RNA SHAPE analysis in living cells. *Nat Chem Biol* 2013;9:18–20.
- [54] Calvet JP, Pederson T. Heterogeneous nuclear RNA double-stranded regions probed in living HeLa cells by crosslinking with the psoralen derivative aminomethyltrioxsalen. *Proc Natl Acad Sci U S A* 1979;76:755–9.
- [55] Cimino GD, Gamper HB, Isaacs ST, Hearst JE. Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry. *Annu Rev Biochem* 1985;54:1151–93.
- [56] Breaker RR. Ancient, giant riboswitches at atomic resolution. *Nat Struct Mol Biol* 2012;19:1208–9.
- [57] Pan T, Sosnick T. RNA folding during transcription. *Annu Rev Biophys Biomol Struct* 2006;35:161–75.
- [58] Warf MB, Berglund JA. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci* 2010;35:169–78.
- [59] Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension. *Cell* 2009;136:719–30.
- [60] Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 2005;361:13–37.
- [61] Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* 2007;8:113–26.
- [62] Boyle J, Robillard GT, Kim SH. Sequential folding of transfer RNA. A nuclear magnetic resonance study of successively longer tRNA fragments with a common 5' end. *J Mol Biol* 1980;139:601–25.
- [63] Kramer FR, Mills DR. Secondary structure formation during RNA synthesis. *Nucleic Acids Res* 1981;9:5109–24.
- [64] Brehm SL, Cech TR. Fate of an intervening sequence ribonucleic acid: excision and cyclization of the *Tetrahymena* ribosomal ribonucleic acid intervening sequence *in vivo*. *Biochemistry* 1983;22:2390–7.
- [65] Pan J, Woodson SA. The effect of long-range loop-loop interactions on folding of the *Tetrahymena* self-splicing RNA. *J Mol Biol* 1999;294:955–65.
- [66] Heilman-Miller SL, Woodson SA. Effect of transcription on folding of the *Tetrahymena* ribozyme. *RNA* 2003;9:722–33.
- [67] Mahen EM, Harger JW, Calderon EM, Fedor MJ. Kinetics and thermodynamics make different contributions to RNA folding *in vitro* and in yeast. *Mol Cell* 2005;19:27–37.
- [68] Mahen EM, Watson PY, Cottrell JW, Fedor MJ. mRNA secondary structures fold sequentially but exchange rapidly *in vivo*. *PLoS Biol* 2010;8:e1000307.
- [69] Repsilber D, Wiese S, Rachen M, Schroder AW, Riesner D, Steger G. Formation of metastable RNA structures by sequential folding during transcription: time-resolved structural analysis of potato spindle tuber viroid (–)-stranded RNA by temperature-gradient gel electrophoresis. *RNA* 1999;5:574–84.
- [70] Yanofsky C. Attenuation in the control of expression of bacterial operons. *Nature* 1981;289:751–8.
- [71] Adilakshmi T, Soper SF, Woodson SA. Structural analysis of RNA in living cells by *in vivo* synchrotron X-ray footprinting. *Methods Enzymol* 2009;468:239–58.
- [72] Alexander JC, Pandit A, Bao G, Connolly D, Rochev Y. Monitoring mRNA in living cells in a 3D *in vitro* model using TAT-peptide linked molecular beacons. *Lab Chip* 2011;11:3908–14.
- [73] Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 2003;72:291–336.
- [74] Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* 2005;6:386–98.

- [75] Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell* 2009;136:701–18.
- [76] Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 2008;14:802–13.
- [77] Graveley BR. Mutually exclusive splicing of the insect *Dscam* pre-mRNA directed by competing intronic RNA secondary structures. *Cell* 2005;123:65–73.
- [78] Buratti E, Muro AF, Giombi M, Gherbassi D, Iaconcig A, Baralle FE. RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol Cell Biol* 2004;24:1387–400.
- [79] Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 2010;463:457–63.
- [80] McManus CJ, Graveley BR. RNA structure and the mechanisms of alternative splicing. *Curr Opin Genet Dev* 2011;21:373–9.
- [81] Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T. *N*(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* 2015;518:560–4.
- [82] Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010;141:129–41.
- [83] Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* 2007;131:174–87.
- [84] Holt CE, Bullock SL. Subcellular mRNA localization in animal cells and why it matters. *Science* 2009;326:1212–6.
- [85] Jambhekar A, Derisi JL. *Cis*-acting determinants of asymmetric, cytoplasmic RNA transport. *RNA* 2007;13:625–42.
- [86] Bullock SL, Ish-Horowitz D. Conserved signals and machinery for RNA transport in *Drosophila* oogenesis and embryogenesis. *Nature* 2001;414:611–6.
- [87] Chartrand P, Meng XH, Huttelmaier S, Donato D, Singer RH. Asymmetric sorting of Ash1p in yeast results from inhibition of translation by localization elements in the mRNA. *Mol Cell* 2002;10:1319–30.
- [88] Macdonald PM, Kerr K, Smith JL, Leask A. RNA regulatory element BLE1 directs the early steps of bicoid mRNA localization. *Development* 1993;118:1233–43.
- [89] Serano T, Cohen RS. A small predicted stem-loop structure mediates oocyte localization of *Drosophila K10* mRNA. *Development* 1995;121:3809–18.
- [90] Doyle F, Zaleski C, George AD, Stenson EK, Ricciardi A, Tenenbaum SA. Bioinformatic tools for studying post-transcriptional gene regulation: the UAlbany TUTR collection and other informatic resources. *Methods Mol Biol* 2008;419:39–52.
- [91] Rabani M, Kertesz M, Segal E. Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. *Proc Natl Acad Sci U S A* 2008;105:14885–90.
- [92] Palazzo AF, Springer M, Shibata Y, Lee CS, Dias AP, Rapoport TA. The signal sequence coding region promotes nuclear export of mRNA. *PLoS Biol* 2007;5:e322.
- [93] Kortmann J, Narberhaus F. Bacterial RNA thermometers: molecular zippers and switches. *Nat Rev Microbiol* 2012;10:255–65.
- [94] Chowdhury S, Maris C, Allain FH, Narberhaus F. Molecular basis for temperature sensing by an RNA thermometer. *EMBO J* 2006;25:2487–97.
- [95] Pelletier J, Sonenberg N. The involvement of mRNA secondary structure in protein synthesis. *Biochem Cell Biol* 1987;65:576–81.
- [96] Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 2009;324:255–8.
- [97] Hou J, Wang X, McShane E, Zauber H, Sun W, Selbach M, et al. Extensive allele-specific translational regulation in hybrid mice. *Mol Syst Biol* 2015;11:825.
- [98] Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 2009;324:218–23.
- [99] Shabalina SA, Ogurtsov AY, Spiridonov NA. A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res* 2006;34:2428–37.
- [100] Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess Jr JW, Swanson R, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 2009;460:711–6.
- [101] Wen JD, Lancaster L, Hodges C, Zeri AC, Yoshimura SH, Noller HF, et al. Following translation by single ribosomes one codon at a time. *Nature* 2008;452:598–603.
- [102] Nudler E, Mironov AS. The riboswitch control of bacterial metabolism. *Trends Biochem Sci* 2004;29:11–7.
- [103] Lusk JE, Williams RJ, Kennedy EP. Magnesium and the growth of *Escherichia coli*. *J Biol Chem* 1968;243:2618–24.
- [104] Ouyang Z, Snyder MP, Chang HY. SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res* 2013;23:377–87.
- [105] Aviran S, Trapnell C, Lucks JB, Mortimer SA, Luo S, Schroth GP, et al. Modeling and automation of sequencing-based characterization of RNA structure. *Proc Natl Acad Sci U S A* 2011;108:11069–74.
- [106] Zou C, Ouyang Z. Joint modeling of RNase footprint sequencing profiles for genome-wide inference of RNA structure. *Nucleic Acids Res* 2015;43:9187–97.
- [107] Selega A, Sirocchi C, Iosub I, Granneman S, Sanguinetti G. Robust statistical modeling improves sensitivity of high-throughput RNA structure probing experiments. *Nat Methods* 2017;14:83–9.
- [108] Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 2014;42:D92–7.
- [109] Hu B, Yang YT, Huang Y, Zhu Y, Lu ZJ. POSTAR: a platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res* 2017;45:D104–14.
- [110] Lempereur L, Nicoloso M, Riehl N, Ehresmann C, Ehresmann B, Bachellerie JP. Conformation of yeast 18S rRNA. Direct chemical probing of the 5' domain in ribosomal subunits and in deproteinized RNA by reverse transcriptase mapping of dimethyl sulfate-accessible. *Nucleic Acids Res* 1985;13:8339–57.
- [111] Tijerina P, Mohr S, Russell R. DMS footprinting of structured RNAs and RNA-protein complexes. *Nat Protoc* 2007;2:2608–23.
- [112] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–42.
- [113] Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 2013;10:584–90.
- [114] Bai XC, Fernandez IS, McMullan G, Scheres SH. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife* 2013;2:e00461.
- [115] Hang J, Wan R, Yan C, Shi Y. Structural basis of pre-mRNA splicing. *Science* 2015;349:1191–8.
- [116] Yan C, Hang J, Wan R, Huang M, Wong CC, Shi Y. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* 2015;349:1182–91.