Recent advances in RNA structurome

Bingbing Xu[1†], Yanda Zhu[1†], Changchang Cao[2†], Hao Chen[3†], Qiongli Jin[4†], Guangnan Li[5†], Junfeng Ma[6†], Siwy Ling Yang[7†], Jieyu Zhao[8†], Jianghui Zhu[9,10†], Yiliang Ding[11*], Xianyang Fang[6*], Yongfeng Jin[1*], Chun Kit Kwok[8,12*], Aiming Ren[3*], Yue Wan[7*], Zhiye Wang[4*], Yuanchao Xue[2,13*], Huakun Zhang[14*], Qiangfeng Cliff Zhang[9,10*] & Yu Zhou[5*]

[1]MOE Laboratory of Biosystems Homeostasis & Protection, Innovation Center for Cell Signaling Network, College of Life Sciences, Zhejiang University, Hangzhou 310058, China;
[2]Key Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China;
[3]Life Sciences Institute, Zhejiang University, Hangzhou 310058, China;
[4]State Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou 310058, China;
[5]State Key Laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan 430072, China;
[6]Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China;
[7]Stem Cell and Regenerative Biology, Genome Institute of Singapore, A*STAR, Singapore, Singapore;
[8]Department of Chemistry, and State Key Laboratory of Marine Pollution, City University of Hong Kong, Kowloon Tong, Hong Kong SAR, China;
[9]MOE Key Laboratory of Bioinformatics, Beijing Advanced Innovation Center for Structural Biology and Frontier Research Center for Biological Structure, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China;
[10]Tsinghua-Peking Center for Life Sciences, Beijing 100084, China;
[11]Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom;
[12]Shenzhen Research Institute of City University of Hong Kong, Shenzhen 518057, China;
[13]University of Chinese Academy of Sciences, Beijing 100101, China;
[14]Key Laboratory of Molecular Epigenetics of the Ministry of Education, Northeast Normal University, Changchun 130024, China

RNA structures are essential to support RNA functions and regulation in various biological processes. Recently, a range of novel technologies have been developed to decode genome-wide RNA structures and novel modes of functionality across a wide range of species. In this review, we summarize key strategies for probing the RNA structurome and discuss the pros and cons of representative technologies. In particular, these new technologies have been applied to dissect the structural landscape of the SARS-CoV-2 RNA genome. We also summarize the functionalities of RNA structures discovered in different regulatory layers —including RNA processing, transport, localization, and mRNA translation—across viruses, bacteria, animals, and plants. We review many versatile RNA structural elements in the context of different physiological and pathological processes (e.g., cell differentiation, stress response, and viral replication). Finally, we discuss future prospects for RNA structural studies to map the RNA structurome at higher resolution and at the single-molecule and single-cell level, and to decipher novel modes of RNA structures and functions for innovative applications.

RNA structurome, high-throughput techniques, genome-wide, 3D structure, RNA secondary structure, SARS-CoV-2, decoding, function

Introduction

RNA structure serves as important genetic information for RNA besides the nucleotide sequence that carries the coding information. Folding endows RNA molecules with catalytic, scaffolding, and perceptive functions. In 1965, Robert Holley and his collaborators published the nucleotide sequence of yeast alanine transfer RNA (tRNA) and proposed its possible secondary structure (Holley et al., 1965). In 1974, two groups independently acquired the 3 Å structure of tRNA by X-ray crystallography (Kim et al., 1974; Robertus et al., 1974), and scientists reached a consensus on a specific L-shaped tertiary structure of tRNA based on its cloverleaf secondary structure. In 1982, Tom Cech's group discovered that ribozymes could splice themselves from host RNA in the absence of protein, which was intrinsic to their RNA structure (Kruger et al., 1982). Over the ensuing decades, RNAs became increasingly recognized as functional macromolecules, leading to the proposal of the RNA world hypothesis.

RNA folding is considered a hierarchical stochastic process driven by the pairing (hydrogen bonding) of a subset of its bases, such as G-C, A-U, or G-U base pairs, to form various RNA secondary structures. Continuous base pairs, called RNA stems or helices, are rigid elements that can interact hierarchically to form complex higher-order three-dimensional (3D) structures with many recurrent tertiary motifs, such as the kink-turn. Moreover, RNA helices can be formed by short- or very long-range interactions, or even by crossing interactions (called pseudoknots). Although these complicated and dynamic features of RNA structure not only endow RNAs with their extensive functional versatility, they also make the algorithmic prediction of RNA structure quite challenging. For decades, researchers have attempted to decipher the structural code of RNA by computationally predicting the 3D structure of its primary sequences. However, this is a huge challenge and much more difficult than predicting the 3D structure of proteins, which has recently made breakthroughs based on deep learning techniques such as AlphaFold (Jumper et al., 2021). One of the difficulties is that the available training data for RNAs with known structures is very limited, especially for large and complex RNA structures.

Although RNA 3D structures are too complicated to be predicted *de novo*, researchers have developed many algorithms and programs to predict RNA secondary structure(s), which can be decomposed into RNA helices, bulges, hairpins, and multiple types of loops. Based on realistic energy functions with experimentally measured parameters, highly efficient algorithms have been implemented to model the RNA structure in terms of minimum free energy (MFE), with representative programs including RNAstructure (Mathews, 2014) and the ViennaRNA package (Lorenz et al., 2011), which have been widely used and continuously updated to date. Although these programs have achieved satisfactory performance on short RNAs less than 100 nt, accuracy becomes limited for long RNA sequences.

RNA secondary structure prediction can be improved by experimental data, and many programs, such as the Vienna RNA package, support computational folding of RNAs with hard or soft constraints by incorporating data from probing experiments. Previously, experimental evidence was often obtained from enzymatic cleavage-based RNA structure probing by running long polyacrylamide gel electrophoresis (PAGE) gel(s) on a single RNA, which was both time-consuming and labor-intensive. Subsequently, a revolutionary technique called PARS (parallel analysis of RNA structure) was developed by coupling high-throughput sequencing with RNA structure-specific RNase V1 and S1 nucleases to provide quantitative cleavage signals for thousands of RNAs simultaneously (Kertesz et al., 2010). Since then, many sequencing-based chemical probing approaches have been created to profile RNA structures *in vitro* and *in vivo* at single nucleotide resolution by detecting chemically-probe modified single-stranded sites, such as Dimethyl sulfate sequencing (DMS-seq) (Rouskin et al., 2014), Structure sequencing (Structure-seq) (Ding et al., 2014), selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-seq) (Lucks et al., 2011), *in vivo* click selective 2′-hydroxyl acylation and profiling experiment (icSHAPE) (Spitale et al., 2015), and azido-kethoxal (N_3-kethoxal) probing with deep sequencing (Keth-seq) (Weng et al., 2020). Meanwhile, RNA proximity ligation-based high-throughput profiling methods have been developed to identify genome-wide RNA-RNA interactions, such as psoralen analysis of RNA interactions and structures (PARIS) (Lu et al., 2016) and RNA *in situ* conformation sequencing (RIC-seq) technology (Cai et al., 2020). These methods have efficiently obtained large-scale base-resolution profiling of RNA structural information *in vivo*, thus improving the fidelity of RNA structure predictions. Recently, biophysical techniques such as cryo-electron microscopy (cryo-EM) (Kappel et al., 2020), atomic force microscopy (AFM) (Spokoini-Stern et al., 2020), and small angle scattering (SAS) (Zhang et al., 2019d) have promoted the direct modeling of RNA 3D structures.

In this review, we summarize recent advances in different techniques for probing the systematic mapping of intramolecular RNA structures and intermolecular RNA-RNA interactions across the transcriptome, referred to as the RNA structurome and interactome, respectively. Note that in some cases, for brevity, we use RNA structurome to represent both intramolecular RNA structures and intermolecular RNA-RNA interactions. We detail the many informative RNA structures found in a wide range of species from viruses, microbes, animals, and plants. In particular, these new technologies have been applied to explore the structural landscapes of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA genome. The functionalities of various RNA structural motifs, such as frameshifting elements and RNA G-quadruplex (rG4), are detailed by their characteristics, locations, and regulated molecular events. Finally, we discuss future technological development trends aimed at cracking the RNA structural code, such as deep learning and integrative methods, and indicate the detailed mechanism that may need to be studied to dissect how RNA secondary structure determines phenotypes for the development of RNA structure-based applications.

Global mapping of RNA secondary structures

Genome-wide mapping of RNA secondary structures

RNA structure is the basis for RNA function and regulation (Bevilacqua et al., 2016). Although biophysical technologies such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-EM are used to obtain RNA 3D structures, generally these methods study only one RNA molecule or complex at a time and are often limited by the crystallization ability of RNA molecules or the flexibility of RNA structures in solution (Leamy et al., 2016). The development of biochemical strategies, including enzymatic and chemical probes that can modify single-stranded and double-stranded bases as well as solvent-accessible bases along RNA, has also greatly expanded our capabilities to study RNA secondary structures in solution (Table 1). In the last decade, the field of RNA biology has witnessed the development of sequencing-based biochemical approaches that have enabled high-throughput profiling of RNA secondary structures (Li et al., 2020).

The first generation of RNA structure probing methods called PARS (Kertesz et al., 2010) uses the membrane impermeable enzymes RNase V1 and S1 nucleases, which recognize double- and single-stranded regions along RNAs, respectively. Upon cleaving at either the double- or single-stranded bases, these enzymes leave behind the 5′ phosphate (5′P) and 3′OH groups, allowing cleavage sites to be directly ligated to an adapter and captured. Since most RNA degraded fragments in the cell contain 5′OH, the 5′P leaving

groups allows accurate capture of cleavage events. Then, 3′ adapter ligation, reverse transcription (RT) and sequencing are performed to determine the location of double- and single-stranded bases in the transcriptome. PARS was first applied *in vitro* to study the RNA structure of the yeast transcriptome and was subsequently utilized to study the thermal stability of yeast RNAs across a series of temperatures. PARS was also applied under lysate conditions to study the human RNA structurome of a family of trios, whereby used to study the heritability of RNA structures passed from father and mother to child. Besides PARS, other enzymatic methods, including FragSeq, dsRNA-seq and ssRNA-seq using the enzymes nuclease P1, RNase V1 and RNase One, respectively, have also been used to probe the transcriptome-wide RNA structures in mice (Underwood et al., 2010), plants (Li et al., 2012b; Zheng et al., 2010), *Drosophila* and *Caenorhabditis elegans* (Li et al., 2012a). However, because these methods do not typically support probing RNA structure *in vivo* (Figure 1A), the way in which RNA folds inside cells remains elusive.

The development of membrane-permeable small molecules to modify the base or sugar-phosphate backbone of unpaired nucleic acids enables high-throughput structure probing of intracellular RNAs (Figure 1B and C). After structure probing and RNA extraction, the following reverse transcription step enables the introduction of stop or mutation sites for the modified sites. Computational pipelines can capture the stop or mismatch profiles of RNAs obtained from sequencing and directly yield RNA secondary structure information *in vivo* and on a genome-wide scale. Considering the cost and complexity of the experiments, most studies to date have used only one kind of small molecule reagents, each with its own advantages in terms of structure probing. Clearly, it seems that "parallel" studies using multiple reagents would enable a more comprehensive profile of RNA structure. We will elaborate on the details below.

DMS-seq (Rouskin et al., 2014) and Structure-seq (Ding et al., 2014) were among the earliest genome-wide approaches to probe the structure of RNA *in vivo*. Both methods use DMS, a small molecule that readily reacts with unpaired adenosine and cytidine on the Watson-Crick interface to modify single-stranded regions of cellular RNAs. Modified positions will block RT and generate a prematurely terminated complementary DNA (cDNA) library. The main difference between DMS-seq and Structure-seq is that DMS-seq requires ligation of priming adapters to fragment RNAs prior to PCR and sequencing, whereas Structure-seq uses random hexanucleotide sequences (N6) to directly prime RNA during RT, where cDNA is derived from and then ligated to DNA adapters. The implication here is that Structure-seq minimizes RNA degradation when constructing cDNA libraries. Based on the sequencing data, the detected 3′ end of the cDNA aligned to the transcriptome indicates the

Table 1   List of the global mapping strategies for RNA spatial conformation and interaction

Strategies	Approach	*In vivo* /*In vitro*	Chemical probe/ Ligand/Enzyme	Species	Reference	Pros/Cons
Enzymatic strategies	PARS	*In vitro*	RNase V1 RNase S1	Yeast, human cells	(Kertesz et al., 2010; Wan et al., 2014)	Pros: ssRNA and dsRNA information Cons: *In vitro* only
	FragSeq	*In vitro*	RNase P1	Mouse nuclear RNA	(Underwood et al., 2010)	Pros: ssRNA information Cons: *In vitro* only
	ssRNA- and dsRNA-seq	*In vitro*	RNaseV1 and RNase One	*Drosophila* cells, *C. elegans, Arabidopsis*	(Li et al., 2012a; Li et al., 2012b; Zheng et al., 2010)	Pros: ssRNA and dsRNA information Cons: *In vitro* only
	PIP-seq	*In vitro*	RNaseV1 and RNase One	HeLa cells, *Arabidopsis*	(Gosai et al., 2015; Silverman et al., 2014)	Pros: Profiles RNA secondary structure and protein binding sites concurrently Cons: *In vitro* only
Chemical probing strategies	SHAPE-Seq	*In vitro*	1M7	Bacteria	(Lucks et al., 2011)	Pros: First coupling of SHAPE compounds to high throughput sequencing Cons: *In vitro*, not transcriptome-wide
	CIRS-seq	*In vitro*	DMS, CMCT	Mouse cells	(Incarnato et al., 2014)	Pros: Probe four nucleotides Cons: *In vitro* only, no dsRNA information
	DMS-seq	*In vitro/in vivo*	DMS	Yeast, human cells, *E. coli*	(Rouskin et al., 2014; Zhang et al., 2018)	Pros: *In vivo* Cons: Only probe A and C, no dsRNA information
	Structure-seq	*In vivo*	DMS	*Arabidopsis*, rice	(Deng et al., 2018; Ding et al., 2015; Ding et al., 2014; Tack et al., 2020)	Pros: *In vivo* Cons: Only probe A and C, no dsRNA information
	Structure-seq2	*In vivo*	DMS	Rice	(Ritchey et al., 2017; Su et al., 2018)	Pros: *In vivo*, optimized structure-seq Cons: Only probe A and C, no dsRNA information
	icSHAPE	*In vivo*	NAI-N3	Mouse cells, human embryonic stem cells (hESCs) SARS-CoV-2	(Spitale et al., 2015; Sun et al., 2021a; Wang et al., 2021b)	Pros: *In vivo*, probe four nucleotides, high signal-to-noise ratio Cons: No dsRNA information
	SHAPE-MaP	*In vivo*	1M7, 2A3, NAI	HIV-1, HEK293, *E. coli*, SARS-CoV-2	(Manfredonia et al., 2020; Marinus et al., 2021; Siegfried et al., 2014; Yang et al., 2021)	Pros: *In vivo*, probe four nucleotides, analysis of low-abundance RNAs Cons: Low mutation ratio, high background, no dsRNA information, requires high dosage of chemical treatments that is very harmful to the living cells
	icSHAPE-MaP	*In vivo*	NAI-N3	HEK293T cells	(Luo et al., 2021)	Pros: *In vivo*, probe four nucleotides, high signal-to-noise ratio Cons: Low mutation ratio, no dsRNA information
	Dance-MAP	*In vivo*	DMS	Jurkat E6-1 cells, RPE-1 cells, HEK293T cells	(Olson et al., 2022)	Pros: *In vivo*, probe four nucleotides, structure probing of RNAs in multiple conformations Cons: Require high-abundance RNAs
	rG4-seq	*In vitro*	PDS	HeLa, HEK293T, *Arabidopsis, E. coli, P. aeruginosa,* plasmodium falciparum	(Dumetz et al., 2021; Kwok et al., 2016a; Shao et al., 2020; Yang et al., 2020b; Yeung et al., 2019)	Pros: Long flanking sequence (250–300 nt), without an input of known PQSs, high resolution. Cons: *In vitro* rG4 mapping only
	RT Stop profiling	*In vitro*	DMS	Mouse embryonic stem cells (mESCs)	(Guo and Bartel, 2016)	Pros: 60–80 nt RNA fragments were selected Cons: Should ideally be performed under single-hit kinetics conditions, CircLigase-mediated intramolecular ssDNA ligation has more bias, *in vitro* rG4 mapping only
	G4RP-seq	*In vivo/ in vitro*	BioTASQ	MCF7	(Yang et al., 2018)	Pros: Ligand-binding based *in vivo* genome-wide rG4 mapping method Cons: BioTASQ ligand may change the G4 landscape, the detection resolution is low
	DMS-MaPseq	*In vivo*	DMS	Yeast, human cells, *Drosophila, Arabidopsis, E. coli,* HIV-1, SARS-CoV-2	(Manfredonia et al., 2020; Tomezsko et al., 2020; Wang et al., 2018; Wang et al., 2019b; Wu and Bartel, 2017; Zhang et al., 2018; Zubradt et al., 2017)	Pros: *In vivo*, analysis of low-abundance RNAs, structure probing of RNAs in multiple conformations Cons: Only probe A and C, low mutation ratio, no dsRNA information
	CAP-STRUC-TURE-seq	*In vivo*	NAI	*Arabidopsis*	(Yang et al., 2020a)	Pros: Obtain the *in vivo* RNA structures of intact mRNAs, probe four nucleotides Cons: No dsRNA information
	Keth-seq	*In vivo*	N3-kethoxal	mESCs	(Weng et al., 2020)	Pros: *In vivo*, reversible labeling of unpaired G Cons: Only probe G, no dsRNA information
	SHALiPE-Seq	*In vivo/ in vitro*	NAI	*Arabidopsis, Oryza sativa*	(Yang et al., 2020a)	Pros: Genome-wide rG4 mapping *in vitro* and *in vivo* Cons: Dependent on the *in vitro* unfolding and folding SHALiPE standards for annotating *in vivo* status
	PORE-cupine	*In vivo*	NAI	hESCs, SARS-CoV-2 virus	(Aw et al., 2021; Yang et al., 2021)	Pros: *In vivo*, capture the RNA structures of isoforms, structural information of full-length transcripts Cons: No dsRNA information

(*Continued*)

Strategies	Approach	*In vivo* /*In vitro*	Chemical probe/ Ligand/Enzyme	Species	Reference	Pros/Cons
	CLASH	*In vivo*	UV crosslinking	Yeast, *C. elegans*	(Helwak et al., 2013; Kudla et al., 2011; Shen et al., 2018)	Pros: Genome-wide *in vivo* RBP-RNA and RNA-RNA interaction Cons: Profile a defined RBP-organized RNA duplexes only, require radioisotopes during library construction, ligation is performed in dilute solution, relies on overexpressed tagged protein
	hiCLIP	*In vivo*	UV crosslinking	Flp-In T-Rex 293 cells	(Sugimoto et al., 2017)	Pros: Genome-wide *in vivo* RBP-RNA and RNA-RNA interaction, high ligation efficiency Cons: Profile a defined RBP-organized RNA duplexes only, require radioisotopes during library construction, ligation is performed in dilute solution, requires high-quality antibody
	irCLASH	*In vivo*	UV crosslinking	HEK293 cells	(Song et al., 2020)	Pros: Genome-wide *in vivo* RBP-RNA and RNA-RNA interaction, high ligation efficiency Cons: Profile a defined RBP-organized RNA duplexes only, ligation is performed in dilute solution, relies on overexpressed tagged protein
	RPL	*In vivo*	Proximity ligation	Yeast, human cells	(Ramani et al., 2015)	Pros: Genome-wide *in vivo* RNA-RNA interactions Cons: limited to determine intra-molecular RNA-RNA interactions due to lack of cross-linking, ligation is performed in dilute solution
	PARIS	*In vivo*	AMT crosslinking	HeLa cells, HEK293 cells, mES cells	(Lu et al., 2016)	Pros: Genome-wide *in vivo* RNA-RNA interactions, near base-pair resolution Cons: Psoralen crosslinking has sequence bias, ligation is performed in dilute solution
Proximity ligation strategies	LIGR-seq	*In vivo*	AMT crosslinking	293T cells	(Sharma et al., 2016)	Pros: Genome-wide *in vivo* RNA-RNA interactions Cons: Psoralen crosslinking has sequence bias, ligation is performed in dilute solution
	SPLASH	*In vivo*	EZ-Link Psoralen-PEG3-Biotin crosslinking	HeLa cells, yeast, *E. coil*	(Aw et al., 2016)	Pros: Genome-wide *in vivo* RNA-RNA interactions, high signal-to-noise ratio Cons: Psoralen crosslinking has sequence bias, ligation is performed in dilute solution
	MARIO	*In vivo*	Dual crosslinking with EthylGlycol bis and formaldehyde	Mouse brain, mES cells	(Nguyen et al., 2016)	Pros: Genome-wide *in vivo* RNA-RNA interactions, only capture the RNAs co-bound with a single protein Cons: RNA duplexes do not associate with any protein are lost, dual crosslinking may introduce false-positive RNA-RNA interactions, ligation is performed in dilute solution
	RIPPLiT	*In vivo*	No crosslinking	HEK293 cells	(Metkar et al., 2018)	Pros: Proximal RNA ligation without any crosslinking treatment, genome-wide *in vivo* RBP-RNA and RNA-RNA interaction Cons: Profile a defined RBP-organized RNA duplexes only, ligation is performed in dilute solution, relies on overexpressed tagged protein
	COMRADES	*In vivo*	Azide-modified psoralen crosslinking	Zika virus, SARS-CoV-2 virus	(Ziv et al., 2018; Ziv et al., 2020)	Pros: Genome-wide *in vivo* RNA-RNA interaction of specific RNA Cons: Psoralen crosslinking has sequence bias, ligation is performed in dilute solution
	RIC-seq	*In vivo*	Formaldehyde crosslinking	HeLa cells, SARS-CoV-2 virus	(Cai et al., 2020; Cao et al., 2021b)	Pros: Genome-wide *in vivo* RNA-RNA interactions, high signal-to-noise ratio, *in situ* proximity ligation, high percentage of chimeric reads Cons: Formaldehyde crosslinking could link protein-protein, along with protein-RNA interactions

single-stranded region of RNAs. For both methods, control experiments must be performed during the modification step using a solvent lacking the modifying reagent so as to distinguish true positive signals from background signals (e.g., those caused by natural RT drop-offs). While DMS-seq and Structure-seq provide a comprehensive profile of genome-wide RNA secondary structure with *in vivo* single-nucleotide precision, it is worth noting that DMS reagents reacts only with unpaired A and C bases.

A recently developed approach termed Keth-seq (Weng et al., 2020) uses N_3-kethoxal to react with guanines on the Watson-Crick interface of single-stranded RNA under mild conditions, ultimately using a similar working principle to the RT-stop-based techniques described above to probe RNA structures *in vivo*. However, it is worth noting that kethoxal modification in guanines is reversible under alkaline or heating conditions. In practice, this means that Keth-seq (Weng et al., 2020) does not require separate preparation of a
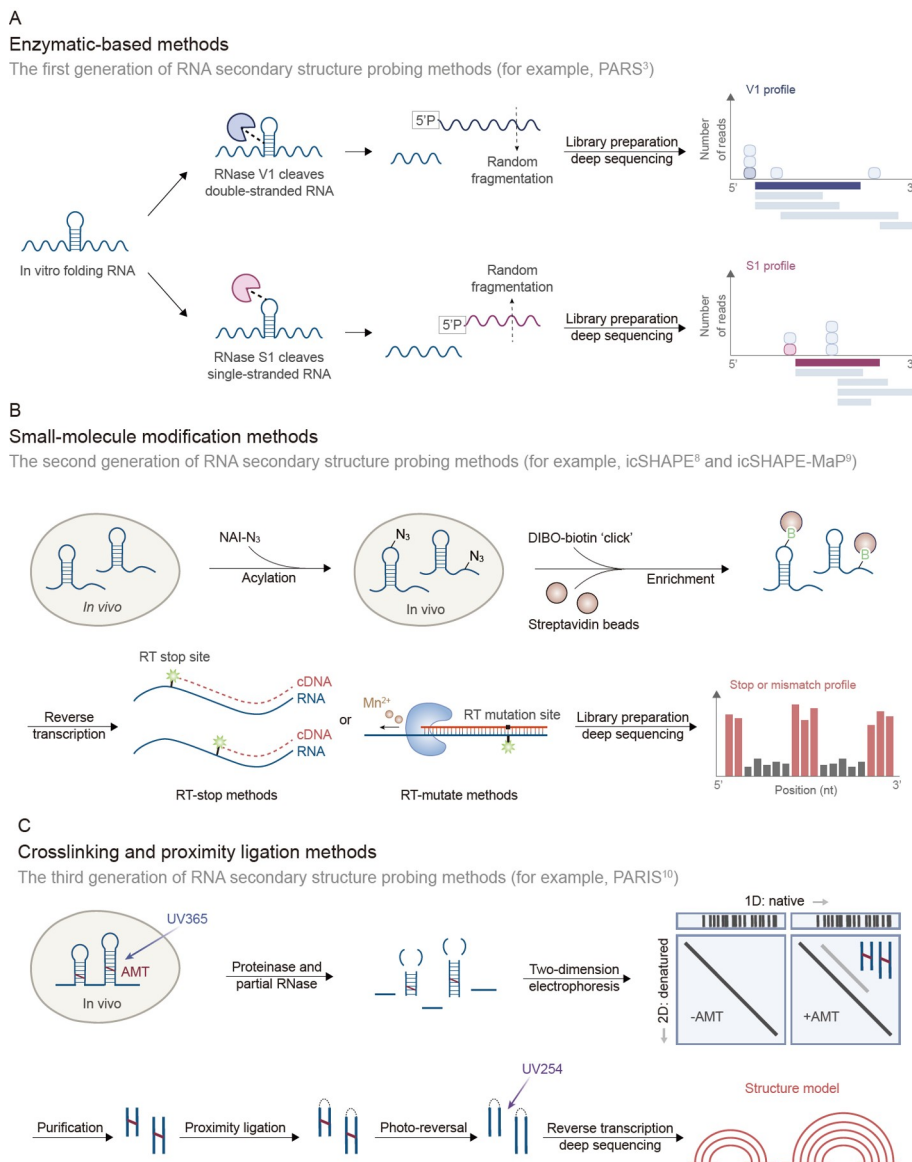
Figure 1 Schematic diagram of RNA secondary structure probing methods. A, Schematic for the enzymatic-based method, e.g., PARS. RNase V1 (purple) and S1 (pink) cleave double-stranded regions and single-stranded regions of RNA molecules *in vitro*, respectively, leaving a 5′P end. Due to limitations in sequencing depth, RNA molecules are randomly fragmented. The resulted fragments are sequenced and then aligned to the genome or transcriptome. Each aligned sequence provides structure information for a single base. The marked purple square or pink square illustrates structure information from one mapped sequence, while the light blue squares represent the mapping of additional sequences (light blue bars). B, Schematic for the small-molecule modification methods, e.g., icSHAPE and icSHAPE-MaP. RNA molecules are modified by NAI-N3 *in vivo*. The modified RNA molecules are attached to a biotin moiety via copper-free click chemistry and enriched using streptavidin beads. During reverse transcription, reverse transcriptase (e.g., Superscript III) stopped (more frequently in the Mg^{2+} buffer) or caused a mutation (more frequently in the Mn^{2+} buffer) at the modified position. After library preparation and deep sequencing, the profile of stop or mismatch sites obtained through computational pipelines provide information about RNA secondary structure *in vivo*. DIBO, dibenzocyclooxtyne. C, Schematic for the crosslinking and proximity ligation method, e.g., PARIS. RNA base pairs are fixed *in vivo* using AMT under UV 365 irradiation. After complete proteinase and partial RNase digestion, the RNA duplexes are purified using two-dimensional electrophoresis. Following proximity ligation and photo-reversal of crosslinks, the duplex RNA fragments are subject to deep sequencing and mapped to the genome. Each read provides evidence for direct base pairing between two RNA fragments. Dashed lines indicate ligations.

reagent-free control sample; the control can be simply produced by reversing the modifications. Furthermore, N_3-kethoxal also has higher labeling activity compared with DMS and can be efficiently permeated into living cells within 1 min.

The SHAPE family of compounds, including NMIA and 1M7, are small molecules that modify all four nucleotides by acylating the 2′-OH of the sugar-phosphate backbone in unpaired regions of RNA, thereby providing stop sites in RT (Spitale et al., 2013). To identify a compound with improved solubility and cell membrane penetrability properties, Chang and coworkers synthesized two new compounds, 2-methyl-nicotinic acid imidazolide (NAI) and 2-methyl-3-furoic acid imidazolide (FAI) (Spitale et al., 2013). These two com-

pounds can effectively modify RNAs from different organisms, and subsequent improvements in the library preparation protocol led to the fusion of the azido group with NAI to generate NAI-N$_3$, which can be attached to a biotin moiety via copper-free click chemistry. NAI-N$_3$-modified RNAs can be enriched by using streptavidin pull-down; this enrichment makes it more effective in probing RNA structures *in vivo*. This method is called icSHAPE (Spitale et al., 2015). A significant advantage of icSHAPE is the unbiased and efficient modification of all four bases in living cells.

All reverse transcriptase stop (RT-stop) methods share a common limitation: targeted RNAs typically lose 3′ terminus structural information due to loss of mapping of too short sequencing reads. In addition, only one of each cDNA molecule can be used as an RT stop site. As an improvement, other chemical probes such as 2A3 have higher reactivity with RNA and increased membrane permeability for *in vivo* probing, and new strategies such as mutational mapping (MaP) have been developed to map RNA structural modifications along RNA (Marinus et al., 2021; Wang et al., 2021b), DMS-MapSeq (Zubradt et al., 2017), and icSHAPE-MaP (Luo et al., 2021), which have recently been demonstrated to use RT mutations rather than RT stops for probing modification sites. These RT mutations are caused by the use of SuperScript II Reverse Transcriptase and Mn^{2+} buffer instead of SuperScript III and Mg^{2+}. A drawback of these RT-mutation methods though is that more in-depth sequencing is needed to support the bioinformatic analysis to reliably call a single-strand position, as random mutations may be incorporated in the library preparation process. Another benefit is that this method allows multiple mutations to co-exist along a single transcript. This generates a single cDNA molecule with multiple misincorporated nucleotides, each representing a distinct RNA structure signal, enabling single-molecule RNA structure analysis. Since RNAs rarely exist in a single conformation, structural heterogeneity along an RNA can be detected. Such studies have revealed structural heterogeneity in the HIV genome (Tomezsko et al., 2020) and the SARS-CoV-2 genome (Manfredonia et al., 2020). Very recently, a new RNA chemical probing strategy, DANCE-MaP, was used to determine dynamic RNA structural ensembles *in vivo*, and revealed that the human 7SK noncoding RNA encodes a structural switch in response to cell growth and stress (Olson et al., 2022).

An additional challenge of current high-throughput RNA structure probing strategies is that many transcriptomes, including the human transcriptome, are extensively alternatively spliced. As such, one gene can produce multiple transcripts (gene-linked isoforms) that can only vary by a short sequence between isoforms. Since some of these isoforms may be differentially regulated and show differences in translation and decay, it is difficult to assess the contribution of RNA structure in RNA regulation using short-read Illumina sequencing. To overcome this limitation, Wan and colleagues coupled NAI chemical probing with nanopore direct RNA sequencing to directly detect single-stranded positions along an RNA (Aw et al., 2021). Machine learning can be used to identify shifts in the current flowing through the pore when a base is modified as opposed to unmodified, thereby identifying multiple positions of single-stranded bases along a single RNA. Since nanopore sequencing allows for longer sequencing reads (the median length of sequencing reads for modified RNAs is ~750 bases compared with 2×150 bases in classical Illumina sequencing), this enables unique mapping reads to different gene-linked isoforms and the determination of isoform-specific structures. However, nanopore sequencing is still much more expensive compared with Illumina sequencing for achieving the same depth of sequencing to detect RNA structures.

While the above-mentioned RNA structure probing methods based on enzymatic cleavage or small molecule modification can effectively distinguish single- or double-stranded regions of RNA molecules, they cannot provide detailed RNA base-pairing information for these double-stranded regions. Since RNA base-pairing can occur between bases that are far apart on a linear sequence, e.g., more than 10 kb apart, in order to bring distant RNA regions into close physical space, the experimental data are needed to determine pairwise RNA base-pairing partners. To this end, many research groups have developed proximity ligation sequencing methods, including RPL, PARIS, LIGR-seq, SPLASH, and COMRADES (Aw et al., 2016; Lu et al., 2016; Ramani et al., 2015; Sharma et al., 2016; Ziv et al., 2018). These methods use variations of a reversible photo-crosslinkable compound psoralen to crosslink pairwise RNA interactions together, fragment the RNA, and then ligate the interacting RNA regions together prior to reverse cross-linking and library preparation. PARIS (Lu et al., 2016) (psoralen analysis of RNA interactions and structures) reveals RNA duplexes with a near base-pair resolution and simultaneously determines RNA interactions and RNA structure. This method crosslinks base pairs in living cells by using the crosslinker psoralen derivative, 4′-aminomethyl-trioxsalen (AMT). It then utilizes partial RNase digestion and two-dimensional gel electrophoresis to enrich cross-linked RNA duplexes. After proximity ligation of RNA duplexes and photo-reversal of crosslinks, RNA fragments are read by sequencing to obtain the RNA structurome and interactome. LIGR-seq also uses AMT for RNA crosslinking and enriches duplex fragments by using RNase R to degrade single-stranded fragments. In SPLASH, a biotinylated version of psoralen is used to enrich cross-linking interactions, while COMRADES uses psoralen azide to selectively enrich RNAs of interest via biotinylated probes. Psoralen azide cross-linked RNA fragments are then captured in later steps by adding biotin using click chemistry (Sharma et al., 2016;

Ziv et al., 2018). Collectively, these methods provide innovative and robust approaches for revealing higher-order and alternative RNA structures.

While RNA secondary structure information can be obtained experimentally by the above probing methods, many computational tools have recently been developed based on machine learning or deep learning to model or predict RNA secondary structures, including E2Efold (Chen et al., 2020), MXfold (Sato et al., 2021), SPOT-RNA2 (Singh et al., 2021), and UFold (Fu et al., 2022). In addition, icSHAPE data are used to better predict RNA-binding protein binding sites along RNAs. The development of the PRISMNET program extends RNA reactivity data beyond purely studying RNA structure to understand RNA function, i.e., how it interacts with RNA-binding proteins inside cells (Sun et al., 2021b).

Global mapping of RNA G-quadruplex structures

Guanine (G)-rich sequences can be assembled into two or more stacks of G-quartets (Figure 2A) and stabilized by Hoogsteen hydrogen bonding and monovalent cations to form a secondary structure commonly referred to as G-quadruplex (G4) (Figure 2B). The stabilizing effect of G4 by monovalent cations decreases in the following order: $K^+>Na^+>Li^+$ (Neidle and Balasubramanian, 2006). Canonical G4 structures can be bioinformatically predicted based on primary sequences (Figure 2B) (Huppert and Balasubramanian, 2005) and categorized into intra- and inter-molecular structures with multiple topologies such as parallel, antiparallel, and hybrid. Non-canonical G4 structures include 2-quartet, bulge, long loop, and G-triplex. Over the years, RNA G-quadruplexes (rG4s) have attracted great interest in the field due to their diverse molecular structures and multifaceted functions in cells (Lyu et al., 2021). G4s in mRNAs are involved in multiple functions, including transcription termination (Skourti-Stathaki et al., 2011), alternative splicing (Didiot et al., 2008), polyadenylation (Beaudoin and Perreault, 2013), mRNA localization and stability (Ishiguro et al., 2016; Zalfa et al., 2007), as well as the translation process (Lyu et al., 2019). Besides, G4s in non-coding RNAs are important for telomere homeostasis and genome stability (Gros et al., 2008; Pennarun et al., 2005), biogenesis of miRNAs (Rouleau et al., 2018), and piRNAs (Balaratnam et al., 2019; Zhang et al., 2019b). In addition to the functions reported in mammalian systems, rG4s are considered to play critical roles in many species, including but not limited to plants (Kwok et al., 2015; Yang et al., 2020b), bacteria (Shao et al., 2020) and viruses (Fleming et al., 2016), one of which is SARS-CoV-2 causing the current pandemic (Ji et al., 2021). Notably, multiple regulatory functions of rG4 structures have been associated with different human diseases, such as cancer and neurological disorders (Cammas and Millevoi, 2017).

The rapid development of high-throughput next-generation sequencing (NGS) technology has facilitated the possibility of global rG4 mapping across the transcriptome (Table 1). Broadly speaking, NGS-based methods for transcriptome-wide rG4 detection can be classified into *in vitro* and *in vivo* rG4 detection.

In 2016, Kwok et al. developed an *in vitro* rG4 sequencing (rG4-seq) technique that exploits reverse transcriptase stalling (RTS) of the rG4 structure followed by NGS readout (Figure 2C) (Kwok et al., 2016a). They first applied this rG4 profiling method to purified polyadenylated (poly(A))-enriched RNA extracted from HeLa cells. By taking advantage of the fact that the rG4 secondary structure motif can either fold in the presence of K^+ or G4-stabilized ligand PDS under K^+ (K^+/PDS), while unfolding when exposed to Li^+ (Kwok and Balasubramanian, 2015), RTS sites can be detected by rG4-seq due to the stops of reverse transcriptase upon encountering G4 under K^+ and K^+/PDS conditions during the reverse transcription step, while less RTS is detected under Li^+ conditions at the same position (Figure 2C). In this study, over 3,000 and 13,000 RT stop sites were detected under K^+ and K^+/PDS conditions in the human transcriptome, respectively, providing prevalent *in vitro* rG4 formation in the transcriptome. Moreover, PDS can stabilize rG4, leading to more RT stops compared with the condition where only K^+ is used. Recently, the experimental procedures of rG4-seq were further evaluated and optimized with lower RNA input, shorter library preparation time and lower transcriptome abundance variation (Yeung et al., 2019). The associated bioinformatics analysis tool, rG4-seeker, was developed to allow for better false-positive discrimination and improved sensitivity to non-canonical rG4s, enabling high-confidence identification of novel and non-canonical rG4 motifs from rG4-seq experiments (Chow et al., 2020). So far, rG4-seq has been applied to many other species (Dumetz et al., 2021; Shao et al., 2020; Yang et al., 2020b).

Roughly concurrent with the invention of rG4-seq, another *in vitro* method called RT-stop profiling was developed to detect rG4 structures in mouse embryonic stem cells (mESCs) (Guo and Bartel, 2016). Compared with the rG4-seq method that incorporates K^+/PDS as one of the positive G4 folding conditions, RT-stop profiling follows the principle that more RT stops can be detected due to rG4 formation under K^+ condition, while fewer RT stops under Na^+ and Li^+ conditions are due to fewer rG4 structures being folded as negative controls. Moreover, DMS probing was also incorporated into the RT-stop profiling method because the N_7 position of Guanine nucleotides can be methylated by DMS to inhibit rG4 formation under K^+ conditions. Despite the presence of K^+ during reverse transcription, DMS probing under denaturing conditions (95°C, 0 mmol L^{-1} K^+) substantially diminishes RT stops at G residues, resulting in less RT stops compared with the case without DMS treatment
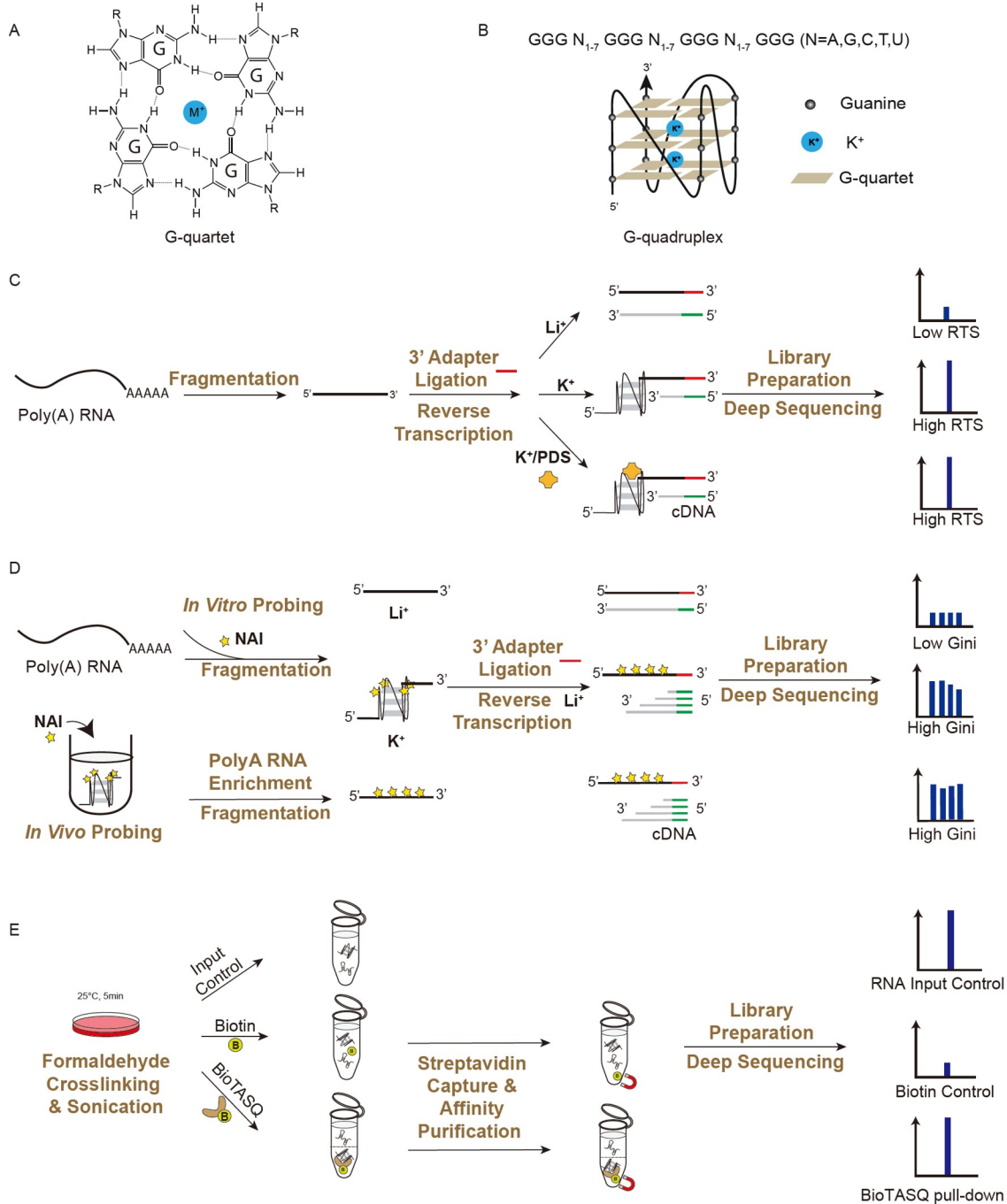
Figure 2 RNA G-quadruplex structure and representative rG4 global mapping techniques. A, Chemical structure of a G-quartet. M^+ represents monovalent cation such as K^+, which stabilizes the G4 formation. B, Canonical G4 sequence consensus and folded G4 structure. C, Experimental flowchart in detail of rG4 sequencing (rG4-seq). D, Selective 2′-hydroxyl acylation with lithium ion-based primer extension sequencing (SHALiPE-seq). E, G4-RNA-specific precipitation and sequencing (G4RP-seq). RNA and cDNA are labeled with black and grey line respectively; the 3′ adapter is labeled with a red line and the reverse primer is labeled with a green line.

under the same RT conditions. Some similarities and differences can be spotted between the two methods. Similar to rG4-seq, a few thousand strong RT stops were detected under K^+ conditions rather than Li^+ and Na^+ conditions, and both methods were developed to explore *in vitro* rG4 in the transcriptome. RT-stop profiling generated 60–80 nt RNA fragments for 3′ adapter ligation, while rG4-seq generated

around 250–300 nt of fragmented RNA to ensure that the flanking sequence is long enough to better cover and represent the rG4 structural subtypes, such as rG4s with long loops (Jodoin et al., 2014). In the study by Guo and Bartel (2016), the folding or unfolding of rG4 *in vivo* was also performed by coupling DMS-seq (Rouskin et al., 2014) and RT-stop profiling based on the DMS specificity of the

methylated N_7 position of G residues (m^7G). In this method, the *in vivo* folding state of rG4s were detected based on structural protection from DMS modification and refolding during reverse transcription, leading to strong RT stop signals. Whereas those residues within unfolded G4s in cells were modified by DMS and thus unable to form G4 during reverse transcription, resulting in different RT patterns from those of *in vivo* folded G4s. Overall, thousands of mammalian G-rich RNA regions have been reported to fold into G-quadruplex structures *in vitro*; however, most of these rG4-forming regions are shown to be unfolded in mammalian cells and yeast. Later, they also used NAI as well as RT-stop profiling and showed similar results in mESCs.

NAI was also used to probe RNA for rG4 mapping, followed by a primer extension reaction consisting of a home-made Li$^+$-containing reaction buffer, a method referred to as selective 2′-hydroxyl acylation with lithium ion-based primer extension (SHALiPE) (Kwok et al., 2016b). Since SHALiPE is a low-throughput method for individual rG4 mapping, it was exploited to develop SHALiPE sequencing (SHALiPE-Seq) to achieve *in vivo* global rG4 mapping (Yang et al., 2020b). This method is based on strong modifications of the last G in the rG4 sequence by NAI, which may lead to an RTS during reverse transcription and further detection by NGS (Figure 2D). In this study, *in vivo* SHALiPE profiles were compared with those of *in vitro* NAI probing profiles in the presence of K$^+$ and Li$^+$ conditions as a benchmark, which were evaluated to determine the folding state of the rG4 of interest *in vivo* (Figure 2D). Using SHALiPE-seq in plants, hundreds of rG4 structures were found to fold in both Arabidopsis and rice, showing for the first time that rG4 forms in living eukaryotic cells. Furthermore, individual rG4 structural candidates were later functionally characterized to play roles in translation and plant growth. Although this study somewhat showed opposite result from Guo and Bartel's study, the authors suggested that the discrepancies may be due to the difference in cellular environment in the studies. rG4s might be able to form in specific cell types and growth conditions according to their special biological function in human diseases (Kharel et al., 2020).

As described in the previous section, Keth-seq was established to map global RNA secondary structures *in vivo* (Weng et al., 2020). In this study, Keth-seq was successfully applied to both mESC and HeLa cell models for high-throughput RNA profiling, including the rG4 structure. By combining PDS treatment in cells, rG4 structures can also be detected *in vivo* by Keth-seq.

In contrast to RT footprinting methods, another high-throughput *in vivo* rG4 mapping method called G4-RNA-specific precipitation and sequencing (G4RP-seq) was developed based on the affinity of G4 ligand binding (Yang et al., 2018). This type of ligand binding method leverages a designed biotinylated TASQ ligand (BioTASQ), which binds to G4-RNA targets via end-quartet stacking due to its specificity for self-assembly into a synthetic G-quartet (Figure 2E). In brief, after formaldehyde crosslinking in human cancer cells, G4-containing RNAs were captured and enriched by the BioTASQ ligand, and then sequenced for target identification (Figure 2E). As such, *in vivo* rG4 detection was carried out by assessing changes in the relative abundance of G4-containing RNA transcripts enriched by BioTASQ relative to the corresponding input (Figure 2E). By using G4RP-seq, several hundreds of G4-RNAs were identified as being formed in the transcriptome within human cells, suggesting an alternative outcome to the study by Guo and Bartel (2016). They postulated that this may be caused by the dynamic formation of transiently folded rG4s in cells, and that these rG4 are then unwound by helicases or other mechanisms. In addition, by comparing the effects of two well-studied G4-stabilizing ligands, BRACO-19 (Moore et al., 2006) and RHPS4 (Salvati et al., 2007), on the G4-RNA folding landscape in the transcriptome followed by Bio-TASQ enrichment, they demonstrated that the G4 transcriptomic landscape, especially in long non-coding RNAs can be induced and changed by different G4 ligands. So far, G4RP-seq is the only transcriptome-wide rG4 mapping method that relies on ligand-based G4-RNA pull-down in cells. However, this method has its own limitations. First, TASQ itself is a G4 ligand, which may induce changes in the RNA landscape, although cross-linking prior to BioTASQ enrichment can minimize the effect. Second, BioTASQ prefers to target parallel rG4s(Karsisiotis et al., 2011), which may limit the capture of other types of G4-RNAs (Huang et al., 2014; Warner et al., 2014; Xiao et al., 2018). Moreover, the resolution of rG4 detection is restricted to the transcriptional level. In other words, G4RP-seq needs to be further optimized due to the inability to determine the detailed rG4 amount and location of each transcript.

To date, chemical probes applied for *in vivo* rG4 mapping have given us an initial glimpse into the rG4 structural landscape in different species. One of the limitations of this approach is that most of the findings so far have been on high-abundance RNAs. To explore rG4s in low-abundance transcripts, it is possible to use enrichment methods (Chu et al., 2012; Simon, 2013; Ziv et al., 2018) to enrich low-abundance transcripts (Baldwin et al., 2021) in order to minimize high-abundance transcripts that do not contain potential rG4s. It is also anticipated that newer and better chemicals will continue to be developed for rG4 structure probing. For example, besides BioTASQ, biotin may be attached to other G4 chemical probes (Einarson and Sen, 2017; Lat et al., 2020), or other chemical moieties such as azide and photo-crosslinker groups may be attached to the G4 probe (Zhang et al., 2021) for subsequent labeling reactions, or these handles may be attached to different classes of G4

binders, such as antibodies (Biffi et al., 2014), peptides (Zheng et al., 2020), and aptamers (Chan and Kwok, 2020; Umar and Kwok, 2020). So far, a dozen of rG4-binding proteins have been identified (Brázda et al., 2014; Kwok and Merrick, 2017) and shown to play key roles in many biological processes and diseases (Dumas et al., 2021; Fay et al., 2017b; Kharel et al., 2020). It is of great interest to ask about the importance of rG4 binding proteins on rG4 folding, for example, by determining variations in the rG4 structural landscape and rearrangement between rG4 and other secondary structures, with/without the protein of interest. From these results, individual rG4 candidates are likely to be nominated for comprehensive biochemical and genetic dissection to reveal the fascinating rG4 biology and associated molecular mechanisms. Recent studies suggest that rG4 formation may be favoured in specific subcellular compartments or membrane-less organelles, potentially linking it to stress conditions and disease (Asamitsu and Shioda, 2021; Falabella et al., 2019; Fay et al., 2017a; Herviou et al., 2020; Lyons et al., 2017; Sauer et al., 2019; Zhang et al., 2019c). Examination of rG4 structure profiles under physiological and stress conditions will help uncover which sets of rG4s are involved in gene regulation and RNA metabolism. Moreover, with recent technological advances in subcellular and membrane-less organelle isolation, it is possible to couple RNA structure probing with the purification of these organelles such as stress granules (Khong et al., 2018) and p-bodies (Hubstenberger et al., 2017), and use them for library preparation and NGS to uncover new rG4 biological insights.

Global mapping of RNA spatial conformations and interactions

RNA, especially non-coding RNAs, often requires the formation of intricate tertiary structures and interactions with their substrates to achieve various regulatory functions. Along with the rapid progress in experimental and computational methods for decoding RNA secondary structures, several high-throughput techniques have been invented to profile RNA spatial conformations and RNA-RNA interactions mediated by RNA binding proteins (RBPs) globally (Table 1). The first method is called CLASH (Cross-linking, Ligation, and Sequencing of Hybrids), in which proximally interacting RNA fragments are cross-linked to RBPs via ultraviolet (UV) light and then ligated *in vitro* for sequencing of the chimeric fragments (Kudla et al., 2011). After sequencing, the two interacted RNA fragments can be computationally recovered from the same chimeric reads (Figure 3A). CLASH has been successfully applied to detect snoRNA-rRNA interactions mediated by several RBPs (Nop1, Nop56, and Nop58) in yeast (Kudla et al., 2011), miRNA-mRNA interactions mediated by Ago1 proteins in human

cells (Helwak et al., 2013), and piRNA-mRNA interactions associated with the PRG-1-piRISC complex in *C. elegans* (Shen et al., 2018). These works have significantly contributed to our understanding of small non-coding RNA-mediated gene regulation in multiple organisms.

CLASH requires the use of radioisotope labeling to visualize and purify protein-RNA complexes, restricting its general application in many laboratories. Recently, by integrating an infrared-dye-conjugated and biotinylated ligation adaptor, Song et al. (2020) revised the CLASH protocol and developed a method called irCLASH to identify RNA secondary structures recognized by human ADAR proteins. As a major technological advancement, irCLASH enables shorter library construction time and higher percentage of chimeric reads compared with CLASH. Conceptually similar to CLASH and irCLASH, another method called hiCLIP (RNA hybrid and individual-nucleotide resolution ultraviolet cross-linking and immunoprecipitation) also uses UV to crosslink RBP-mediated RNA-RNA proximal interactions (Sugimoto et al., 2015). However, hiCLIP incorporates an additional adaptor between two different RNA fragments for proximity ligation, thereby reducing physical constraints, increasing ligation efficiency, and improving the assignment of duplex positions. The hiCLIP method has been successfully used to profile RNA duplexes associated with STAU1, a double-stranded RNA binding protein, and has revealed the prevalence of long-range duplexes in 3′ UTRs of mRNAs for modulating their splicing or degradation.

Unlike CLASH and hiCLIP, another biochemical method called RIPPLiT (RNA immunoprecipitation and proximity ligation in tandem) directly ligates proximal RNA fragments in purified EJC-containing ribonucleoproteins (RNPs) under native conditions (Metkar et al., 2018). RIPPLiT enables efficient mapping of RNA-RNA interactions mediated by a FLAG-tagged Magoh protein without any cross-linking reagent treatment. Analysis of RIPPLiT data shows that mRNAs tend to lack locus-specific structures, which, in contrast, predominate in non-coding RNAs. Polymer simulation using identified RNA proximity information unexpectedly deciphered a flexible rod-like conformation of mRNAs. More importantly, most translated mRNAs do not appear to form a closed-loop conformation, which challenges the traditional model of productive translation (Metkar et al., 2018). Specifically, the CLASH, irCLASH, hiCLIP, and RIPPLiT methods require ectopic expression of defined proteins in cell lines. Overexpressed proteins may interfere with native RNA-RNA interaction networks. Moreover, the adoption of proximity ligation in dilute solutions may potentially increase false positive rates.

Most, if not all, endogenous RNA-RNA interactions are mediated by RBPs. The methods mentioned above, CLASH, irCLASH, hiCLIP, and RIPPLiT, can only analyze a defined RBP-organized RNA duplex. Therefore, how to map all
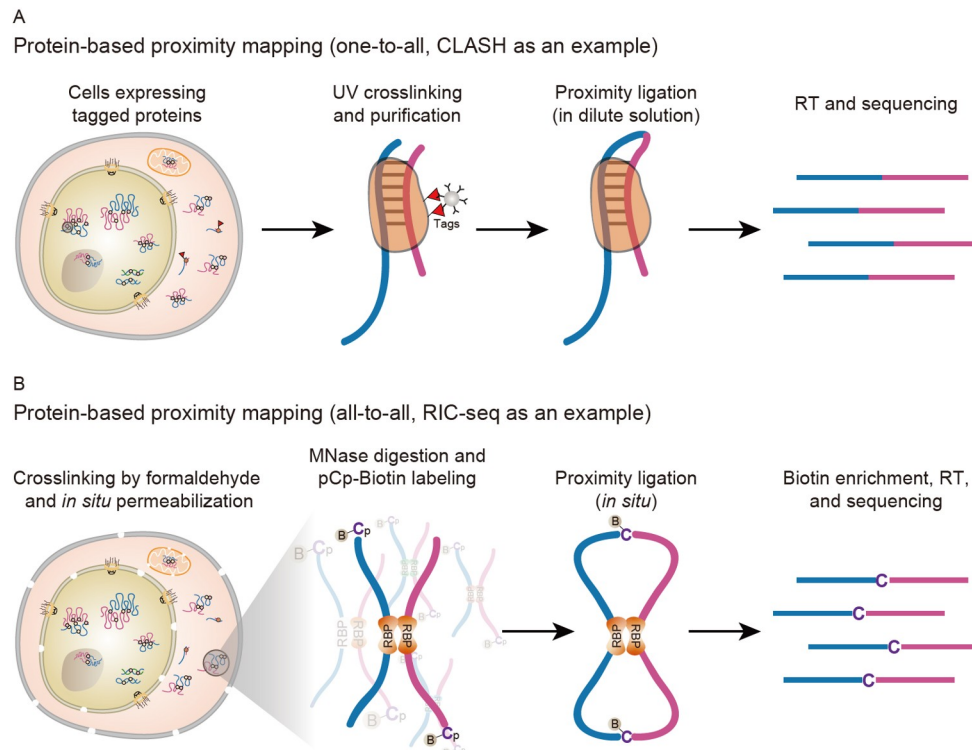
A
Protein-based proximity mapping (one-to-all, CLASH as an example)

Cells expressing tagged proteins · UV crosslinking and purification · Proximity ligation (in dilute solution) · RT and sequencing

B
Protein-based proximity mapping (all-to-all, RIC-seq as an example)

Crosslinking by formaldehyde and *in situ* permeabilization · MNase digestion and pCp-Biotin labeling · Proximity ligation (*in situ*) · Biotin enrichment, RT, and sequencing

Figure 3 RBP-based RNA spatial interaction mapping methods. A, Schematic diagram of CLASH method for detecting RNA duplexes mediated by a defined RBP. Cells expressing tagged proteins (e.g., HTA-tagged Ago1) are used for UV crosslinking. Base-paired RNAs in purified RNA-protein complexes are ligated in dilute solution and then sequence of the hybrids. B, Schematic illustration of RIC-seq technology for unbiased mapping of RNA-RNA proximal interactions in situ. Living cells are crosslinked by formaldehyde and permeabilized in situ. After MNase digestion, proximal RNA fragments are ligated in situ and then sequenced. Of note, the pCp-biotin is applied to label the junctures and enrich chimeric RNAs for sequencing.

RBP-mediated intra- and inter-molecular RNA-RNA interactions globally remains a major challenge. RPL (RNA proximity ligation) is the first method developed to simultaneously map the complete repertoire of RNA duplexes mediated by proteins (Ramani et al., 2015). RPL ligates proximally interacting RNA fragments under native conditions for subsequent sequencing of RNA hybrids. As a significant technological advance, RPL has shown good performance in mapping the secondary structure of many abundant RNAs such as ribosomal RNAs, snRNAs and snoRNAs. Nevertheless, unfortunately, RPL is primarily limited to determining intramolecular RNA-RNA interactions because of the omission of cross-linking steps. To identify intra- and inter-molecular RNA-RNA interactions simultaneously, Nguyen et al. (2016) developed the MARIO (mapping RNA interactome *in vivo*) method. They used two cross-linking reagents, formaldehyde and DSG, to fix and stabilize protein-mediated RNA duplexes. MARIO has revealed tens of thousands of different types of intra- and inter-molecular RNA-RNA interactions in the brain, embryonic fibroblasts and embryonic stem cells of mice. Nonetheless, proximity ligation in MARIO is performed in dilute solution, which may result in spurious ligation and cause high false-positive RNA-RNA interactions.

To address the above challenges, an RNA *in situ* con-

formational sequencing technology called RIC-seq has recently been developed to map intracellular RNA-RNA spatial interactions in an unbiased manner (Cai et al., 2020; Cao et al., 2021b). RIC-seq performs proximity ligation *in situ* and uses pCp-biotin to mark the junction of chimeric RNAs for enrichment, resulting in high efficiency, accuracy and minimal false positive rates when mapping RNA-RNA spatial interactions (Figure 3B). In addition to faithfully capturing RNA secondary and tertiary structures, RIC-seq data analysis enables the identification of pervasive RNA topological domains (i.e., isolated blocks with extensive interactions). The tremendous pairwise RNA-RNA interactions also enable the discovery of over 600 hub RNAs that show intense interactions with other RNA molecules. More importantly, RIC-seq identifies direct enhancer-promoter RNA interactions and demonstrates that these pairwise interactions can be used to assign the correct enhancer-promoter connectivity. Recently, a variant of RIC-seq called vRIC-seq (virion RNA *in situ* conformation sequencing) has been developed and shown to have great potential for probing the RNA spatial interactome within 60–80 nm virions. Notably, the proximal interactions revealed by vRIC-seq can be used for the first time to reconstruct the 3D conformation of the SARS-CoV-2 RNA genome (Cao et al., 2021a).

Despite the tremendous progress in experimental methods for globally mapping protein-associated RNA spatial interactions, several challenges remain to be tackled in the future. First, the overall percentage of useful data (i.e., chimeric reads) remains low, with a maximum percentage of about 16%, which is achieved by the RIC-seq method. Hence, further titration and optimization of various experimental steps will be required in the near future. Second, the methods mentioned above have limited coverage of low-abundance RNAs, for which an antisense oligo-based pull-down approach should help to enrich specific RNA interactions. Third, sophisticated software that is compatible with all available experimental methods and can accurately identify chimeric reads, distinguish high-confidence interactions from random noise, and screen differentially interacting pairwise RNAs between distinct conditions is still lacking. Fourth, it is challenging to convert the RNA spatial conformations obtained by these experimental methods into high-resolution tertiary structures. However, as the number of solved high-resolution RNA structures increases, artificial intelligence and deep learning are hopeful for solving this profound problem.

3D structure determination of RNA

It is now increasingly recognized that higher-order tertiary and quaternary structures (rather than just secondary structures) are keys to biological functions. However, our knowledge about the higher order structures of RNA is very limited. As of October 8, 2021, a total of 182,949 structures were deposited in the Protein Data Bank (PDB, http://www.rcsb.org). The majority are proteins (87.4%), but RNA-only and RNA-containing complex structures account for only 0.9% and 2.1%, respectively. Given the functional importance of RNA and the scarcity of experimentally determined RNA 3D structures, alternative biophysical, biochemical, and computational approaches have been developed and applied to obtain RNA 3D structural information (Table 2).

Biophysical approaches for RNA 3D structure determination

A number of biophysical techniques can be used to determine the overall 3D structure of RNA. Depending on the biophysical technique used, high-resolution 3D atomic structures or intermediate-to-low resolution 3D shapes have been obtained (Figure 4A–C).

X-ray crystallography (XRC) is the most widely used technique for structure determination of biomacromolecules at atomic resolution (Garman, 2014). Indeed, approximately 88% of the structures in PDB (http://www.rcsb.org) are determined using XRC. However, there are still many obstacles

Table 2 List of approaches to determine 3D RNA structures

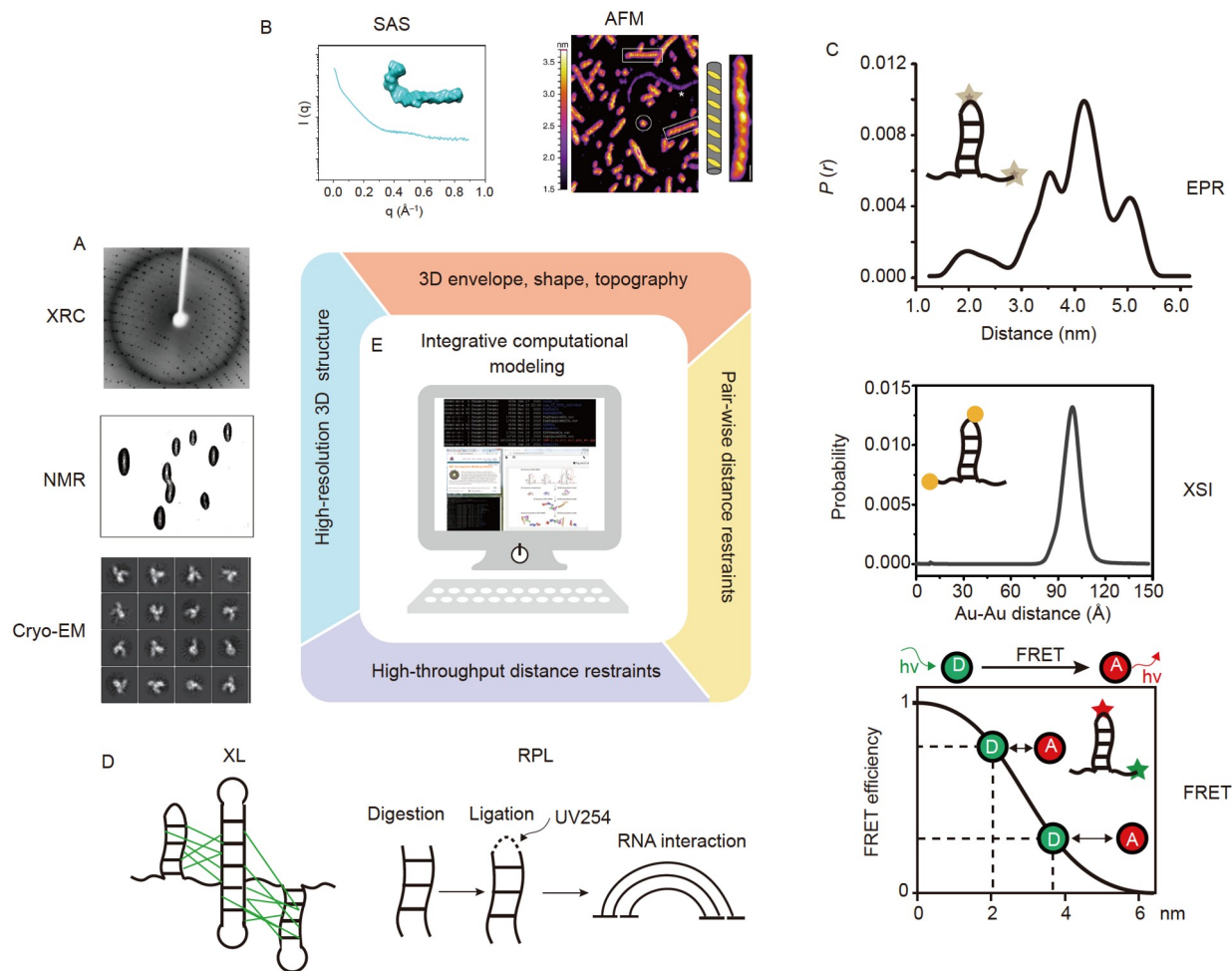| Strategies	Approach	Description	Pros	Cons
Biophysical approaches	XRC (Garman, 2014)	The most widely used structural technique for RNA and RNA-protein complex	High resolution structure information	Need well-diffracting crystal; viable for well-folded and rigid RNAs; phase problem
	NMR (Barnwal et al., 2017)	Another common tool for RNA structure determinination	Both structure and dynamics information; solution technique	Limited to smaller RNAs.
	cryo-EM (Carroni and Saibil, 2016; Danev et al., 2019; Nogales, 2016; Yang et al., 2022)	Applied on large biomolecules embedded in vitreous water	Only limited sample is required; heterogeneity is allowed	Hard to obtain high-resolution structures
	SAS (Jacques and Trewhella, 2010)	Including SAXS and SANS in solution	Providing global and dynamic structural information in solution; fast and time-resolved processes	Low-resolution map or structure
	AFM (Engel and Müller, 2000; Schön, 2016)	A type of scanning probe microscopy	Studying the structure and dynamics in solution	Intermediate resolution
	EPR (Duss et al., 2014)	Electron spin instead of the spin of atomic nuclei being excited	No size restrictions; Providing distance distribution information	Need efficient labeling of spins; Limited distance range
	XSI (Shi et al., 2015; Zettl et al., 2016)	Based on SAXS measurements of gold nanoparticle labeled biomolecules	Providing high-resolution distance distributions	Need AuNPs labeled at specific sites
	smFRET (Dimura et al., 2016; Helm et al., 2009)	Studying the structure and dynamics	Providing distance restraints; single-molecule analysis	Need two fluorophores for orthogonal labeling; Limited distance distributions
Biochemical approaches	XL (Harris and Christian, 2009)	Analyzing macromolecular interactions	Providing a range of through-space restraints	Need to use radioisotope labeling or other bireactive reagent
	RPL (Kudla et al., 2011)	Recording the physical proximity of two nucleic acids	Mapping intra-molecular RNA-RNA spatial interactions	Limited to intra-molecular RNA-RNA interactions

Figure 4 A variety of experimental and computational approaches that can be used to obtain RNA 3D structural information. A, Among the biophysical approaches, XRC, NMR, and cryo-EM can determine high-resolution 3D atomic structure for RNA. B, SAS, and AFM can provide intermediate-to-low resolution 3D envelope, shape or topography structure for RNA. C, EPR, XSI, and smFRET are three molecular rulers that can provide pair-wise distance distributions for RNAs in the nanometer range, where the RNAs are site-specifically labeled with spin label, gold nanoparticle, or fluorescent tag, respectively. D, XL, and RPL are high throughput biochemical approaches to identify potential regions of intra- and inter-molecular interactions in RNAs, which can be transformed into structural restraints for direct structural modeling. E, Structural information from different biophysical, biochemical, and computational approaches can be combined to determine RNA structure through integrative computational modeling.

to overcome in macromolecular XRC. One of the most significant bottlenecks is the need to obtain high-quality single crystals for X-ray diffraction experiments. The growth of high-quality single crystals requires the preparation of homogenous biomacromolecules in the same states. Since RNA is relatively dynamic, the preparation of RNA samples for crystallization is more sophisticated. Solid and extensive biochemical experiments on RNA molecules should be performed before designing new RNA constructs for crystallization. Studies of highly conserved RNA motifs such as riboswitches and ribozymes have provided many successful XRC examples for the structural determination of RNA molecules. Besides, some RNA molecules can also be crystallized in complex with protein binding partners. Currently, approximately 62% of the 1,569 RNA-only structures and 61% of the 3,946 RNA-protein complex structures have

been determined using XRC in the PDB. In most cases, these RNA molecules with resolved structures are less than 200 nt and adopt a relatively compact scaffold. Since XRC is not limited to the size of biomacromolecules, it remains a powerful technique in RNA 3D structure determination and can be used for a wider range of RNA molecules. However, for many RNA molecules, particularly larger ones, they are inherently flexible and can sample multiple conformations, making it difficult to crystallize (Reyes et al., 2009) or even be crystallized but diffracted at a lower resolution. To overcome these problems, more effort is needed to resolve the relationships between multiple conformations of dynamic large RNA molecules and to optimize the conditions for catching individual stable homogenous conformations before XRC structural studies can begin.

NMR spectroscopy is another common tool for de-

termining the structure of RNA at atomic resolution (Barnwal et al., 2017). Since many biologically important RNAs have dynamic conformations that prevent them from being crystallized, they are best studied using solution techniques. Compared with proteins, solution NMR spectroscopy has made a greater contribution to RNA 3D structure determination, using solution NMR to resolve roughly 35% of RNA-only structures but only 7% of protein-only structures in PDB (http://www.rcsb.org). Compared with proteins, NMR of RNA remains challenging because the proton density is much smaller and the chemical diversity of its building blocks—the four nucleotides—is much reduced, resulting in severe signal overlap in NMR spectra (Barnwal et al., 2017). These challenges become increasingly severe for larger RNAs, which tumble more slowly in solution, leading to severe broadening of signals and significant loss of signal intensity. As a consequence, NMR spectroscopy is generally limited to smaller RNAs. To facilitate structure determination of larger RNAs using solution NMR, numerous advances have been made over the past decades, including NMR data acquisition and RNA sample preparation, such as nucleotide-specific and segmental isotope labeling strategies, and divide-and-conquer approaches (Barnwal et al., 2017).

Cryo-EM has become a powerful technique for structure determination of large biomolecules and biomolecular machines (Carroni and Saibil, 2016; Danev et al., 2019; Nogales, 2016). Cryo-EM can be applied on samples cooled to cryogenic temperatures and embedded in an environment of vitreous water. Advances in detector technology and imaging processing algorithms have drastically improved the resolution of structures determined by cryo-EM, resulting in an explosive growth in the number of structures resolved by cryo-EM, especially those recently obtained at near-atomic resolution over the past five years. In comparison to other methods, cryo-EM has several important advantages, requiring only a minimal amount of sample, without the need for crystallization, and allowing the sorting of single-particle cryo-EM data from extremely heterogeneous samples during image classification. However, to date, cryo-EM has been applied mainly to protein-only or protein-nucleic acid complexes (e.g., ribosomes and spliceosomes). In fact, about 35% of the RNA-protein complex structures in the PDB can be solved by cryo-EM, and most of them (about 70%) have a resolution better than 4.0 Å. In contrast, there are a few RNA-only structures in the PDB derived from Cryo-EM data, and only one of these has a resolution lower than 4.0 Å (Zhang et al., 2019a). It is likely that the high intrinsic flexibility of pure RNA makes it difficult to obtain high-resolution structures by cryo-EM.

Despite the remarkable progress in cryo-EM, a substantial fraction of density maps is still determined at intermediate or low resolution (worse than 4 Å) (Zhang et al., 2020a). To extract structural information from these low-resolution density maps, advanced computational methods have been developed to create and refine atomic-level structural models (Zhang et al., 2020a). Recently, a deep learning-based method named Emap2sec+ has been developed to detect both protein secondary structure elements and nucleic acids in medium-resolution EM maps (Wang et al., 2021a). A software called auto-DRRAFTER has been developed to facilitate the building of RNA coordinates into low-resolution cryo-EM densities starting from an RNA sequence and secondary structure. Advances in cryo-EM, multidimensional chemical mapping, and Rosetta DRRAFTER computational modeling have been combined to form the Ribosolve pipeline that is expected to accelerate the cryo-EM-guided determination of 3D RNA-only structures (Kappel et al., 2020).

SAS is a collective term for several biophysical techniques, including small-angle X-ray (SAXS) and neutron (SANS) scattering (Jacques and Trewhella, 2010). The basic principles of SAXS and SANS are similar in that both techniques involve illuminating samples of biomolecules with an incident beam of radiation and recording the intensity and angle of the scattered beam to yield a one-dimensional scattering curve containing a wealth of information about the particle in solution. SAXS and SANS offer several advantages in the structural characterization of biomolecules, including providing unique structural information such as molecular mass, radius of gyration, aggregation state, folding state and flexibility, domain organization and overall shape, without the need for crystallization, and applicable to a wide range of molecular sizes and varying solution conditions (Brosey and Tainer, 2019). SANS and SAXS are complementary techniques that, when combined, allow for detailed structural characterization of biomolecules. Using contrast variation, SANS, in combination with deuterium labeling, is able to define the positions of specific components within a complex, while SAXS provides more precise data on the overall shape (Mahieu and Gabel, 2018). So far, SAXS has been widely used for direct structural analysis of RNAs (Fang et al., 2015), including the HIV-1 Rev response element (Fang et al., 2013), the T-box riboswitch core (Fang et al., 2017), and flaviviral subgenomic RNAs (Zhang et al., 2019d), or used as constraints for comprehensive computational modeling of RNA 3D structures (Wang et al., 2009).

AFM is a type of scanning probe microscopy that can be used to study the structure and dynamics of biomolecules at nanometer resolution (Bose et al., 2018; Engel and Müller, 2000; Schön, 2016). AFM offers several advantages over electron microscopy-based techniques in the study of biomacromolecules, including its ability to operate in liquid environments, thereby allowing imaging of biomolecular samples under physiologically relevant conditions, and

minimal requirements for sample preparation (no fixing, staining, coating, or labeling), thus minimizing structural disruption of the biological samples under study. Since AFM images of specimen surfaces have intermediate resolution and are not sufficiently robust to construct an atomic structure model from scratch, both rigid-body and flexible fitting methods have been developed to obtain molecular structures that are suitable for AFM images (Chaves and Pellequer, 2013; Niina et al., 2020; Trinh et al., 2012). Over the past 30 years, AFM has matured to offer 3D surface topography at nanometer spatial resolution for RNA molecules of different lengths and structures as well as for RNA-RNA or RNA-protein complexes (Schön, 2018). There are several examples of AFM applications for RNA structural characterization, including Mg^{2+}-induced conformational switch in the 5′-untranslated region of the hepatitis C virus genomic RNA (García-Sacristán et al., 2015), the structure of the HIV-1 Rev response element (RRE) itself and complexation with the virion regulator (Rev) (Pallesen et al., 2009), the native structures of single viroid RNA molecules (Moreno et al., 2019), the structure and motion of the long non-coding RNA HOTAIR under nucleus-like conditions (Spokoini-Stern et al., 2020).

Electron paramagnetic resonance (EPR) spectroscopy, X-ray scattering interferometry (XSI) and single-molecule Förster resonance energy transfer (smFRET) are three molecular rulers that provide pairwise inter- or intra-molecular distance distributions in the nanometer range (Figure 4C), thus suitable to probe the structure and dynamics of biomolecules under biologically relevant conditions.

The principle of EPR is generally similar to that of NMR, with the difference that the electron spins are excited instead of the spins of atomic nuclei. The advantages of EPR over NMR include higher sensitivity, no restrictions on the size of biomolecules, and the need for only a small amount of material. Given that most biomolecules are diamagnetic, strategies for efficient and site-specific labeling of spins such as nitroxide into biomolecules have been developed. On RNAs, several different strategies for covalent attachment of spin labels are described. With short RNAs, spin-labeled RNA oligonucleotides can be prepared by solid-phase chemical synthesis, or by using spin-labeled phosphoramidite nucleotides during synthesis, or by post-synthetic spin-labeling of pre-functionalized specific nucleotides by chemical reactions via Sonogashira cross-coupling, Click chemistry, or a photolabel protection group (Jakobsen et al., 2010; Kerzhner et al., 2016; Weinrich et al., 2018). However, solid-phase synthesis is generally limited to short RNAs of less than 100 nucleotides. Longer spin-labeled RNAs have been achieved by ligation techniques (Kerzhner et al., 2018) or, more recently, by post- (Wang et al., 2020b) or co- (Domnick et al., 2020) transcriptional spin labeling using the NaM-TPT3 unnatural base pairs (UBPs) system, which opens up the

possibility of investigating long RNAs by EPR spectroscopy. While continuous-wave EPR measurements provide the distance constraints of approximately 20 Å, pulsed EPR methods, such as pulsed electron-electron double resonance (PELDOR or DEER) spectroscopy, are capable of measuring long-range pairwise distance distributions between nitroxides up to 80 Å. Many studies have shown that EPR-measured distance constraints, coupled with NMR and computational modeling, provide a wealth of information in studying the global structure and conformational dynamics of RNA and RNA-protein complexes (Duss et al., 2014).

XSI is an emerging molecular ruler technique based on SAXS measurements of biomolecules labeled with pairs of gold nanoparticles (AuNPs) (Shi et al., 2015; Zettl et al., 2016). XSI can provide high-resolution label-label distance distributions ranging from 50 to 400 Å in a model-independent manner (Mathew-Fenn et al., 2008), and can therefore be used to quantify ensembles of macromolecular structures and correlate directly with three-dimensional structures (Shi et al., 2015). In the past few years, XSI has been fruitfully used in nucleic acids and nucleic acid/protein complex studies (Zettl et al., 2018), such as probing the structure and conformational changes of DNA origami, DNA, or DNA-protein complexes (Hura et al., 2013; Mathew-Fenn et al., 2008), and conformational landscapes of a complex RNA motif in response to changes in solution condition and protein binding (Shi et al., 2016; Shi et al., 2017). A prerequisite for the successful application of XSI is the preparation of biomolecules conjugated to a single and/or a pair of AuNPs at specific sites (Shi et al., 2015; Zettl et al., 2018). Recently, an efficient and universally applicable strategy for site-specific nanoparticle labeling of large RNAs has been developed, which is authorized by transcription of an expanded genetic alphabet containing TPT3-NaM UBP. This strategy overcomes the size constraints of conventional RNA labeling methods, thus implementing XSI for large RNA structural studies (Wang et al., 2020a). Because XSI can measure much broader distance distributions (ranging from 50 to 400 Å) than other molecular rulers such as EPR and smFRET, it is expected that XSI will play important roles in structural studies of large RNAs and RNA complexes.

smFRET is a popular and versatile technique that can be used to study the structure, dynamics and interactions of biomolecules, including RNAs, at both the ensemble and single-molecule levels (Dimura et al., 2016; Helm et al., 2009). FRET makes use of the mechanism of "Förster Resonance Energy Transfer" to transfer energy from donor to acceptor fluorophores (Roy et al., 2008). Since RNA has low intrinsic fluorescence, a prerequisite for FRET analysis is the specific labeling of RNA sites with two fluorophores. In recent years, a plethora of chemical, enzymatic and UBP-based labeling methods have been developed to achieve site-

specific fluorescent labelling of RNAs, in many cases under near-native conditions (Niu et al., 2021; Paredes et al., 2011). To be applicable as a FRET pair, the emission spectrum of the donor and the excitation spectrum of the acceptor should overlap. As the distance between donor and acceptor is within 10–80 Å, energy transfer can occur via FRET, and the efficiency of FRET is inversely proportional to the sixth power of the distance between the donor and acceptor. Therefore, FRET can be applied to measure intra- or inter-molecular distance as a sensitive spectroscopic ruler, which can be used to track conformational changes or distance restraints for integrative structural modeling (Dimura et al., 2016). In particular, smFRET has become a powerful technique widely applied to study the conformational dynamics and interactions of RNAs (Helm et al., 2009). FRET-derived distance restraints combined with computational modeling have been utilized to define the 3D structures of large RNAs (Stephenson et al., 2016; Stephenson et al., 2013). Recently, a freely accessible software suite was developed to demonstrate the possibility of determining protein structures based on FRET-assisted coarse-grained structural modeling (Dimura et al., 2020). It is expected that this FRET-assisted structure modeling approach can be extended to RNA 3D structure determination.

Biochemical approaches to probe RNA 3D structure

Biochemical approaches such as chemical and enzymatic probing experiments have been used to determine RNA structure (Kertesz et al., 2010; Mailler et al., 2019). In the past decades, a wide variety of probing reagents and experimental pipelines have been constantly improved to probe the structures of RNAs *in vitro* and in cells. Both secondary and tertiary structural information can be obtained from such biochemical approaches (Christy et al., 2021; Harris and Christian, 2009; Homan et al., 2014; Kudla et al., 2020; Ramani et al., 2015), which can be transformed into structural restraints for direct structural modeling (Figure 4D).

Chemical cross-linking (XL) is a well-established method for analyzing macromolecular interactions, which has been widely used in the study of proteins and nucleic acids (Harris and Christian, 2009). For the study of RNA structure and interactions, crosslinking is generally achieved using UV light to induce the formation of covalent bonds between proximal unmodified RNAs or between RNA and a photo-affinity reagent incorporated randomly or at specific sites in the RNA structure. Both long-range (>9 Å) and short-range (~3 Å) photoaffinity agents have been utilized for RNA photo-crosslinking studies. The cross-linked products can be isolated, followed by mapping of the cross-linked sites and assessment of the crosslinking data (Harris and Christian, 2009). Crosslinking sites can be identified by alkaline hydrolysis, primer extension, PAGE or RNA sequencing techniques. The power of crosslinking lies in the ability to rapidly identify potential regions of intra- and inter-molecular interactions and to generate a range of spatial restraints that can be used for higher-order structural modeling. An alternative approach is to employ bifunctional crosslinking reagents to chemically link the proximal nucleotides in an RNA structure. More recently, a strategy called SHAPE-JuMP has been introduced for monitoring RNA tertiary structures by means of the bi-reactive SHAPE reagent, *trans*-bis-isatoic anhydride (TBIA), which reacts preferentially with the 2′-hydroxyl groups of two proximal nucleotides. Distance restraints derived from SHAPE-JuMP can be used to model the global folding of large RNAs (Christy et al., 2021).

RPL uses the proximity ligation of native RNA followed by deep sequencing to yield chimeric reads with ligation junctions in the vicinity of structurally proximate bases (Kudla et al., 2011). Proximity ligation records the physical proximity of two nucleic acid termini via their ligation and has been applied to detect DNA aptamer-bound proteins, as well as for targeted or global chromosome conformation capture (3C) (Ramani et al., 2015). Similar to 3C methods for DNA conformation, RIC-seq has recently been recently developed to map intra- and inter-molecular RNA-RNA spatial interactions within cells in an unbiased manner (Figure 3B), whereas RPL uses digestion and re-ligation of RNA but omits crosslinking, relying instead on the inherent spatial proximity of intra-molecular RNA nucleobases in secondary structural features (i.e., stem-loops). By deeply sequencing these resulting fragments and quantifying the relative abundance of specific intra-molecular ligation junctions, pairwise contact maps reflecting the short- and long-range stem-loop and pseudoknot interactions of intra-molecular RNA secondary structures can be obtained (Ramani et al., 2015). Some variants of this approach, including SPLASH, PARIS, and COMRADES, add an inter- or intra-molecular crosslinking step to hold the interacting fragments together, and these methods present different strategies for fragmentation and enrichment of cross-linked RNA molecules (Kudla et al., 2020). Similar to XL experiments, information on pairwise RNA interactions obtained from RNA proximity ligation can be integrated to create accurate structural models.

Computational approaches for RNA 3D structure prediction

Given the rapid growth of RNA sequence information and the very slow increase in the number of experimentally determined RNA structures, a variety of computational approaches have been developed to predict RNA 3D structures from sequence information (Somarowthu, 2016).

Manual modeling is an interactive process in which a user manually manipulates or assembles individual nucleotides or RNA segments to construct RNA 3D structures with a

graphical interface. A variety of programs, including RNA2D3D (Martinez et al., 2008) and S2S/Assemble (Jossinet et al., 2010), are available for manually building RNA models. Although these methods can be employed to build 3D structures for large RNAs with hundreds of nucleotides due to their manual tools, the structure building process can be very time consuming.

Comparative modeling, also known as homology modeling or template-based modeling, predicts the 3D structure of a "target" RNA based on the alignment of a primary sequence to one or more RNAs with known structures (templates). Similar to proteins, homologous RNAs also share a similar tertiary structure, and thus RNAs can be predicted using a known homologous structure in the database as a template. In recent years, various tools, such as ModeRNA (Rother et al., 2011), RNAbuilder (Flores and Altman, 2010), and Rosetta (Watkins et al., 2019), have been developed to build RNA homology models. Among these methods, the first and most crucial step is the identification of the template structure and the alignment of target and template sequences. With the increasing availability of experimentally resolved template structures, comparative modeling becomes increasingly accurate.

De novo modeling is a collective term for methods that predict RNA 3D structures directly from the sequence, without any template. *De novo* modeling methods can be categorized into three groups, including (i) all-atom-based, (ii) coarse-grained, and (iii) fragment-assembly techniques (Somarowthu, 2016). All-atom-based methods predict the structure by simulating RNA folding processes with physics-based atomic force fields such as AMBER (Case et al., 2005) and CHARMM (Brooks et al., 2009). These simulations are very detailed and, in principle, enable the actual motion of the atoms in the molecule to be seen. Coarse-grained methods simplify the representation of nucleotides by grouping the functional atoms as a single bead, thus reduce computational time during RNA folding simulations. A group of coarse-grained methods, such as NAST (Jonikas et al., 2009), iFold (Sharma et al., 2008), Vfold (Xu et al., 2014), have been used to model various RNA targets. Fragment-assembly methods first decompose known RNA structures into fragments and make a library of fragments, and then build RNA models by searching for fragments that are similar to the target sequence and assembling them together. Several approaches have been developed to perform all-atom assembling using known fragments, such as FARNA/FAR-FAR/FARFAR2 (Das and Baker, 2007; Das et al., 2010; Watkins et al., 2020), RNAComposer (Popenda et al., 2012), and 3dRNA (Zhang et al., 2020c; Zhao et al., 2012).

The future is hybrid

Given the scarcity of experimentally determined RNA 3D structures, computational RNA 3D predictions can be helpful in designing new structure-function experiments to accelerate discovery. However, "purely theoretical" structures usually have limited accuracy and often require experimental validation. While the inherent flexibility of RNA poses significant technological challenges and often exposes the limitations of individual structural biology techniques, it is still possible to obtain extensive heterogeneous structural information about RNAs from different biophysical, biochemical, and computational approaches. To overcome the shortcomings of a single technique, the field is increasingly inclined to develop a framework that combines data from multiple structural approaches, physical theories, and statistical analyses to form a more complete picture of dynamic biological assemblies. This approach is often referred to as hybrid or integrative modeling (Ward et al., 2013). To obtain a 3D structural model for RNA using integrative methods (Figure 4E), optimized smaller RNA subdomains or well-folded regions of functional importance can be studied using high-resolution techniques such as XRC, NMR or cryo-EM. For the remaining subdomains or regions whose homologous templates can be identified from the PDB, comparative modeling approaches can be used to predict the 3D structure. When no homologous template is available, *de novo* modeling approaches aided by additional data, such as secondary structures from bioinformatics analysis, enzymatic or chemical probing, can be used for 3D structure prediction. These theoretical RNA 3D structures can be validated, evaluated or filtered using structural information from different sources, such as low-resolution density maps or shape structures from cryo-EM, AFM or SAS, or pairwise distance restraints from biophysical (EPR, XSI or smFRET) or biochemical approaches (XL or RPL). Alternatively, integrative modeling algorithms can be used to combine all existing data to compute and refine structural models of RNAs, guide structure elucidation, and subsequently determine mechanisms of interactions and functions. A bunch of integrative modeling algorithms, including Integrative Modeling Platform (IMP) (Russel et al., 2012), M3 (Karaca et al., 2017), Xplor-NIH (Schwieters et al., 2018), etc., can be used to combine multiple types of restraints applicable to RNA modeling. Although many of these restraints are low-resolution structural information, they are able to resolve ambiguities when combined, thus further increasing the confidence of the final structural model.

AlphaFold has recently made a breakthrough in protein structure prediction through a new machine learning method (Jumper et al., 2021; Tunyasuvunakool et al., 2021) that relies on large database of structures. Although it is difficult to computationally predict RNA structures due to the small number of available structures, a novel machine learning approach, ARES, significantly improves 3D RNA structure prediction compared with previous approaches (Townshend

et al., 2021). In the future, an increasing number of RNA structures solved by different methods will drive artificial intelligence and deep learning techniques forward.

The functionality of RNA structures in micro-organism

The functionality of RNA structures in RNA viruses

RNA viruses include numerous human pathogens that cause severe diseases worldwide, such as human immunodeficiency virus (HIV), influenza A virus (IAV), dengue virus (DENV), Zika virus (ZIKV), and SARS-CoV-2. Information essential for regulating different viral processes is encoded both in the RNA genome sequence and in the higher-order structures of the genome itself (Boerneke et al., 2019). Various structured RNA elements have been characterized in the viral genome, particularly the 5′ and 3′ UTRs, which function via different mechanisms (Figure 5). For instance, in the case of hepatitis C virus, the internal ribosomal entry site that drives translation of the viral polyprotein forms a specific structure in the 5′ UTR of the viral genome (Khawaja et al., 2015). The 5′-leader RNA of HIV-1 contains RNA structures such as the trans-activation responsive region (TAR) and the dimer initiation site (DIS), as well as the φ hairpin (Figure 5A), which allows the genome to bind to the Tat protein to enable transcription. The latest research shows that the number of 5′ cap base G determines alternative conformations of 5′ leader, and then determines the fate of HIV-1 genome: packaging or translation (Brown et al., 2020). The 5′ leader with [Cap]3G and [Cap]2G tends to form a conformation with DIS being sequestered and cap/SD/AUG being accessible, in favorable to function as mRNA (Figure 5A), whereas 5′ leader with [Cap]1G tends to form an alternative conformation with DIS being exposed and cap/SD/AUG being sequestered, facilitating genome dimerization and packaging into viral particles (Figure 5B). A hairpin structure formed at the 5′ UTR of Sindbis virus RNA was found to regulate viral replication (Niesters and Strauss, 1990), while a cloverleaf structure at the 5′ end of the poliovirus RNA genome was shown to be central to RNA synthesis and formed the binding site for several proteins (Andino et al., 1990; Herold and Andino, 2001). Exoribonuclease-resistant RNA structural elements located within the 3′ UTR of the flavivirus genome prevent the degradation of downstream RNA sequences by stalling the 5′–3′ exoribonuclease Xrn1 (Ochsenreiter et al., 2019; Pijlman et al., 2008). In addition, the 5′ and 3′ UTRs of dengue and Zika viruses can be base-paired to enable genome circularization for efficient replication of viral genomes (Figure 5C). High-throughput experiments, including ic-SHAPE, SHAPE-MaP, and proximity ligation experiments performed on whole viral genomes, have also identified

many RNA structures in different viruses, including influenza, HIV, dengue, Zika, and coronavirus genomes including SARS-CoV-2 (Figure 5D), highlighting the architectural complexity of these genomes. Some of these structures can bind to RNA binding proteins as well as snoRNAs, enabling a better understanding of the structure-function relationships in RNA viruses.

Functional RNA structures have also been found within the coding regions of viral genomes. For instance, well-studied ribosomal frameshifting signal-associated pseudoknots can regulate translation of coronaviruses (Brierley et al., 1989), as well as poliovirus *cis*-acting replication elements involved in viral RNA encapsulation (Goodfellow et al., 2000; Rieder et al., 2000). In addition, the HIV-1 genome contains a triple helix RNA structure that results in a programmed −1 ribosomal frameshifting (−1 PRF) in the HIV-1 genome (Figure 5A) (Watts et al., 2009). Moreover, the RRE structure inside the env gene can bind to the Rev protein and facilitate the nuclear export of the HIV genome. RNA structures in viruses can facilitate other cellular processes, such as splicing. In human H3N2 influenza A virus, the 3′ ss of M2 is hindered by the RNA structure, thus limiting M2 expression (Bogdanow et al., 2019). The panhandle structure not only promotes the initiation of IAV transcription (Fodor et al., 1994), but also binds and activates RIG-I (Figure 5E) (Liu et al., 2015). IAV may adopt some strategy to balance these two functions. Upon infection with IAV, Z-RNA is produced, but the mechanism for this is currently unknown. In the nucleus, the Zα domain protein ZBP1 binds and recognizes Z-RNA, activates RIPK3-mediated MLKL activation, and triggers cellular necroptosis (Zhang et al., 2020b). Another Zα domain protein, DAI, can also induce RIPK3-mediated apoptosis by recognizing IAV genomic RNA (Thapa et al., 2016).

The complexity of RNA structure in RNA viruses is further highlighted by the observation that the same RNA can have numerous different structures within cells. Proximity ligation experiments with dengue and Zika viruses revealed that the same RNA region can pair with two or more other regions, indicating the existence of multiple alternative viral conformations within cells. In addition, using the DMS-MaP and DREEM algorithms, researchers quantitatively analyzed the variable RNA structure at the position of the HIV-1 Tat mRNA A3 splicing site (Tomezsko et al., 2020). About 33%–37% of the transcripts used the A3 stem-loop (A3SL) conformation, while 63%–67% of the transcripts used the A3-exposed conformation. Mutants that stabilize the A3SL conformation inhibit splicing at this site, indicating that RNA structure plays an important role in the splicing process of HIV-1 RNA. In addition, other variable RNA structures have been identified as widely distributed in the HIV-1 genome, including RRE and splice sites such as A4a, A4b, A4c, and A5 (Figure 5A). Besides affecting translation and splicing,
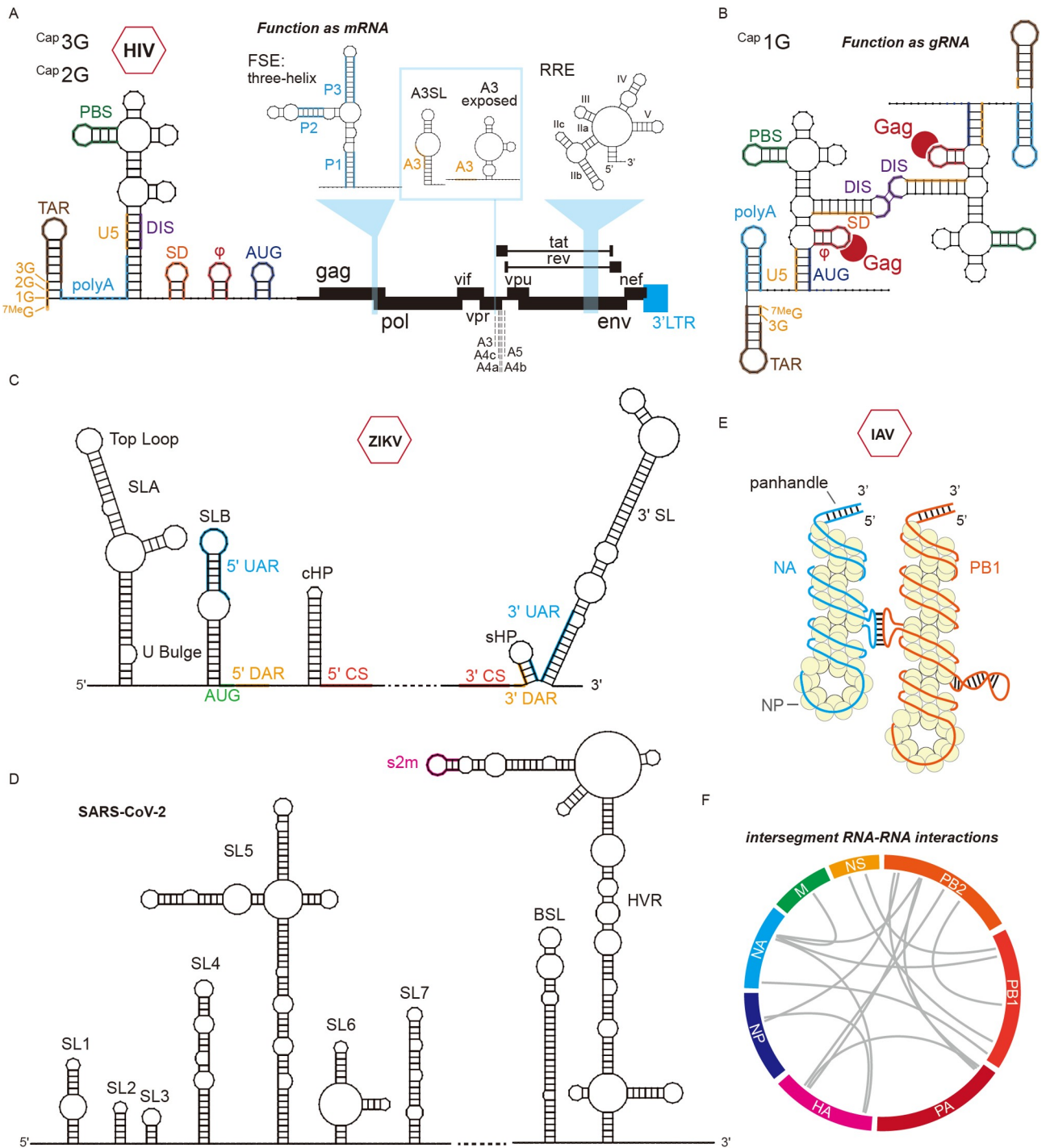
Figure 5   RNA structure-based strategies functioning in representative viruses. A, Schematic representation of HIV-1 RNA structure and DIS-sequestered 5′ leader RNA with ^{Cap}3G in functioning as mRNA. ^{7Me}G represents the cap structure of HIV-1 RNA. 1G, 2G and 3G represent numerical order of guanine at the 5′ end. The structural elements in the 5′ leader include TAR, polyA, U5, PBS, DIS, SD, φ, and AUG. The frameshifting element (FSE) in the gag coding region, alternative structures regulating splicing (A3SL and A3 exposed) in tat mRNA, and Rev-responsive element (RRE) are indicated with light blue inserts. A3 represents splicing acceptor site 3 in HIV. Other alternative structures regulating splicing, such as A4a, A4b, A4c, and A5, are indicated with dashed lines. B, Schematic representation of DIS-exposed 5′ leader with ^{Cap}1G of HIV-1 in functioning as gRNA. Two single gRNA molecules interact with each other through exposed DIS regions to form RNA dimer. Gag proteins bind packaging signals (φ) to mediate HIV-1 genome packaging. C, Conserved RNA structural elements in DENV and ZIKV viruses. Complementary regions responsible for genome cyclization are highlighted with different colors: blue for 5′–3′ UAR, yellow for 5′–3′ DAR, and red for 5′–3′ CS. Dashed lines indicate omitted genome sequence. D, Conserved RNA elements in the 5′ and 3′ UTR of the SARS-CoV-2. The stem-loop II-like motif (s2m) is shown in pink. Dashed lines indicate omitted genome sequence. E, Schematic representation of NA-PB1 RNA-RNA interactions in Udorn (H3N2) strain. Bold colored lines represent gRNA segments and yellow balls represent nucleoprotein (NP). The Panhandle region lies at the end of each segment. The NA-PB1 RNA-RNA interaction is marked at the approximate location. F, Schematic representation of intersegment RNA-RNA interactions in Udorn (H3N2) strain. Gray arch lines indicate main intersegment interaction regions detected by SPLASH.

RNA structures on the HIV-1 genome have the potential to inhibit small interfering RNA (siRNA) targeting. Cultivation of HIV-1 with siRNA targeting the Nef gene resulted in mutant strains that escaped siRNA targeting. In these strains, the R8 mutant has a G to A substitution at the 7 nt upstream position of the siRNA target site. This mutation induces the target site to form a stable stem-loop structure, which limits the recognition of the target site by siRNA and thus escapes siRNA targeting (Westerhout et al., 2005). These results show that RNA structures can be used as a universal strategy for viral survival in helping to adapt to the host cell environment. Using DRACO and DMS-MaP, the researchers also identified extensive amounts of RNA heterogeneity in SARS-CoV-2, suggesting that the formation of alternative structures is prevalent in RNA viruses (Morandi et al., 2021).

In addition to identifying RNA structures intramolecularly along viral genomes, high-throughput structural experiments have shown how different fragments of virus genomes interact to package and segregate together. Through SPLASH, the phenomenon of segment co-segregation of influenza A viruses was found to be consistent with intersegmental RNA-RNA interactions (Figure 5F) (Dadonaite et al., 2019). The NA segment of the Udorn strain in H3N2 specifically co-separated from the PB1 segments of other strains in H3N2, but without PB1 in H1N1, which was achieved through NA-PB1 RNA-RNA interactions. However, the Wyo03 strain in H3N2 did not have this specificity. If the corresponding interaction site on the NA segment of the Wyo03 strain was mutated to the Udorn strain sequence, the NA-PB1 RNA-RNA interaction was restored and the NA-PB1 co-segregation phenomenon was recovered. Similar intersegmental interactions were found using the 2CIMPL (dual cross-linking, immunoprecipitation, and proximity ligation) method (Le Sage et al., 2020).

Sequence-based RNA folding algorithms and phylogenetic analyses have been used to predict functional viral RNA genome structural elements; many of the predicted putative structures have been validated as functional by mutational analysis (Damgaard et al., 2004; Manzano et al., 2011; Marz et al., 2014; Sasaki et al., 2001; Tuplin et al., 2002). With the advent of sequencing-based enzymatic and chemical probing methods over the last decade, it is now possible to evaluate the structure of an intact viral RNA genome on a genome-wide scale, *in vitro*, and/or *in vivo* (Burrill et al., 2013; Mauger et al., 2015; Watts et al., 2009; Wu et al., 2013). Many viral RNA genome structures within native viral particles or host cells have been elucidated (Diaz-Toledano et al., 2017; Huber et al., 2019; Li et al., 2018; Pirakitikulr et al., 2016; Sun et al., 2021a).

The functionality of RNA structures in bacteria

RNA molecules with well-defined structures are widely used

as important tools by bacteria to regulate various biochemical processes in cells. Riboswitches, located in the 5′-untranslated region of mRNA, are among the most abundant RNA structure-based regulators that mediate gene expression at the transcriptional or translational level (Howe et al., 2015; Lund et al., 2020; Serganov and Nudler, 2013; Sherlock et al., 2018). Typically, riboswitches consist of two main functional domains, the sensing domain and the adjacent expression platform. The sensing domain can bind specifically to cognate metabolites and induce conformational changes throughout the riboswitch structure, which results in the formation of terminator/anti-terminator stems (in transcription) or sequester/anti-sequester stems (in translation) in the expression platform and, finally, regulates the turning on or off of related gene expression (Serganov and Nudler, 2013). Since the first discovery of riboswitches in 2002 (Mironov et al., 2002; Nahvi et al., 2002; Winkler et al., 2002a; Winkler et al., 2002b), more than 45 riboswitches have been identified to date (McCown et al., 2017). Based on the bound cognate metabolites, riboswitches can be classified into six families, including enzyme cofactors and their derivatives, purines and their derivatives, amino acids, metals, anions, and glmS riboswitches (Serganov and Nudler, 2013). Riboswitches participate in the regulation of gene expression of the related proteins, which are involved in the transport or biosynthesis of their cognate metabolites in a structure-based feedback mechanism (Serganov and Nudler, 2013). Each riboswitch has its characteristic conserved sequence, a specific secondary structure and distinct tertiary structures. In the higher order structures of riboswitches, long-range interactions usually play an essential role in stabilizing the overall structure and recognizing cognate ligands (Jones and Ferré-D'Amaré, 2017). As shown in Figure 6A, the FMN riboswitch in *E. coli* is able to bind to FMN (Flavin MonoNucleotide) to form an alternative RNA tertiary fold containing multiple long-range interactions followed by a stable sequester stem unfavorable to translation initiation, thus negatively regulating the expression of genes associated with FMN biosynthesis (Howe et al., 2015; Serganov et al., 2009; Winkler et al., 2002b). It is notable that most riboswitches exist in bacteria. The isolation of transcription and translation may hinder the evolution of riboswitches in eukaryotes. Only one type of thiamine pyrophosphate (TPP) riboswitch has been identified in plants to date, which regulates the gene expression by alternative splicing (Wachter et al., 2007). In evolution, the cellular metabolite-sensing and gene-regulation function of some ancient eukaryotic RNAs may have been substituted with various functional proteins. It is an intriguing scientific question whether there are more types of eukaryotic riboswitches await to be identified in the future.

Additional structure-based RNA regulators include T-box RNAs (Breaker, 2018; Sherwood and Henkin, 2016; Zhang

and Ferré-D'Amaré, 2016), temperature-sensing RNA thermometers (Kortmann and Narberhaus, 2012; Sherwood and Henkin, 2016), and pH-sensing RNA-based pH sensors (Sherwood and Henkin, 2016). T-box RNAs are involved in amino acid starvation. However, rather than interacting with free amino acids, T-box RNAs bind to tRNA that are not aminoacyl modified (Breaker, 2018; Sherwood and Henkin, 2016). Tertiary structural intermolecular interactions are involved in the recognition of tRNA by T-box RNAs (Breaker, 2018; Zhang and Ferré-D'Amaré, 2016). RNA thermometers do not form any specific ligand binding sites, but only un-

dergo structural changes when sensing different temperatures. In *E. coli*, the *rpoH* gene expresses the heat shock transcription factor σ32, whose translation initiation is regulated by temperature (Kortmann and Narberhaus, 2012). At 30°C or 37°C, the ribosome binding site (RBS) and the AUG region form an RNA structure that inhibits ribosome binding and limits its translation, while at 42°C, the RBS and AUG are completely open to maintain a high level of translation (Figure 6B). The RNA-based pH sensors are pH-sensitive, one of which is located in the 5′ UTR region of the *E. coli alx* gene. The results show that RNA-based pH sen-
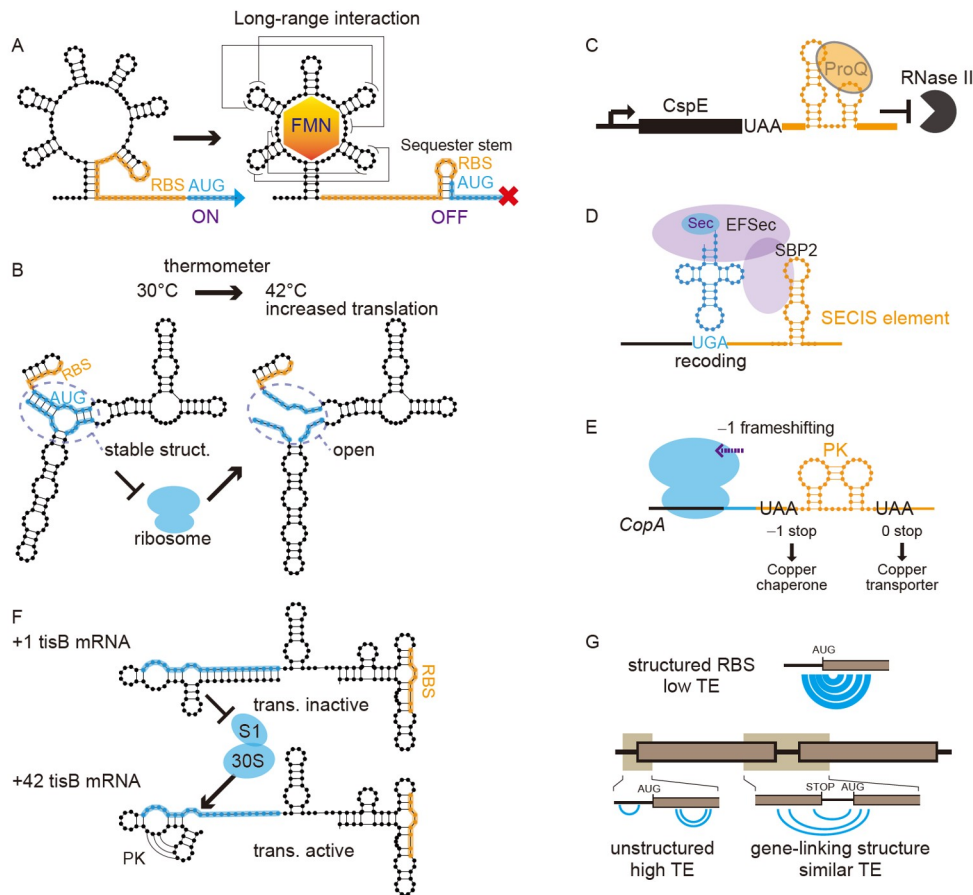

Figure 6 RNA structure-based strategies functioning in bacteria. A, Schematic representation of FMN riboswitch. The absence of FMN binding results in translation-ON conformation that upstream sequence of RBS is involved in 5′ structure, thereby RBS is released for ribosome binding (Left); in the existence of FMN, the sensing domain of FMN riboswitch forms a compact bound-form structure with extensive long-range interactions (labeled with thin lines), which induce the formation of the translation-OFF conformation containing the stable sequester stem between RBS and start codon region (AUG), thereby preventing translation. RBS represents ribosome binding site. B, Schematic representation of RNA thermometer of *rpoH* gene. At 30°C, the start codon region (AUG) pairs with downstream CDS region which results in limited ribosome trafficking and limited expression (Left). A rise in temperature to 42°C opens loop structure and liberates AUG start codon, thereby increasing translation (Right). C, RNA structure predicted to exist in *cspE* 3′UTR in *Salmonella* and bound by ProQ in preventing RNA decay through counteracting exoribonuclease activity. D, SECIS element located downstream of the UGA stop codon binding to SBP2 protein in recruiting selenocysteine-specific translation elongation factor (EFSec) to recode the UGA stop codon as Sec. E, −1 PRF in *E. coli CopA* gene. A pseudoknot (PK) structure that located at the CDS region facilitates −1 frameshifting, and produces a premature translation termination product CopA(Z) protein. F, *E. coli tisB* mRNA isoforms with differential translation efficiency. The blue bold line represents S1 and 30S binding sites. The +1 isoform of *tisB* mRNA contains intact 5′ UTR in which S1 and 30S binding site is restricted due to pairing with 5′ region, thereby preventing RBS to be released and recognized by ribosome (top). The +42 isoform forms a pseudoknot (PK) at 5′ end which facilitates S1 and 30S binding, thereby promoting RBS releasing and translation (bottom). G, Gene linking structure and coordinated translation. The blue arch line represents RNA-RNA interactions. Highly structured RBS and start codon region (AUG) restrict ribosome binding and low translation efficiency (TE) (top). The RNA-RNA interactions between two genes in polycistron tend to restrict TE of the downstream gene, however, translation of upstream gene will eliminate the RNA-RNA interactions and activate translation of the downstream gene, resulting in a similar TE of polycistron (bottom).

sors exhibit a structural shift between neutral pH and high pH (Nechooshtan et al., 2009). Under alkaline conditions, the RNA-based pH sensor forms a ribosome-accessible structure that allows the expression of *alx* genes (Nechooshtan et al., 2014; Nechooshtan et al., 2009).

RNA structure plays an important role in the post-transcriptional control of bacterial RNA dynamics. Both *Salmonella* and *E. coli* encode a highly conserved RNA structure-binding protein, ProQ, which specifically recognizes the RNA structural motif in 3′ UTRs and inhibits the exonuclease from degrading mRNA (Figure 6C) (Holmqvist et al., 2018). In addition to regulating RNA turnover, RNA structure is used more extensively for translational regulation. In *E. coli*, the *FdhF* gene has a selenocysteine insertion sequence (SECIS), a 17 nt stem-loop structure located downstream of the UGA stop codon, which mediates codon recoding to selenocysteine (Figure 6D) (Liu et al., 1998). Another type of translational recoding is the programmed ribosomal frameshifting (PRF) mediated by a slippery sequence and downstream pseudoknot in the coding sequence (CDS) region. The copper ion transporter gene *CopA* in *E. coli* uses the −1 frameshifting to generate an early translation termination protein CopA(Z), which is the chaperone protein of CopA itself (Figure 6E) (Meydan et al., 2017). A complex pseudoknot RNA structure is required for activating *tisB* translation in *E. coli* in response to DNA damage. The RBS site of the *tisB* gene is restricted to structured RNA, and the pseudoknot is responsible for binding S1 protein and 30S to help expose the RBS site (Romilly et al., 2020) (Figure 6F).

By means of SHAPE-MaP, a recent study found that RNA structure widely affects translation in *E. coli* and that translation efficiency is positively correlated with SHAPE reactivity (Mustoe et al., 2018). This correlation was reduced with the addition of the translation initiation inhibitor Kasugamycin, indicating that RNA structure is a limiting factor that extensively affects mRNA translation. The presence of RNA structures in RBS has a significant effect on translation efficiency, while structured RBS inhibits translation efficiency (Figure 6G top). Interestingly, *E. coli* uses this structural feature to regulate the translational coupling between polycistrons. When the RBS of the downstream gene interacts with the CDS of the upstream gene, the translation efficiency of both genes is similar (Figure 6G bottom) (Mustoe et al., 2018). Moreover, bacteria have evolved the phenomenon of N-terminal codon bias to avoid the appearance of RNA structures around the translation initiation site (Bhattacharyya et al., 2018).

In summary, RNA structure-based strategies are widely used in the life cycle of viruses and bacteria. Thus, RNA functional structural elements are potent targets for developing antiviral and antibacterial drugs. For example, ribocil, an antibiotic that selectively targets the riboflavin riboswitch,

was screened out to effectively inhibit bacterial cell growth (Howe et al., 2015). Identifying these functional structural elements using the RNA structurome and functional screening assays will be challenging and very important in future studies.

The functionality of RNA structures in animal

RNA structures tend to change dynamically to orchestrate gene expression in various species. At a global level, the structures and interactions of RNAs in RNP machines, such as ribosome and spliceosome, undergo dynamic changes during their assembly and function to ensure faithful and highly efficient cellular progresses in animal cells (Anger et al., 2013; Wilkinson et al., 2020). Many structural elements are present in the 5′ UTR and 3′ UTR of mammalian genes to modulate their expression post-transcriptionally in response to environmental stimuli or stresses. For example, the hypoxia stability region (HSR), which can undergo dynamic conformational changes, is present in the 3′ UTR of *VEGF* (also known as *VEGFA*) mRNA (Figure 7A). In the presence of IFN-γ and hypoxia, HSR adopts a translation-permissive conformation that allows the binding of HNRNPL and excludes the binding of EPRS for the efficient translation of *VEGF* (Ray et al., 2009). However, the HSR secondary structure is switched to a translation-silent conformation at normoxia condition which inhibits HNRNPL binding but permits EPRS binding, thus blocking *VEGF* induction (Ray et al., 2009). Since *VEGF* is typically induced by hypoxic stress in many pathological conditions, such as tumor centers and atherosclerotic lesions, this elegant stress-inducible RNA switching mechanism may also be applicable to various VEGF-related diseases. A similar RNA switching element was observed in another transcript, GNPDA1, in which a stem-loop structure was formed between the binding site of the RNA-binding protein PTB and several miRNAs targeting sites (Xue et al., 2013). In the presence of PTB, the stem-loop structure is disrupted and miRNAs can efficiently access their binding sites to repress translation. Conversely, the structure is stably formed in the absence of PTB, thereby blocking miRNA targeting and allowing efficient translation of GNPDA1 (Figure 7B). This mechanism may also occur during PTB-mediated neuronal programming and reprogramming processes (Xue et al., 2013).

In addition to translational regulation, RNA structural dynamics have emerged as a fundamental mechanism to regulate RNA decay. Global mapping of RNA structure by DMS-seq in zebrafish embryos revealed that mRNA structures are significantly remodeled during the maternal-to-zygotic transition (MZT). In contrast to the reduced structure of the 5′ UTR and coding regions, the 3′ UTRs form intricate structures *in vivo* that affect maternal mRNA expression by
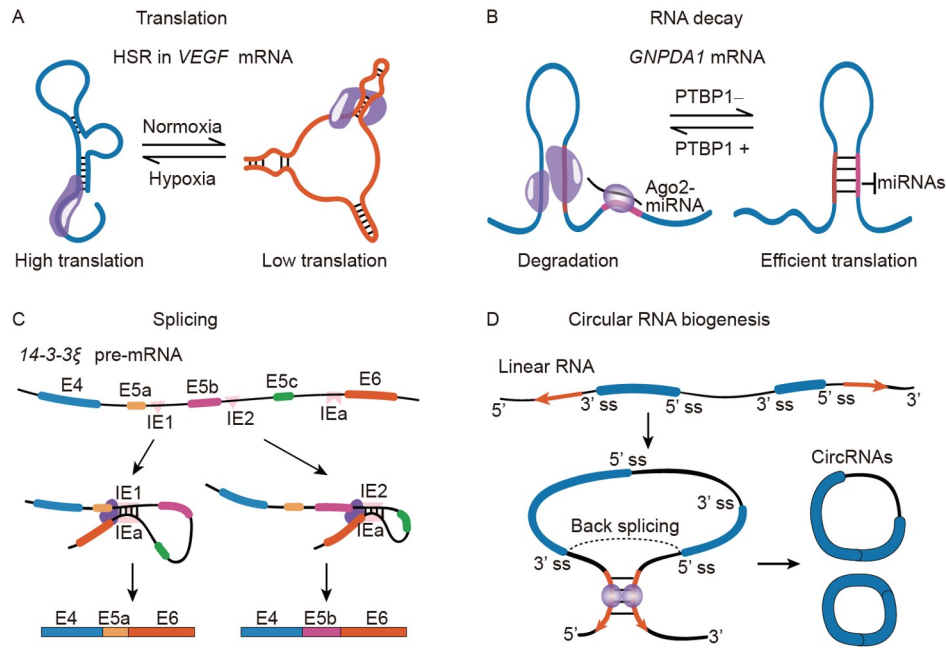
Figure 7 Dynamic RNA structures in post-transcriptional regulation. A, The structural changes in the HSR at the 3′ UTR in *VEGF* mRNA regulate its translation efficiency upon hypoxic stress. Purple, RBPs. B, The alterations in the structure of 3′ UTR can enhance or repress the degradation of mRNA by impacting its interplay with RBPs (purple ellipses) or miRNAs. C, The competing base-pairing in *14-3-3ξ* pre-mRNA determines the choice of alternative exons. E4, exon 4; E5a-c, exon 5a-c; E6, exon6; IE1 and IE2: intronic elements 1 and 2; IEa, docking site. D, The duplex formed between orientation-opposite complementary sequences (yellow arrows) in outsides introns bring splice sites into proximal, and promote back splicing and circRNA biogenesis. 5′ ss, 5′ splice site; 3′ ss, 3′ splice site; circRNA, circular RNA.

modulating the activity of microRNA miR-430 (Beaudoin et al., 2018). Notably, miR-430 is a crucial factor in promoting the deadenylation and clearance of maternal mRNAs (Giraldez et al., 2006). Another structural mapping of mRNAs during early zebrafish embryogenesis by icSHAPE reveals a similar phenomenon, where structurally variable regions are also enriched in 3′ UTRs containing *cis*-regulatory RNA elements essential for MZT (Shi et al., 2020).

RNA structural changes also play critical roles in regulating splicing. In *Drosophila melanogaster*, mutually exclusive splicing of the *Dscam* (down syndrome cell adhesion molecule) gene can produce 38,016 different mRNA isoforms from 95 variable exons. Comparative genomics revealed two conserved *cis*-elements, docking sites and selector sequences in the *Dscam* exon 6 cluster. Alternative base pairing between the docking site in intron 5 and one of the selector sequences determines which variable exon is included (Graveley, 2005). Surprisingly, a novel set of RNA secondary structures was found in the exon 6 cluster of *Dscam1* from the two-midge species (*Belgica antarctica* and *Clunio marinus*). These new pairings are equivalent to docking site-selector sequences base-pairing based on their locations, but the paired primary sequences are completely different from the docking sites and selector sequences (Hong et al., 2020). Recently, a set of hidden RNA secondary structures, termed balancer RNA pairings, was found to drive the stochastic choice of the exon 6 cluster in *Dscam1*

(Dong et al., 2022). Moreover, such competing RNA structural mechanisms were later found and experimentally validated in exon clusters 4 and 9 of *Dscam* and in other insect genes such as *14-3-3ξ* pre-mRNA (Figure 7C) (Dong et al., 2021; Hong et al., 2021; Yang et al., 2011; Yue et al., 2016). In addition to canonical splicing, a kind of non-canonical splicing called back splicing has been shown to be the fundamental mechanism for generating circular RNAs (Kristensen et al., 2019). Intronic Alu elements or protein-mediated long-range looping dynamically modulates the ratio of canonical to back splicing, resulting in the generation of large amounts of circular RNAs (Figure 7D) (Conn et al., 2015; Ivanov et al., 2015; Liang and Wilusz, 2014; Zhang et al., 2014).

RNA structural dynamics also appear to be critical for efficient primary and precursor miRNA processing. A hairpin structure at the 3′ end of long non-coding RNA (lncRNA) *NEAT1* has been demonstrated to resemble a primary miRNA for attracting the Drosha-DGCR8 Microprocessor to paraspeckles, thus enhancing the processing of hundreds of primary miRNAs localized in the subnuclear body (Jiang et al., 2017). The RNA-binding protein hnRNPA1 has been shown to bind at the apical loop of pri-miR-18a, destabilizing base pairs in the loop-proximal stem regions and thus promoting Drosha cleavage (Kooshapur et al., 2018). A recent parallel functional assay termed Dro-seq (Drosha sequencing) also revealed that the structural ensemble of

pri-miRNA hairpins is vital for their efficient processing (Rice et al., 2020). In addition, stem-loop structures in *cis*-acting RNA elements can also control the specific subcellular localization of mRNAs (Martin and Ephrussi, 2009). These *cis*-acting RNA elements, also known as "localization elements" or "zip codes", are mainly positioned in the 3′ UTR of mRNAs. For example, *bicoid* mRNA is known to be transported to the anterior pole of *Drosophila* oocytes, and a 50 nt localization element in the 3′ UTR containing a stem-loop structure is essential for its specific localization (Macdonald et al., 1993).

Besides classical stem-loop motifs, rG4 is a kind of four-stranded RNA structure in the 3′ UTR that can also act as a localization element, regulating the localization of PSD-95 and CaMKIIa mRNAs in cortical neurons (Subramanian et al., 2011). In addition to those elements in the 3′ UTR, several structural *cis*-acting elements in the coding region or 5′ UTR have been shown to possibly affect RNA localization. In *Saccharomyces cerevisiae*, four stem-loop-containing sequences in the coding region of *ASH1* mRNA have been demonstrated to serve as localization elements for regulating its transport to the bud tip (Chartrand et al., 1999). Beyond the secondary structural motifs, the specific higher-order 3D conformation of these *cis*-acting elements is also crucial for RNA localization. The NMR spectroscopy derived a tertiary structure of the localization elements in *fs(1) K10* mRNA, a 44 nt stem loop, revealing an A'-form conformation with widened major grooves. This specific 3D conformation can be recognized by the dynein motor and mediates the transport of *fs(1)K10* mRNA in *Drosophila* (Bullock et al., 2010). Notably, the dynamic modulation of these structural features emerges as an important way to regulate RNA localization and expression.

Dynamic RNA structures not only regulate the metabolism of their host RNAs, but also determine the fate of other RNA molecules. Most recently, Xiang et al. found that the locally open region of a circular intronic RNA named *ciankrd52* enables its competition with the structured linear pre-mRNA *ANKRD52* to form R-loops with DNAs, thus protecting pre-*ANKRD52* from RNase H1 degradation and facilitating its transcriptional elongation (Li et al., 2021). Besides circular RNAs, many lncRNAs adopt specific secondary and tertiary conformations to control transcription, RNA splicing, RNA stability and mRNA translation (Statello et al., 2021). One of the most famous examples is the lncRNA *roX* in *Drosophila*, which utilizes several tandem stem-loops to attract MLE proteins and form the male-specific lethal (MSL) complex to promote transcription of genes on the X chromosome in males (Ilik et al., 2013). Many other transcription-regulating lncRNAs adopt a similar mechanism, such as *Bvht* in mice (Xue et al., 2016), *XIST* and *HOTAIR* in human (Lu et al., 2016; Somarowthu et al., 2015). Few abundant lncRNAs, such as lncRNA *MALAT1*, contain many structured elements

and may undergo conformational changes to organize the assembly of nuclear condensates through multivalent interactions with distinct proteins (McCown et al., 2019; Statello et al., 2021). LncRNA *MALAT1* is an essential component of the nuclear speckle that affects pre-mRNA splicing and transcription. However, the functional importance and regulatory mechanism of the structures of these lncRNAs in the formation of nuclear condensates remain to be investigated.

In summary, RNA structures and their dynamic changes can determine the fate of almost all types of RNAs and are emerging as a crucial way to modulate gene expression in a variety of species and biological processes. Technical advances in recently invented methods have significantly broadened our understanding of the intricate RNA structurome and interactome within cells. Approaches that combine these experimental efforts with sophisticated deep learning and artificial intelligence may finally uncover a panoramic view of the RNA structural world.

The functionality of RNA structures in plant

Plants comprise a large number of species that form one of the kingdoms of eukaryotes. By virtue of their unique cellular structure, plants have distinct morphologies for growth and development. As sessile organisms, plants must adapt to diverse environmental conditions in order to survive in widely distributed landscapes. Here, we review recent advances in methodologies and knowledge in the field of plant RNA structure. We provide an overview of RNA structural functionality in plant growth and development as well as plant responses to varying environmental conditions.

Advances in understanding the functions of plant RNA structures

In recent years, some RNA structure probing methods have been designed specifically to study RNA structure in plants (Deng et al., 2018; Ding et al., 2014; Foley et al., 2015; Su et al., 2018; Wang et al., 2018). Some are new RNA structure probing methods that were originally developed in the plant system (Foley and Gregory, 2016; Gosai et al., 2015; Liu et al., 2021; Yang et al., 2020a; Yang et al., 2020b; Yang et al., 2021). Protein interaction profiling sequencing (PIP-seq) complements PARS-seq by including an additional step in which RNAs and proteins are cross-linked via formaldehyde or UV light prior to enzymatic probing, providing extra information about RNA-protein interactions (Gosai et al., 2015). This method was initially developed in the *Arabidopsis* nucleus, which is also applicable to the mammalian species (Shan et al., 2021; Silverman et al., 2014). A DMS-based *in vivo* RNA structure profiling method was simultaneously established in yeast and *Arabidopsis* (Ding et al.,

2015; Ding et al., 2014; Rouskin et al., 2014). Due to the unique cellular architecture of plant cells, high DMS concentrations of over 2% or prolonged reactions of over 30 min led to the browning of *Arabidopsis* seedlings, probably due to chlorophyll molecules reacting with DMS (Wang et al., 2019b). For SHAPE-based methods, only one SHAPE reagent, NAI, has been successfully applied to *Arabidopsis* seedlings (Yang et al., 2020a; Yang et al., 2020b).

To understand the possibility of a causal link between RNA structure and biological processes, it is important to distinguish the RNA structures of different mRNA populations at different stages of the mRNA life cycle. In plant systems, an NAI-based *in vivo* RNA structure probing method, Nuc-SHAPE-Structure-Seq, was developed to enrich the RNA structure of pre-mRNAs (i.e., before mRNA processing) for studying mRNA processing (Liu et al., 2021). This study revealed that a two-nucleotide single-stranded RNA structure feature upstream of 5′ ss is important for recognizing 5′ ss by splicing and selecting alternative 5′ ss. A comprehensive mutagenesis analysis revealed that this RNA structural feature is sufficient to alter splicing fate (Figure 8) (Liu et al., 2021). During mRNA

processing, the recognition of polyadenylation sites in plants is also not always dependent on sequence content (Loke et al., 2005). Nuc-SHAPE-Structure-Seq identified two adjacent single-stranded regions (from −28 nt to −17 nt upstream of the poly(A) site and from −4 nt to +1 nt across the poly(A) site) that are important for recognizing both polyadenylation and alternative polyadenylation (Figure 8) (Liu et al., 2021). The maintenance of single-strandedness between PAS sites and poly(A) sites is likely to offer binding sites for those single-stranded RNA-binding proteins involved in polyadenylation.

Following mRNA processing, the mRNAs are subjected to the translation process. Both *in vitro* and *in vivo* RNA structure studies in *Arabidopsis* revealed a single-stranded region upstream of the start codon in the 5′ UTR, which is strongly associated with higher translation efficiency (Ding et al., 2014; Li et al., 2012b). Furthermore, in both *Arabidopsis* and rice (*Oryza sativa*), mRNAs with more single-strandedness as well as a triple-cycle trend in RNA structural patterns across CDS regions are closely associated with high translation efficiency (Deng et al., 2018; Ding et al., 2014). Apart from global RNA structural features, rG4 has recently
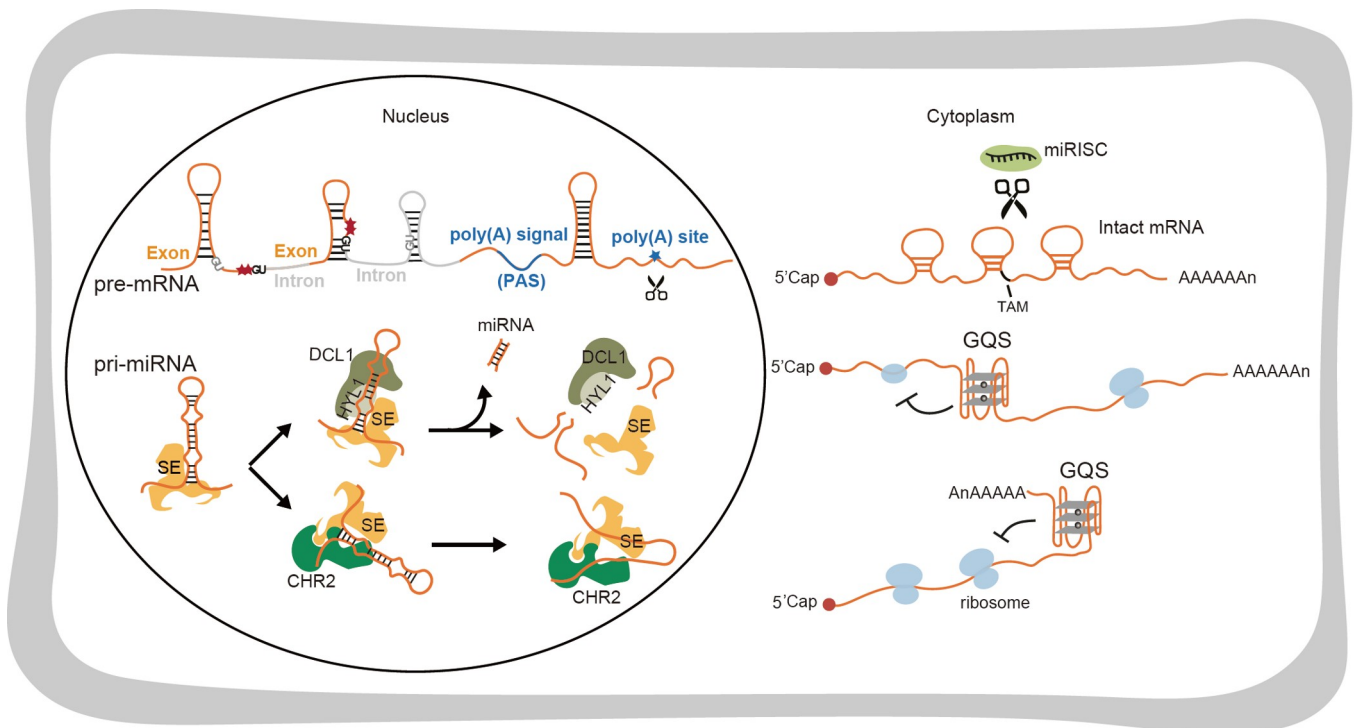

Figure 8   Schematic of some exampled novel molecular functions of RNA structure in plants. In nucleus, a two-nucleotide single-stranded RNA structure (red stars) upstream of 5′ splice sites (5′ ss, upstream of the first two-dinucleotide "GU" in the intron) is important for pre-mRNA splicing. If two nucleotides upstream of GU (in dark) are single-stranded, the splicing is switched on. If two nucleotides upstream of GU (in grey) are double-stranded, the splicing is switched off. Additionally, two close-by single-stranded regions across polyadenylation sites in pre-mRNAs are critical for both polyadenylation and alternative polyadenylation events. The first single-stranded region (highlighted in the blue line) is from −28 nt to −17 nt upstream of the polyadenylation (poly(A)) sites, overlapped with the PAS motif. The other region (blue star) is from −4 nt to +1 nt across the poly(A) site which might facilitate the cleavage. Moreover, chromatin remodeler CHR2/BRM binds pri-miRNAs via SE and unwinds their RNA secondary structure, inhibiting processing by DCL1 and HYL1. In cytoplasm, RNA G-quadruplexes (GQS) localized in either 5′ UTR or 3′ UTR of the intact mRNA are capable of repressing translation. In the miRNA-targeted mRNAs, the single-strandedness of the two nucleotides immediately downstream of the target site, TAM, is capable of triggering miRNA cleavage. Small blue circles represent small ribosomal subunits, whilst big blue circles indicate large ribosomal subunits.

been highlighted as an important RNA structural motif that regulates translation (Fay et al., 2017b). The first highly conserved plant rG4 was found in the 5′ UTR of *ATAXIA TELANGIECTASIA-MUTATED AND RAD3-RELATED (ATR)* in *Arabidopsis* and functions as a translational suppressor (Figure 8) (Kwok et al., 2015). Recent studies have identified hundreds of RNA G-quadruplex structures that fold globally in both *Arabidopsis* and rice, which represent both dicots and monocots (Yang et al., 2020b). A single folded rG4 located in the 3′ UTR of the gene *HIRD11*, which encodes a KS-type dehydrin, was able to suppress its own translation (Figure 8). In contrast, rG4 structures were not detected in either yeast or mice (Guo and Bartel, 2016). One possible reason for this is temperature, which is a key factor affecting the folding of rG4s. The optimal environmental temperature for plants (21–22°C for *Arabidopsis* and 26–28°C for rice) is much lower than the body temperature of animals. Therefore, these relatively lower temperatures may allow stable formation of rG4 structures in plant cells. Thus, ideal physiological conditions in plants may confer on plants the ability to adopt rG4 structures as translational regulators.

The microRNA (miRNA)-mediated RNA degradation pathway is one of the main RNA degradation pathways in plants (Yu et al., 2017). In *Arabidopsis*, chromatin remodeling factor 2 (CHR2), the ATPase subunit of the large SWI/SNF chromatin-remodelling complex, was determined to directly bind and unwind pri-miRNA structures (Wang et al., 2018). Remodelling of the pri-miRNA structure inhibited pri-miRNA processing by DCL1 (Microprocessor-Dicing complex includes Dicer-like 1), which recognizes hairpin structures. This work uncovered an additional regulatory layer that controls miRNA biogenesis via RNA structural remodelling (Wang et al., 2018). A newly developed SHAPE-based method, CAP-STRUCTURE-Seq, revealed landscapes of intact RNA structures in *Arabidopsis* to demonstrate a link between RNA structure and miRNA-mediated cleavage (Yang et al., 2020a). Surprisingly, miRNA target sites are structurally inaccessible for "miRNA-Induced Silencing Complexes" (miRISCs) binding prior to cleavage. Target site unfolding was found to be the rate-limiting step in determining miRISC activity *in vivo*. Besides the RNA structure within target sites, the RNA structure in the flanking region also contributed to miRNA-mediated cleavage, where the single-strandedness of the two nucleotides adjacent to the 3′ end of the miRNA target site called "target adjacent nucleotide motif" (TAM) is sufficient to trigger miRNA-mediated cleavage (Figure 8). However, TAM does not affect miRISC binding to the target.

Novel insights on the functions of RNA structure in plant growth and development

Recent RNA structure studies have discovered diverse functional roles of RNA structure across various aspects of plant growth and development. Flowering is one of the key developmental stages in plants. A key gene, *FLOWERING LOCUS C* (*FLC*), represses the initiation of flowering. The antisense lncRNA of *FLC*, *COOLAIR*, plays a major role in regulating *FLC* transcription in response to vernalization. A recent *in vitro* RNA structure study revealed that the *COOLAIR* transcript has a complex structural architecture (Figure 9) (Hawkes et al., 2016). Notably, the single nucleotide polymorphisms (SNPs) in a natural *Arabidopsis* variant, VAR2-6, resulted in a shorter helix H4 affecting RNA stability, thereby promoting increased *FLC* transcription and late flowering.

Another unique plant developmental feature is the plant vascular system, a remarkable evolutionary step, which allows plants to colonize the land (Ruonala et al., 2017). A recent study identified a novel RNA structure-mediated translational regulatory module for phloem development (Cho et al., 2018). This work revealed that *JULGI* (*JUL*), a key regulator of phloem differentiation, encodes a plant-specific protein with three RanBP2-type zinc finger (ZnF) domains, which bind directly to the 5′ UTRs of the *SUPPRESSOR OF MAX2 1-LIKE4/5* (*SMXL4/5*) mRNA (Figure 9) (Cho et al., 2018). This RNA-protein interaction helps to limit phloem differentiation. Within the 5′ UTRs of *SMXL4/5*, there are rG4 motifs found exclusively in vascular plant species. Further validation showed that *JUL* binds directly to these rG4 motifs, suppressing *SMXL4/5* translation and inhibiting phloem differentiation. Interestingly, abolishing the JUL-RNA G-quadruplex interaction led to an increase in phloem cells and enhanced the capacity of phloem transport, thereby enhancing the sink strength of seeds and leading to an increase in seed size and weight (Cho et al., 2018). The mRNAs are transported via the phloem to distant tissues that can potentially act as non-cell autonomous signals (Ruonala et al., 2017). A recent transcriptome-wide study found that a tRNA-like structure (TLS), harboured inside mRNAs, facilitates mRNA mobility (Figure 9) (Zhang et al., 2016). Remarkably, these efficiently transported mRNAs can also be translated into proteins after transport. Reporter transcripts fused to TLS were found to be enriched in the phloem stream. Comprehensive structural mutations suggest that stem-bulge-stem-loop structures within the TLS are sufficient to mediate mRNA transport (Zhang et al., 2016).

Unlike animals, plants survive via photosynthesis by converting light energy, carbon dioxide from air and water into glucose (Schlüter and Weber, 2020). A recent study revealed that *psbA*, which encodes the D1 protein of photosystem II (PSII), relies on an RNA structure on its own 5′ UTR to switch its translation on and off so as to turn photosynthesis on and off (Gawroński et al., 2021). The "translation-on" mode requires a high degree of accessibility to the translation initiation region. Besides *psbA*, other
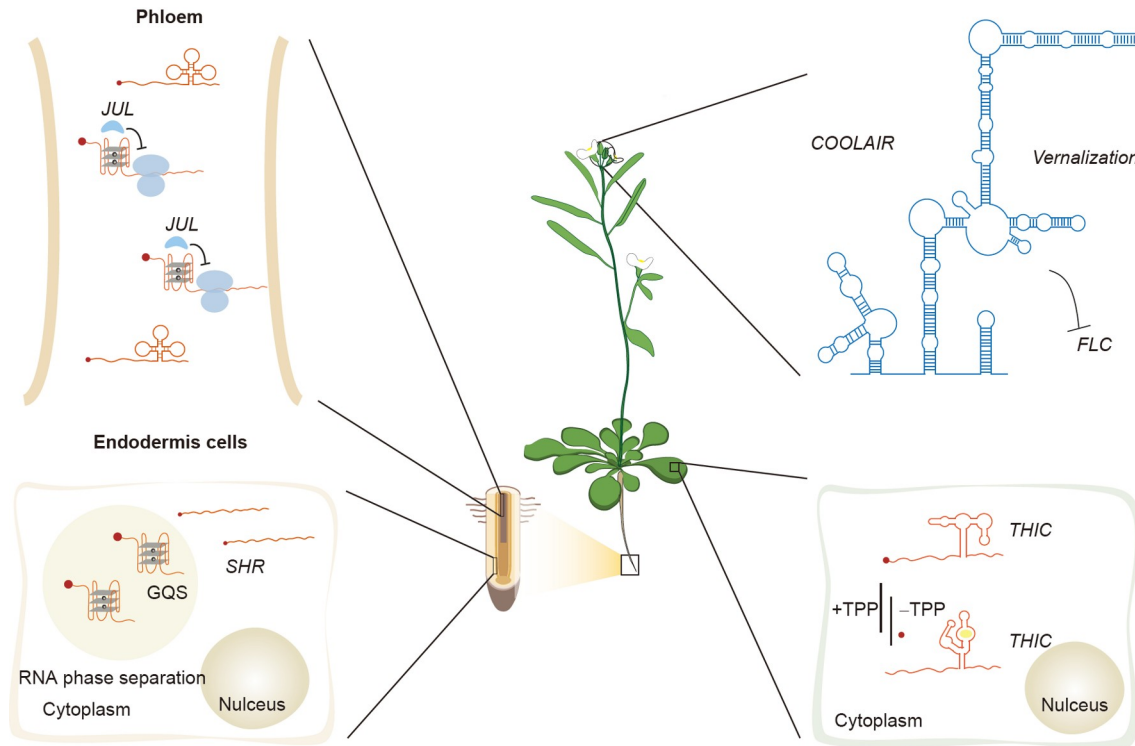
Figure 9 Schematic of some exampled novel functions of RNA structure in plant growth and development. The antisense lncRNA of *FLC*, *COOLAIR*, plays a major role in suppressing *FLC* transcription in response to vernalization, initiating flowering. Its complex structure was suggested to be critical for its function. In the plant phloem, a key regulator of phloem differentiation, *JUL* directly binds to the GQS of the *SMXL4/5* 5′ UTR, suppresses its translation and inhibits phloem differentiation. In the phloem stream, a tRNA-like structure (TLS), harboured inside mRNAs, facilitates mRNA mobility. In the endodermis cells, *SHR* mRNA contains an RNA GQS capable of triggering liquid-liquid phase separation. Plant vitamin B1 derivative TPP riboswitches are found in the 3′ end of pre-mRNA involved in splicing and alternative 3′ end processing. Under low TPP concentrations, the RNA structure conformation in the 3′ end of pre-mRNA contributes to short 3′ UTR, which increases *THIC* RNA levels. Under high TPP concentrations, TPP binds to the 3′ end of pre-mRNA, changing the RNA structure conformation. This TPP-induced RNA structural alteration prevents RNA splicing, resulting in long 3′ UTR. This longer 3′ UTR promotes RNA degradation, thereby reducing *THIC* RNA levels.

chloroplast mRNAs with weak Shine-Dalgarno sequence content also utilize this RNA structure-dependent translational regulation (Gawroński et al., 2021).

Below ground, plant roots are essential for plant productivity and serve a variety of functions, such as water and nutrient uptake (Motte et al., 2019). A recent single-molecule RNA FISH (smFISH) study on the cellular distribution of the key root development gene *SHORT ROOT* (*SHR*) has revealed a striking phenomenon in which phase separation-like signals of *SHR* mRNA are aggregated in endodermis cells (Zhang et al., 2019c). Further characterization of *SHR* mRNA features revealed that *SHR* mRNA contains rG4 capable of triggering liquid-liquid phase separation (Figure 9), suggesting that *SHR* RNA might adopt rG4 to store *SHR* mRNAs in endodermis cells for acquiring cellular identity during root development. Another rG4, located in the 3′ UTR of *HIRD11* mRNA, encodes a KS-type dehydrin and was recently identified from a high-throughput *in vivo* profiling study (Yang et al., 2020b). Further phenotypic assessment found that abolishing this rG4 in plants resulted in significantly longer primary roots, suggesting that this rG4 has a negative regulatory effect on plant root growth.

This example illustrates the putative RNA structure-dependent regulation of plant root development, offering the potential to improve plant root development via RNA structural alterations (Yang et al., 2020b). During root growth, epidermal precursor cells differentiate into either root hair or non-hair cells (Marin et al., 2021). A recent RNA-protein interaction study on both root hair and non-hair cell nuclei showed very distinct RNA-protein binding sites and corresponding *in vitro* RNA structural features between hair and non-hair cells, suggesting that cell type-specific RNA structures may lead to different binding affinities for RNA-binding proteins and thus regulate root hair cell fate (Foley et al., 2017).

Another unique feature of plants is that they comprise a large number of distinct metabolites, with some plant species containing up to one million metabolic compounds (Wang et al., 2019a). The TPP riboswitch, a vitamin B1 derivative in both algae and high plants, was identified through RNA structure conservation (Croft et al., 2007; Mehrshahi et al., 2020; Wachter et al., 2007). At low TPP concentrations, the RNA structural conformation at the 3′ end of pre-mRNA contributes to a shortened 3′ UTR, which

increases *THIC* RNA levels (Wachter et al., 2007) (Figure 9). At high TPP concentrations, TPP binds to the 3′ end of pre-mRNA and changes the RNA structural conformation. This TPP-induced RNA structural alteration prevents RNA splicing, which results in a long 3′ UTR. This longer 3′ UTR promotes RNA degradation, which in turn reduces *THIC* RNA levels.

The role of RNA secondary structure in abiotic stress responses in plants

Environmental factors, including abiotic and biotic stresses, are major limiters of plant growth and crop production (Gong et al., 2020). Recent technological advances have enabled the investigation of the effects of abiotic stresses on *in vivo* RNA secondary structure reprogramming and the correlation of RNA structural changes with gene regulation in plants (Zhu et al., 2021). Temperature is not only a key environmental parameter for plants, but also a major factor in RNA folding (Bevilacqua et al., 2016). Su and his colleagues employed Structure-seq2 to analyze rice RNA structurome under normal and acute heat shock conditions (Su et al., 2018) (Figure 10). Heat rapidly unfolded RNA structure *in vivo*, and the 3′ UTR was more susceptible to melting regions during acute heat shock than the 5′ UTR and CDS regions (Su et al.,

2018). Furthermore, heat-induced RNA structural melting was negatively correlated with heat shock-induced changes in mRNA abundance, so that more single-strandedness would reduce mRNA abundance. This suggests that the unfolding of heat shock-induced transcripts facilitates RNA degradation. Surprisingly, there was no correlation between heat-induced global RNA structural reprogramming and changes in translation efficiency. However, the authors suggest that such an RNA thermometer mechanism may occur in certain transcripts or may be revealed under long-term heat treatment (Su et al., 2018). Another recent study investigated the response of plants to increased daytime temperatures (Chung et al., 2020) (Figure 10). Approximately 700 translationally increased genes were identified in response to high temperatures, including *PHYTOCHROME-INTERACTING FACTOR 7* (*PIF7*). This temperature-responsive translational regulation of *PIF7* is caused by a putative hairpin secondary structure located upstream of the start codon site. This hairpin structure is flexible and can shift conformation between low and high temperature conditions, thereby regulating translational initiation. Notably, two other important regulators in thermomorphogenesis, *WRKY22* and *HEAT SHOCK FACTOR A2* (*HSFA2*), possess a similar hairpin structure that thermodynamically controls translation in response to higher temperatures (Chung et al.,
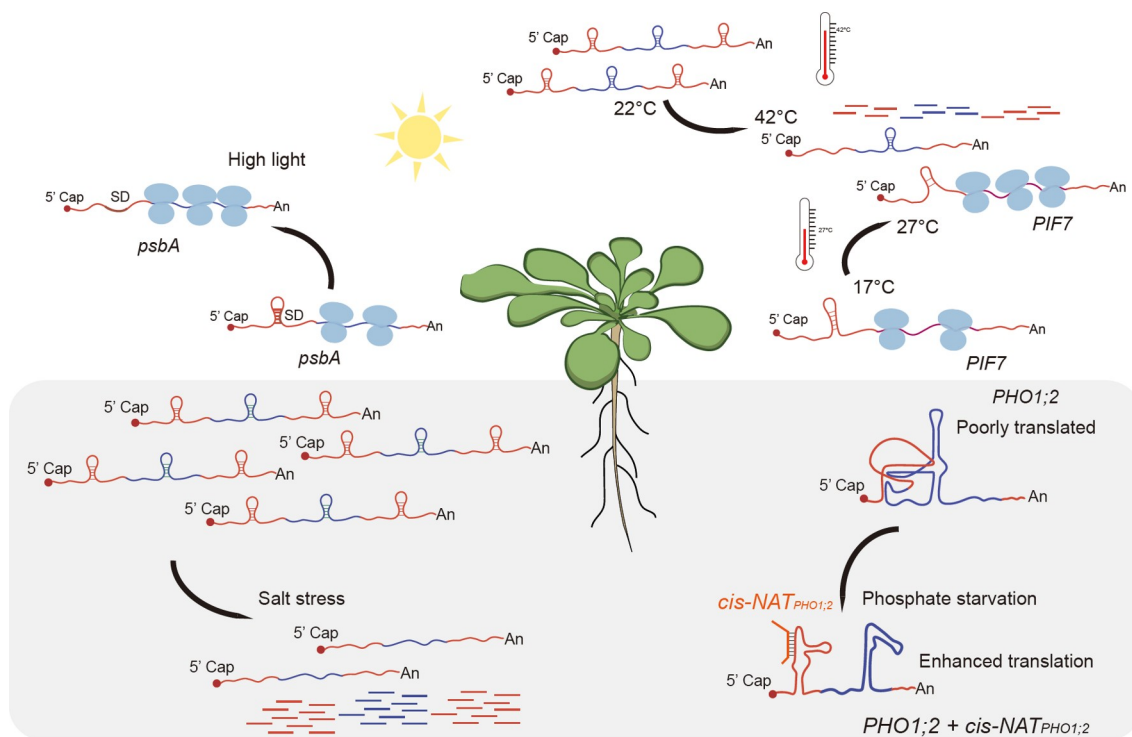

Figure 10 Schematic of some exampled new roles of RNA structure in plant response to abiotic stress. In response to heat stress, heat-induced RNA structural melting in the 5′ UTRs and 3′ UTRs could promote RNA degradation. In response to increased daytime temperature, the RNA hairpin within the 5′ UTR of *PIF7* was changed into an alternative conformation, leading to increased translation of *PIF7*. The RNA structure of *psbA* in 5′ UTR was changed to more single-strandedness under high light stress, increasing the translation of *psbA*. In response to salt stress, salt-induced RNA structural unfolding in the whole transcripts could promote RNA degradation. In response to phosphate starvation, a sense-antisense inter-molecular interaction between *PHO1;2* and *cis-NAT_pho1;2* rearranged the RNA structure to enhance *PHO1;2* translation.

2020). The above two studies seem to draw paradoxical conclusions about the relationship between alterations in RNA secondary structure and changes in translational efficiency in response to varying temperatures. Possible reasons could be due to different heat treatments, different methods, different species, and dimensionality of the analysis. Su et al. (2018) performed an acute heat shock treatment (treating 14-day-old rice at 42°C for 10 min) followed by genome-wide DMS probing with high-throughput data analysis, while Chung et al. (2020) performed a mild heat treatment (treating 14-day-old Arabidopsis seedlings by shifting them from 17°C to 27°C for 15 min) and focused on certain genes associated with *in vitro* folded RNA secondary structure-mediated translational regulation. This also indicates that plants adopt comprehensive RNA structure-mediated strategies to elaborately deal with various heat stresses.

Salinity stress limits the growth and productivity of crops worldwide (Mickelbart et al., 2015). A recent genome-wide *in vivo* RNA structurome study *in Arabidopsis* shoot and root tissues under both unstressed and salt-stressed conditions showed that salinity stress triggered alterations in RNA secondary structure (Tack et al., 2020) (Figure 10). This salinity-induced RNA structural unfolding reduces expression levels. Moreover, tissue-specific transcripts exhibited a more pronounced correlation than tissue-shared transcripts, indicating a robust tissue-specific influence on *in vivo* RNA structure (Tack et al., 2020). Another study later assessed the effect of N^6-methyladenosine (m^{6}A) modification on modulating nuclear RNA secondary structure during long-term salinity stress using PIP-seq (Kramer et al., 2020). The RNA secondary structure of protein-coding transcripts showed large-scale changes under salt stress, particularly in the 3′ UTR, where the RNA structure was significantly reduced. Notably, this salinity stress-induced RNA secondary structure reprogramming was not associated with changes in protein binding sites, indicating that the salt stress-induced global RNA secondary structure reprogramming is not due to protein binding alterations (Kramer et al., 2020). Since m^{6}A modification weakens base pairing (Spitale et al., 2015; Sun et al., 2019) and its deposition is dynamic during salinity stress (Anderson et al., 2018), there is a significant negative correlation between m^{6}A deposition and RNA structure (Anderson et al., 2018). In particular, salinity stress-induced m^{6}A deposition in the 3′ UTR of mRNA was accompanied by an obvious reduction in RNA structure (Anderson et al., 2018). However, in this study, changes in RNA secondary structure were not correlated with changes in transcripts and protein abundance during salinity stress (Anderson et al., 2018). This is inconsistent with the above study, which suggests a strong negatively correlated alteration between RNA secondary structure and transcript abundance under salt stress (Tack et al., 2020). In addition to the different salt treatments (short-term *vs*. long-term), Tack et al. analyzed

the RNA secondary structure of total RNA (mainly cytoplasmic RNA) by Structure-seq, while Kramer et al. investigated the structure of nuclear RNA by PIP-seq. These differences in experimental design and approaches may be the reasons for these different conclusions.

Phosphate (Pi) is essential for crop growth and yield. In plants, soil soluble Pi is absorbed by the root and transferred to above-ground tissues (Oldroyd and Leyser, 2020). Rice PHOSPHATE1;2 (PHO1;2) acts as a Pi exporter and participates in Pi transport from roots to shoots and Pi reallocation during grain filling (Jabnoune et al., 2013; Ma et al., 2021). *PHO1;2* has an associated cis-Natural antisense RNA called *cis-NAT$_{PHO1;2}$*. Unlike *cis-NAT*, whose primary role is to inhibit expression (Borsani et al., 2005; Zhou et al., 2009), Pi starvation-induced *cis-NAT$_{PHO1;2}$* enhances the translation of *PHO1;2* and facilitates Pi transport from roots to shoots (Jabnoune et al., 2013). A recent study adopted *in vitro* SHAPE-MaP to analyze the RNA secondary structure of *PHO1;2* and *cis-NAT$_{PHO1;2}$* individually or together (Reis et al., 2021) (Figure 10). Surprisingly, there was no evidence of widespread inter-molecular RNA-RNA interactions. Instead, compared with *PHO1;2* alone, the RNA structure of a specific local region covering the third and fourth exons of *PHO1;2* was rearranged in the presence of *cis-NAT$_{PHO1;2}$*. This rearranged region contains an unusually high GC content, which suggests the formation of a structure that inhibits *PHO1;2* translation in the absence of *cis-NAT$_{PHO1;2}$*. The authors hypothesized a model in which *cis-NAT$_{PHO1;2}$* interacts with *PHO1;2* mRNA to alter the RNA structure of the high GC inhibitory region, thereby enhancing the accessibility of *PHO1;2* mRNA to 60S (Reis et al., 2021). Although this model requires the support of *in vivo* RNA structure data, the present study provides a possible mechanism for the function of *cis-NAT*s in translation enhancement through dynamic RNA structure switch.

Future efforts in studying plant RNA structure could be extended to different plant species, especially crops, providing the potential to inform future crop improvement strategies. In the natural environment, plants are constantly challenged by diverse environmental conditions. In addition to temperature, light, and nutrients, future studies could focus on other agronomically relevant abiotic and biotic stresses, such as pathogens and microbial infections.

Concluding remarks and future perspectives

In summary, the rapid emergence of high-throughput sequencing and bioinformatics tools are now widely used to decode RNA structuromes and provide an ongoing stream of novel insights to elucidate their functions and mechanisms. Given the extent and diversity of newly available methods, the selection of appropriate analytical methods for different

experiments should be carefully assessed. Addressing specific biological questions requires further optimization and innovation. This review illustrates multiple examples of using different technologies, including global mapping of RNA secondary structures and RNA spatial conformations, and determination of RNA 3D structures. Some of the technical bottlenecks in studying RNA structures are discussed for further improvement.

In contrast to data resources for protein structures, structures for RNA-only and RNA-containing complexes are incredibly scarce. Furthermore, the inherent flexibility of RNA confers significant technological challenges and discloses the limitations of individual structural biology techniques, however, different biophysical, biochemical, and computational approaches can provide a wide range of heterogeneous structural information for RNAs to compute and refine structural models of RNA. Furthermore, the 3D structures can be validated, assessed, or filtered with the structural information from different sources, such as low-resolution density maps or shape structure from cryo-EM, AFM or SAS, or pair-wise distance restraints from biophysical (EPR, XSI, or smFRET) or biochemical approaches (XL or RPL), therefore further increasing the confidence of the final structural model. Also, the mechanisms of interactions and functions deduced from 3D model can be cross-validated by different mutations to enhance structural accuracy.

Moreover, artificial intelligence and deep learning techniques have recently been applied to solve problems in biology and other fields, including the prediction of molecular structures. AlphaFold, a new machine learning method, has been widely used to accurately predict protein structures at the atomic level (Jumper et al., 2021; Tunyasuvunakool et al., 2021). However, since few RNA structures are known, it is difficult to predict them computationally. Recently, a novel machine learning approach, ARES, was able to identify accurate RNA structural models after training with only 18 known RNA structures (Townshend et al., 2021). ARES is a suitable method to respond to the current situation where only a few known RNA structure repertoires exist. And, with the increase of available high-resolution RNA structures in the future, artificial intelligence and deep learning techniques will be a subversive prediction method for decoding accurate RNA structures.

Since the discovery of tRNA secondary structure, the functionalities of many RNA structures have been widely deciphered in microorganisms, animals and plants. In recent years, the exploration of non-coding RNAs has become a focus of research. They are no longer considered as "junk" transcripts, but have important physiological functions (Xue et al., 2020). Specifically, RNA structures are involved in the biogenesis mechanism and functional regulation of noncoding RNAs. Future directions may require more attention to elucidate the non-coding RNA structuromes and their functional roles.

Furthermore, the therapeutic targeting of RNA has currently attracted extensive interest from researchers. RNA-targeted therapeutics is considered as a potential strategy to modulate mRNA of many "undruggable" proteins or disease-implicated non-coding RNAs. The strategies of RNA-targeted therapeutics, such as oligonucleotide therapeutics and small molecule therapy targeting RNA, have been applied in cellular experiments and clinical treatments. Antisense technology as a type of oligonucleotide therapeutics is a brand-new drug design method developed in recent years to revert the effects of genetic mutations by complementary pairing of antisense molecules with target functional regions. Indeed, small molecules have emerged as another promising candidate for altering RNA function by targeting RNA structure. For example, small molecules targeting the oncogenic RNA hairpin precursor could recruit nucleases to cleave the oncogenic RNA and thus modulate oncogenic pathways (Costales et al., 2020; Costales et al., 2018). In addition, diverse naturally produced small molecules can target the bacterial riboswitch and thus regulate gene expression. These small molecules can be designed as potentially effective antibacterial agents. In the future, the functional RNA structure is expected to become a brand novel drug target for disease treatment. Recent studies have shown that RNA secondary structure design can effectively reduce the *in vitro* hydrolysis of RNA and improve the stability of mRNA vaccines. Beyond medical applications, further exploration and exploitation of RNA structure-guided molecular breeding will provide new avenues for crop improvement and sustainable agriculture.

References

Anderson, S.J., Kramer, M.C., Gosai, S.J., Yu, X., Vandivier, L.E., Nelson, A.D.L., Anderson, Z.D., Beilstein, M.A., Fray, R.G., Lyons, E., et al. (2018). N^6-methyladenosine inhibits local ribonucleolytic cleavage to

stabilize mRNAs in *Arabidopsis*. Cell Rep 25, 1146–1157.e3.

Andino, R., Rieckhof, G.E., and Baltimore, D. (1990). A functional ribonucleoprotein complex forms around the 5′ end of poliovirus RNA. Cell 63, 369–380.

Anger, A.M., Armache, J.P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and *Drosophila* 80S ribosome. Nature 497, 80–85.

Asamitsu, S., and Shioda, N. (2021). Potential roles of G-quadruplex structures in RNA granules for physiological and pathological phase separation. J Biochem 169, 527–533.

Aw, J.G.A., Shen, Y., Wilm, A., Sun, M., Lim, X.N., Boon, K.L., Tapsin, S., Chan, Y.S., Tan, C.P., Sim, A.Y.L., et al. (2016). *In vivo* mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. Mol Cell 62, 603–617.

Aw, J.G.A., Lim, S.W., Wang, J.X., Lambert, F.R.P., Tan, W.T., Shen, Y., Zhang, Y., Kaewsapsak, P., Li, C., Ng, S.B., et al. (2021). Determination of isoform-specific RNA structure with nanopore long reads. Nat Biotechnol 39, 336–346.

Balaratnam, S., Hettiarachchilage, M., West, N., Piontkivska, H., and Basu, S. (2019). A secondary structure within a human piRNA modulates its functionality. Biochimie 157, 72–80.

Baldwin, A., Morris, A.R., and Mukherjee, N. (2021). An easy, cost-effective, and scalable method to deplete human ribosomal RNA for RNA-seq. Curr Protoc 1, e176.

Barnwal, R.P., Yang, F., and Varani, G. (2017). Applications of NMR to structure determination of RNAs large and small. Arch Biochem Biophys 628, 42–56.

Beaudoin, J.D., Novoa, E.M., Vejnar, C.E., Yartseva, V., Takacs, C.M., Kellis, M., and Giraldez, A.J. (2018). Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. Nat Struct Mol Biol 25, 677–686.

Beaudoin, J.D., and Perreault, J.P. (2013). Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. Nucleic Acids Res 41, 5898–5911.

Bevilacqua, P.C., Ritchey, L.E., Su, Z., and Assmann, S.M. (2016). Genome-wide analysis of RNA secondary structure. Annu Rev Genet 50, 235–266.

Bhattacharyya, S., Jacobs, W.M., Adkar, B.V., Yan, J., Zhang, W., and Shakhnovich, E.I. (2018). Accessibility of the shine-dalgarno sequence dictates N-terminal codon bias in *E. coli*. Mol Cell 70, 894–905.e5.

Biffi, G., Di Antonio, M., Tannahill, D., and Balasubramanian, S. (2014). Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat Chem 6, 75–80.

Boerneke, M.A., Ehrhardt, J.E., and Weeks, K.M. (2019). Physical and functional analysis of viral RNA genomes by SHAPE. Annu Rev Virol 6, 93–117.

Bogdanow, B., Wang, X., Eichelbaum, K., Sadewasser, A., Husic, I., Paki, K., Budt, M., Hergeselle, M., Vetter, B., Hou, J., et al. (2019). The dynamic proteome of influenza A virus infection identifies M segment splicing as a host range determinant. Nat Commun 10, 5518.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell 123, 1279–1291.

Bose, K., Lech, C.J., Heddi, B., and Phan, A.T. (2018). High-resolution AFM structure of DNA G-wires in aqueous solution. Nat Commun 9, 1959..

Brázda, V., Hároníková, L., Liao, J.C.C., and Fojta, M. (2014). DNA and RNA quadruplex-binding proteins. Int J Mol Sci 15, 17493–17517.

Breaker, R.R. (2018). Riboswitches and translation control. Cold Spring Harb Perspect Biol 10, a032797.

Brierley, I., Digard, P., and Inglis, S.C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell 57, 537–547.

Brooks, B.R., Brooks Iii, C.L., Mackerell Jr., A.D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al. (2009). CHARMM: the biomolecular simulation program. J Comput Chem 30, 1545–1614.

Brosey, C.A., and Tainer, J.A. (2019). Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. Curr Opin Struct Biol 58, 197–213.

Brown, J.D., Kharytonchyk, S., Chaudry, I., Iyer, A.S., Carter, H., Becker, G., Desai, Y., Glang, L., Choi, S.H., Singh, K., et al. (2020). Structural basis for transcriptional start site control of HIV-1 RNA fate. Science 368, 413–417.

Bullock, S.L., Ringel, I., Ish-Horowicz, D., and Lukavsky, P.J. (2010). A′-form RNA helices are required for cytoplasmic mRNA transport in *Drosophila*. Nat Struct Mol Biol 17, 703–709.

Burrill, C.P., Westesson, O., Schulte, M.B., Strings, V.R., Segal, M., and Andino, R. (2013). Global RNA structure analysis of poliovirus identifies a conserved RNA structure involved in viral replication and infectivity. J Virol 87, 11670–11683.

Cai, Z., Cao, C., Ji, L., Ye, R., Wang, D., Xia, C., Wang, S., Du, Z., Hu, N., Yu, X., et al. (2020). RIC-seq for global *in situ* profiling of RNA-RNA spatial interactions. Nature 582, 432–437.

Cammas, A., and Millevoi, S. (2017). RNA G-quadruplexes: emerging mechanisms in disease. Nucleic Acids Res 45, 1584–1595.

Cao, C., Cai, Z., Xiao, X., Rao, J., Chen, J., Hu, N., Yang, M., Xing, X., Wang, Y., Li, M., et al. (2021a). The architecture of the SARS-CoV-2 RNA genome inside virion. Nat Commun 12, 3917.

Cao, C., Cai, Z., Ye, R., Su, R., Hu, N., Zhao, H., and Xue, Y. (2021b). Global in situ profiling of RNA-RNA spatial interactions with RIC-seq. Nat Protoc 16, 2916–2946.

Carroni, M., and Saibil, H.R. (2016). Cryo electron microscopy to determine the structure of macromolecular complexes. Methods 95, 78–85.

Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005). The Amber biomolecular simulation programs. J Comput Chem 26, 1668–1688.

Chan, C.Y., and Kwok, C.K. (2020). Specific Binding of a d-RNA G-quadruplex structure with an l-RNA aptamer. Angew Chem Int Ed 59, 5293–5297.

Chartrand, P., Meng, X.H., Singer, R.H., and Long, R.M. (1999). Structural elements required for the localization of ASH1 mRNA and of a green fluorescent protein reporter particle *in vivo*. Curr Biol 9, 333–338.

Chaves, R.C., and Pellequer, J.L. (2013). DockAFM: benchmarking protein structures by docking under AFM topographs. Bioinformatics 29, 3230–3231.

Chen, X., Li, Y., Umarov, R., Gao, X., and Song, L. (2020). RNA secondary structure prediction by learning unrolled algorithms. ArXiv, abs/2002.05810.

Cho, H., Cho, H.S., Nam, H., Jo, H., Yoon, J., Park, C., Dang, T.V.T., Kim, E., Jeong, J., Park, S., et al. (2018). Translational control of phloem development by RNA G-quadruplex-JULGI determines plant sink strength. Nat Plants 4, 376–390.

Chow, E.Y.C., Lyu, K., Kwok, C.K., and Chan, T.F. (2020). rG4-seeker enables high-confidence identification of novel and non-canonical rG4 motifs from rG4-seq experiments. RNA Biol 17, 903–917.

Christy, T.W., Giannetti, C.A., Houlihan, G., Smola, M.J., Rice, G.M., Wang, J., Dokholyan, N.V., Laederach, A., Holliger, P., and Weeks, K. M. (2021). Direct mapping of higher-order RNA interactions by SHAPE-JuMP. Biochemistry 60, 1971–1982.

Chu, C., Quinn, J., and Chang, H.Y. (2012). Chromatin isolation by RNA purification (ChIRP). J Vis Exp 61, 3912.

Chung, B.Y.W., Balcerowicz, M., Di Antonio, M., Jaeger, K.E., Geng, F., Franaszek, K., Marriott, P., Brierley, I., Firth, A.E., and Wigge, P.A. (2020). An RNA thermoswitch regulates daytime growth in *Arabidopsis*. Nat Plants 6, 522–532.

Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A., and Goodall, G.J. (2015). The RNA binding protein quaking regulates formation of circRNAs. Cell 160, 1125–1134.

Costales, M.G., Matsumoto, Y., Velagapudi, S.P., and Disney, M.D. (2018).

Small molecule targeted recruitment of a nuclease to RNA. J Am Chem Soc 140, 6741–6744.

Costales, M.G., Aikawa, H., Li, Y., Childs-Disney, J.L., Abegg, D., Hoch, D.G., Pradeep Velagapudi, S., Nakai, Y., Khan, T., Wang, K.W., et al. (2020). Small-molecule targeted recruitment of a nuclease to cleave an oncogenic RNA in a mouse model of metastatic cancer. Proc Natl Acad Sci USA 117, 2406–2411.

Croft, M.T., Moulin, M., Webb, M.E., and Smith, A.G. (2007). Thiamine biosynthesis in algae is regulated by riboswitches. Proc Natl Acad Sci USA 104, 20770–20775.

Dadonaite, B., Gilbertson, B., Knight, M.L., Trifkovic, S., Rockman, S., Laederach, A., Brown, L.E., Fodor, E., and Bauer, D.L.V. (2019). The structure of the influenza A virus genome. Nat Microbiol 4, 1781–1789.

Damgaard, C.K., Andersen, E.S., Knudsen, B., Gorodkin, J., and Kjems, J. (2004). RNA interactions in the 5′ region of the HIV-1 genome. J Mol Biol 336, 369–379.

Danev, R., Yanagisawa, H., and Kikkawa, M. (2019). Cryo-electron microscopy methodology: current aspects and future directions. Trends Biochem Sci 44, 837–848.

Das, R., and Baker, D. (2007). Automated *de novo* prediction of native-like RNA tertiary structures. Proc Natl Acad Sci USA 104, 14664–14669.

Das, R., Karanicolas, J., and Baker, D. (2010). Atomic accuracy in predicting and designing noncanonical RNA structure. Nat Methods 7, 291–294.

Deng, H., Cheema, J., Zhang, H., Woolfenden, H., Norris, M., Liu, Z., Liu, Q., Yang, X., Yang, M., Deng, X., et al. (2018). Rice *in vivo* RNA structurome reveals RNA secondary structure conservation and divergence in plants. Mol Plant 11, 607–622.

Diaz-Toledano, R., Lozano, G., and Martinez-Salas, E. (2017). In-cell SHAPE uncovers dynamic interactions between the untranslated regions of the foot-and-mouth disease virus RNA. Nucleic Acids Res 45, 1416–1432.

Didiot, M.C., Tian, Z., Schaeffer, C., Subramanian, M., Mandel, J.L., and Moine, H. (2008). The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. Nucleic Acids Res 36, 4902–4912.

Dimura, M., Peulen, T.O., Hanke, C.A., Prakash, A., Gohlke, H., and Seidel, C.A. (2016). Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. Curr Opin Struct Biol 40, 163–185.

Dimura, M., Peulen, T.O., Sanabria, H., Rodnin, D., Hemmen, K., Hanke, C.A., Seidel, C.A.M., and Gohlke, H. (2020). Automated and optimally FRET-assisted structural modeling. Nat Commun 11, 5394.

Ding, Y., Kwok, C.K., Tang, Y., Bevilacqua, P.C., and Assmann, S.M. (2015). Genome-wide profiling of *in vivo* RNA structure at single-nucleotide resolution using structure-seq. Nat Protoc 10, 1050–1066.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature 505, 696–700.

Domnick, C., Eggert, F., Wuebben, C., Bornewasser, L., Hagelueken, G., Schiemann, O., and Kath-Schorr, S. (2020). EPR distance measurements on long non-coding RNAs empowered by genetic alphabet expansion transcription. Angew Chem Int Ed 59, 7891–7896.

Dong, H., Li, L., Zhu, X., Shi, J., Fu, Y., Zhang, S., Shi, Y., Xu, B., Zhang, J., Shi, F., et al. (2021). Complex RNA secondary structures mediate mutually exclusive splicing of coleoptera *Dscam1*. Front Genet 12, 644238.

Dong, H., Xu, B., Guo, P., Zhang, J., Yang, X., Li, L., Fu, Y., Shi, J., Zhang, S., Zhu, Y., et al. (2022). Hidden RNA pairings counteract the "first-come, first-served" splicing principle to drive stochastic choice in *Dscam1* splice variants. Sci Adv 8, eabm1763.

Dumas, L., Herviou, P., Dassi, E., Cammas, A., and Millevoi, S. (2021). G-quadruplexes in RNA biology: recent advances and future directions. Trends Biochem Sci 46, 270–283.

Dumetz, F., Chow, E.Y.C., Harris, L.M., Liew, S.W., Jensen, A., Umar, M.I., Chung, B., Chan, T.F., Merrick, C.J., and Kwok, C.K. (2021). G-quadruplex RNA motifs influence gene expression in the malaria parasite *Plasmodium falciparum*. Nucleic Acids Res 49, 12486–12501.

Duss, O., Yulikov, M., Jeschke, G., and Allain, F.H.T. (2014). EPR-aided approach for solution structure determination of large RNAs or protein-RNA complexes. Nat Commun 5, 3669.

Einarson, O.J., and Sen, D. (2017). Self-biotinylation of DNA G-quadruplexes via intrinsic peroxidase activity. Nucleic Acids Res 45, 9813–9822.

Engel, A., and Müller, D.J. (2000). Observing single biomolecules at work with the atomic force microscope. Nat Struct Biol 7, 715–718.

Falabella, M., Fernandez, R.J., Johnson, F.B., and Kaufman, B.A. (2019). Potential roles for G-quadruplexes in mitochondria. Curr Med Chem 26, 2918–2932.

Fang, X., Wang, J., O′Carroll, I.P., Mitchell, M., Zuo, X., Wang, Y., Yu, P., Liu, Y., Rausch, J.W., Dyba, M.A., et al. (2013). An unusual topological structure of the HIV-1 Rev response element. Cell 155, 594–605.

Fang, X., Stagno, J.R., Bhandari, Y.R., Zuo, X., and Wang, Y.X. (2015). Small-angle X-ray scattering: a bridge between RNA secondary structures and three-dimensional topological structures. Curr Opin Struct Biol 30, 147–160.

Fang, X., Michnicka, M., Zhang, Y., Wang, Y.X., and Nikonowicz, E.P. (2017). Capture and release of tRNA by the T-loop receptor in the function of the T-box riboswitch. Biochemistry 56, 3549–3558.

Fay, M.M., Anderson, P.J., and Ivanov, P. (2017a). ALS/FTD-associated C9ORF72 repeat RNA promotes phase transitions *in vitro* and in cells. Cell Rep 21, 3573–3584.

Fay, M.M., Lyons, S.M., and Ivanov, P. (2017b). RNA G-quadruplexes in biology: principles and molecular mechanisms. J Mol Biol 429, 2127–2147.

Fleming, A.M., Ding, Y., Alenko, A., and Burrows, C.J. (2016). Zika virus genomic RNA possesses conserved G-quadruplexes characteristic of the flaviviridae family. ACS Infect Dis 2, 674–681.

Flores, S.C., and Altman, R.B. (2010). Turning limited experimental information into 3D models of RNA. RNA 16, 1769–1778.

Fodor, E., Pritlove, D.C., and Brownlee, G.G. (1994). The influenza virus panhandle is involved in the initiation of transcription. J Virol 68, 4092–4096.

Foley, S.W., Gosai, S.J., Wang, D., Selamoglu, N., Sollitti, A.C., Köster, T., Steffen, A., Lyons, E., Daldal, F., Garcia, B.A., et al. (2017). A global view of RNA-protein interactions identifies post-transcriptional regulators of root hair cell fate. Dev Cell 41, 204–220.e5.

Foley, S.W., and Gregory, B.D. (2016). Protein interaction profile sequencing (PIP-seq). Curr Protoc Mol Biol 116, 27.25.21.

Foley, S.W., Vandivier, L.E., Kuksa, P.P., and Gregory, B.D. (2015). Transcriptome-wide measurement of plant RNA secondary structure. Curr Opin Plant Biol 27, 36–43.

Fu, L., Cao, Y., Wu, J., Peng, Q., Nie, Q., and Xie, X. (2022). UFold: fast and accurate RNA secondary structure prediction with deep learning. Nucleic Acids Res 50, e14.

García-Sacristán, A., Moreno, M., Ariza-Mateos, A., López-Camacho, E., Jáudenes, R.M., Vázquez, L., Gómez, J., Martín-Gago, J.Á., and Briones, C. (2015). A magnesium-induced RNA conformational switch at the internal ribosome entry site of hepatitis C virus genome visualized by atomic force microscopy. Nucleic Acids Res 43, 565–580.

Garman, E.F. (2014). Developments in X-ray crystallographic structure determination of biological macromolecules. Science 343, 1102–1108.

Gawroński, P., Enroth, C., Kindgren, P., Marquardt, S., Karpiński, S., Leister, D., Jensen, P.E., Vinther, J., and Scharff, L.B. (2021). Light-dependent translation change of *Arabidopsis psbA* correlates with RNA structure alterations at the translation initiation region. Cells 10, 322.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish miR-430 promotes deadenylation and clearance of maternal mRNAs. Science 312, 75–79.

Gong, Z., Xiong, L., Shi, H., Yang, S., Herrera-Estrella, L.R., Xu, G., Chao, D.Y., Li, J., Wang, P.Y., Qin, F., et al. (2020). Plant abiotic stress response and nutrient use efficiency. Sci China Life Sci 63, 635–674.

Goodfellow, I., Chaudhry, Y., Richardson, A., Meredith, J., Almond, J.W.,

Barclay, W., and Evans, D.J. (2000). Identification of a *cis*-acting replication element within the poliovirus coding region. J Virol 74, 4590–4600.

Gosai, S.J., Foley, S.W., Wang, D., Silverman, I.M., Selamoglu, N., Nelson, A.D.L., Beilstein, M.A., Daldal, F., Deal, R.B., and Gregory, B.D. (2015). Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the *Arabidopsis* nucleus. Mol Cell 57, 376–388.

Graveley, B.R. (2005). Mutually exclusive splicing of the insect *Dscam* pre-mRNA directed by competing intronic RNA secondary structures. Cell 123, 65–73.

Gros, J., Guédin, A., Mergny, J.L., and Lacroix, L. (2008). G-quadruplex formation interferes with P1 helix formation in the RNA component of telomerase hTERC. Chembiochem 9, 2075–2079.

Guo, J.U., and Bartel, D.P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science 353, aaf5371.

Harris, M.E., and Christian, E.L. (2009). RNA crosslinking methods. In: Methods in Enzymology. New York: Academic Press. 127–146.

Hawkes, E.J., Hennelly, S.P., Novikova, I.V., Irwin, J.A., Dean, C., and Sanbonmatsu, K.Y. (2016). COOLAIR antisense RNAs form evolutionarily conserved elaborate secondary structures. Cell Rep 16, 3087–3096.

Helm, M., Kobitski, A.Y., and Nienhaus, G.U. (2009). Single-molecule Förster resonance energy transfer studies of RNA structure, dynamics and function. Biophys Rev 1, 161–176.

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153, 654–665.

Herold, J., and Andino, R. (2001). Poliovirus RNA replication requires genome circularization through a protein-protein bridge. Mol Cell 7, 581–591.

Herviou, P., Le Bras, M., Dumas, L., Hieblot, C., Gilhodes, J., Cioci, G., Hugnot, J.P., Ameadan, A., Guillonneau, F., Dassi, E., et al. (2020). hnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma. Nat Commun 11, 2661.

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., and Zamir, A. (1965). Structure of a ribonucleic acid. Science 147, 1462–1465.

Holmqvist, E., Li, L., Bischler, T., Barquist, L., and Vogel, J. (2018). Global maps of ProQ binding *in vivo* reveal target recognition via RNA structure and stability control at mRNA 3′ ends. Mol Cell 70, 971–982.e6.

Homan, P.J., Tandon, A., Rice, G.M., Ding, F., Dokholyan, N.V., and Weeks, K.M. (2014). RNA tertiary structure analysis by 2′-hydroxyl molecular interference. Biochemistry 53, 6825–6833.

Hong, W., Shi, Y., Xu, B., and Jin, Y. (2020). RNA secondary structures in *Dscam1* mutually exclusive splicing: unique evolutionary signature from the midge. RNA 26, 1086–1093.

Hong, W., Zhang, J., Dong, H., Shi, Y., Ma, H., Zhou, F., Xu, B., Fu, Y., Zhang, S., Hou, S., et al. (2021). Intron-targeted mutagenesis reveals roles for *Dscam1* RNA pairing architecture-driven splicing bias in neuronal wiring. Cell Rep 36, 109373.

Howe, J.A., Wang, H., Fischmann, T.O., Balibar, C.J., Xiao, L., Galgoci, A.M., Malinverni, J.C., Mayhood, T., Villafania, A., Nahvi, A., et al. (2015). Selective small-molecule inhibition of an RNA structural element. Nature 526, 672–677.

Huang, H., Suslov, N.B., Li, N.S., Shelke, S.A., Evans, M.E., Koldobskaya, Y., Rice, P.A., and Piccirilli, J.A. (2014). A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. Nat Chem Biol 10, 686–691.

Huber, R.G., Lim, X.N., Ng, W.C., Sim, A.Y.L., Poh, H.X., Shen, Y., Lim, S.Y., Sundstrom, K.B., Sun, X., Aw, J.G., et al. (2019). Structure mapping of dengue and Zika viruses reveals functional long-range interactions. Nat Commun 10, 1408.

Hubstenberger, A., Courel, M., Bénard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J.B., Munier, A., Fradet, M., et al. (2017). P-body purification reveals the condensation of repressed mRNA regulons. Mol Cell 68, 144–157.e5.

Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. Nucleic Acids Res 33, 2908–2916.

Hura, G.L., Tsai, C.L., Claridge, S.A., Mendillo, M.L., Smith, J.M., Williams, G.J., Mastroianni, A.J., Alivisatos, A.P., Putnam, C.D., Kolodner, R.D., et al. (2013). DNA conformations in mismatch repair probed in solution by X-ray scattering from gold nanocrystals. Proc Natl Acad Sci USA 110, 17308–17313.

Ilik, I.A., Quinn, J.J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R.C., Luscombe, N., Backofen, R., et al. (2013). Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. Mol Cell 51, 156–173.

Incarnato, D., Neri, F., Anselmi, F., and Oliviero, S. (2014). Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. Genome Biol 15, 491.

Ishiguro, A., Kimura, N., Watanabe, Y., Watanabe, S., and Ishihama, A. (2016). TDP-43 binds and transports G-quadruplex-containing mRNAs into neurites for local translation. Genes Cells 21, 466–481.

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., Dieterich, C., et al. (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. Cell Rep 10, 170–177.

Jabnoune, M., Secco, D., Lecampion, C., Robaglia, C., Shu, Q., and Poirier, Y. (2013). A rice *cis*-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. Plant Cell 25, 4166–4182.

Jacques, D.A., and Trewhella, J. (2010). Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. Protein Sci 19, 642–657.

Jakobsen, U., Shelke, S.A., Vogel, S., and Sigurdsson, S.T. (2010). Site-directed spin-labeling of nucleic acids by click chemistry: detection of abasic sites in duplex DNA by EPR spectroscopy. J Am Chem Soc 132, 10424–10428.

Ji, D., Juhas, M., Tsang, C.M., Kwok, C.K., Li, Y., and Zhang, Y. (2021). Discovery of G-quadruplex-forming sequences in SARS-CoV-2. Brief Bioinform 22, 1150–1160.

Jiang, L., Shao, C., Wu, Q.J., Chen, G., Zhou, J., Yang, B., Li, H., Gou, L.T., Zhang, Y., Wang, Y., et al. (2017). NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing. Nat Struct Mol Biol 24, 816–824.

Jodoin, R., Bauer, L., Garant, J.M., Mahdi Laaref, A., Phaneuf, F., and Perreault, J.P. (2014). The folding of 5′-UTR human G-quadruplexes possessing a long central loop. RNA 20, 1129–1141.

Jones, C.P., and Ferré-D'Amaré, A.R. (2017). Long-range interactions in riboswitch control of gene expression. Annu Rev Biophys 46, 455–481.

Jonikas, M.A., Radmer, R.J., Laederach, A., Das, R., Pearlman, S., Herschlag, D., and Altman, R.B. (2009). Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. RNA 15, 189–199.

Jossinet, F., Ludwig, T.E., and Westhof, E. (2010). Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. Bioinformatics 26, 2057–2059.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.

Kappel, K., Zhang, K., Su, Z., Watkins, A.M., Kladwang, W., Li, S., Pintilie, G., Topkar, V.V., Rangan, R., Zheludev, I.N., et al. (2020). Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. Nat Methods 17, 699–707.

Karaca, E., Rodrigues, J.P.G.L.M., Graziadei, A., Bonvin, A.M.J.J., and Carlomagno, T. (2017). M3: an integrative framework for structure determination of molecular machines. Nat Methods 14, 897–902.

Karsisiotis, A.I., Hessari, N.M., Novellino, E., Spada, G.P., Randazzo, A.,

and Webba da Silva, M. (2011). Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. Angew Chem Int Ed 50, 10645–10648.

Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y., and Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. Nature 467, 103–107.

Kerzhner, M., Abdullin, D., Więcek, J., Matsuoka, H., Hagelueken, G., Schiemann, O., and Famulok, M. (2016). Post-synthetic spin-labeling of RNA through click chemistry for PELDOR measurements. Chem Eur J 22, 12113–12121.

Kerzhner, M., Matsuoka, H., Wuebben, C., Famulok, M., and Schiemann, O. (2018). High-yield spin labeling of long RNAs for electron paramagnetic resonance spectroscopy. Biochemistry 57, 2923–2931.

Kharel, P., Balaratnam, S., Beals, N., and Basu, S. (2020). The role of RNA G-quadruplexes in human diseases and therapeutic strategies. WIREs RNA 11, e1568.

Khawaja, A., Vopalensky, V., and Pospisek, M. (2015). Understanding the potential of hepatitis C virus internal ribosome entry site domains to modulate translation initiation via their structure and function. WIREs RNA 6, 211–224.

Khong, A., Jain, S., Matheny, T., Wheeler, J.R., and Parker, R. (2018). Isolation of mammalian stress granule cores for RNA-Seq analysis. Methods 137, 49–54.

Kim, S.H., Suddath, F.L., Quigley, G.J., McPherson, A., Sussman, J.L., Wang, A.H.J., Seeman, N.C., and Rich, A. (1974). Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. Science 185, 435–440.

Kooshapur, H., Choudhury, N.R., Simon, B., Mühlbauer, M., Jussupow, A., Fernandez, N., Jones, A.N., Dallmann, A., Gabel, F., Camilloni, C., et al. (2018). Structural basis for terminal loop recognition and stimulation of pri-miRNA-18a processing by hnRNP A1. Nat Commun 9, 2479.

Kortmann, J., and Narberhaus, F. (2012). Bacterial RNA thermometers: molecular zippers and switches. Nat Rev Microbiol 10, 255–265.

Kramer, M.C., Janssen, K.A., Palos, K., Nelson, A.D.L., Vandivier, L.E., Garcia, B.A., Lyons, E., Beilstein, M.A., and Gregory, B.D. (2020). N^6-methyladenosine and RNA secondary structure affect transcript stability and protein abundance during systemic salt stress in *Arabidopsis*. Plant Direct 4, e00239.

Kristensen, L.S., Andersen, M.S., Stagsted, L.V.W., Ebbesen, K.K., Hansen, T.B., and Kjems, J. (2019). The biogenesis, biology and characterization of circular RNAs. Nat Rev Genet 20, 675–691.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell 31, 147–157.

Kudla, G., Granneman, S., Hahn, D., Beggs, J.D., and Tollervey, D. (2011). Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. Proc Natl Acad Sci USA 108, 10010–10015.

Kudla, G., Wan, Y., and Helwak, A. (2020). RNA conformation capture by proximity ligation. Annu Rev Genom Hum Genet 21, 81–100.

Kwok, C.K., and Balasubramanian, S. (2015). Targeted detection of G-quadruplexes in cellular RNAs. Angew Chem Int Ed 54, 6751–6754.

Kwok, C.K., Ding, Y., Shahid, S., Assmann, S.M., and Bevilacqua, P.C. (2015). A stable RNA G-quadruplex within the 5′-UTR of *Arabidopsis thaliana ATR* mRNA inhibits translation. Biochem J 467, 91–102.

Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S., and Balasubramanian, S. (2016a). rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat Methods 13, 841–844.

Kwok, C.K., and Merrick, C.J. (2017). G-quadruplexes: prediction, characterization, and biological application. Trends Biotechnol 35, 997–1013.

Kwok, C.K., Sahakyan, A.B., and Balasubramanian, S. (2016b). Structural analysis using SHALiPE to reveal RNA G-quadruplex formation in human precursor microRNA. Angew Chem Int Ed 55, 8958–8961.

Lat, P.K., Liu, K., Kumar, D.N., Wong, K.K.L., Verheyen, E.M., and Sen, D. (2020). High specificity and tight spatial restriction of self-

biotinylation by DNA and RNA G-Quadruplexes complexed *in vitro* and *in vivo* with Heme. Nucleic Acids Res 48, 5254–5267.

Le Sage, V., Kanarek, J.P., Snyder, D.J., Cooper, V.S., Lakdawala, S.S., and Lee, N. (2020). Mapping of influenza virus RNA-RNA interactions reveals a flexible network. Cell Rep 31, 107823.

Leamy, K.A., Assmann, S.M., Mathews, D.H., and Bevilacqua, P.C. (2016). Bridging the gap between *in vitro* and *in vivo* RNA folding. Quart Rev Biophys 49, e10.

Li, F., Zheng, Q., Ryvkin, P., Dragomir, I., Desai, Y., Aiyer, S., Valladares, O., Yang, J., Bambina, S., Sabin, L.R., et al. (2012a). Global analysis of RNA secondary structure in two metazoans. Cell Rep 1, 69–82.

Li, F., Zheng, Q., Vandivier, L.E., Willmann, M.R., Chen, Y., and Gregory, B.D. (2012b). Regulatory impact of RNA secondary structure across the *Arabidopsis* transcriptome. Plant Cell 24, 4346–4359.

Li, P., Wei, Y., Mei, M., Tang, L., Sun, L., Huang, W., Zhou, J., Zou, C., Zhang, S., Qin, C.F., et al. (2018). Integrative analysis of Zika virus genome RNA structure reveals critical determinants of viral infectivity. Cell Host Microbe 24, 875–886.e5.

Li, X., Liang, Q.X., Lin, J.R., Peng, J., Yang, J.H., Yi, C., Yu, Y., Zhang, Q. C., and Zhou, K.R. (2020). Epitranscriptomic technologies and analyses. Sci China Life Sci 63, 501–515.

Li, X., Zhang, J.L., Lei, Y.N., Liu, X.Q., Xue, W., Zhang, Y., Nan, F., Gao, X., Zhang, J., Wei, J., et al. (2021). Linking circular intronic RNA degradation and function in transcription by RNase H1. Sci China Life Sci 64, 1795–1809.

Liang, D., and Wilusz, J.E. (2014). Short intronic repeat sequences facilitate circular RNA production. Genes Dev 28, 2233–2247.

Liu, G.Q., Park, H.S., Pyo, H.M., Liu, Q., and Zhou, Y. (2015). Influenza A virus panhandle structure is directly involved in RIG-I activation and interferon induction. J Virol 89, 6067–6079.

Liu, Z., Liu, Q., Yang, X., Zhang, Y., Norris, M., Chen, X., Cheema, J., Zhang, H., and Ding, Y. (2021). *In vivo* nuclear RNA structurome reveals RNA-structure regulation of mRNA processing in plants. Genome Biol 22, 11.

Liu, Z., Reches, M., Groisman, I., and Engelberg-Kulka, H. (1998). The nature of the minimal 'selenocysteine insertion sequence' (SECIS) in *Escherichia coli*. Nucleic Acids Res 26, 896–902.

Loke, J.C., Stahlberg, E.A., Strenski, D.G., Haas, B.J., Wood, P.C., and Li, Q.Q. (2005). Compilation of mRNA polyadenylation signals in *Arabidopsis* revealed a new signal element and potential secondary structures. Plant Physiol 138, 1457–1468.

Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA package 2.0. Algorithms Mol Biol 6, 26.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., et al. (2016). RNA duplex map in living cells reveals higher-order transcriptome structure. Cell 165, 1267–1279.

Lucks, J.B., Mortimer, S.A., Trapnell, C., Luo, S., Aviran, S., Schroth, G.P., Pachter, L., Doudna, J.A., and Arkin, A.P. (2011). Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc Natl Acad Sci USA 108, 11063–11068.

Lund, P.E., Chatterjee, S., Daher, M., and Walter, N.G. (2020). Protein unties the pseudoknot: S1-mediated unfolding of RNA higher order structure. Nucleic Acids Res 48, 2107–2125.

Luo, Q.J., Zhang, J., Li, P., Wang, Q., Zhang, Y., Roy-Chaudhuri, B., Xu, J., Kay, M.A., and Zhang, Q.C. (2021). RNA structure probing reveals the structural basis of Dicer binding and cleavage. Nat Commun 12, 3397.

Lyons, S.M., Gudanis, D., Coyne, S.M., Gdaniec, Z., and Ivanov, P. (2017). Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. Nat Commun 8, 1127.

Lyu, K., Chen, S.B., Chan, C.Y., Tan, J.H., and Kwok, C.K. (2019). Structural analysis and cellular visualization of APP RNA G-quadruplex. Chem Sci 10, 11095–11102.

Lyu, K., Chow, E.Y.C., Mou, X., Chan, T.F., and Kwok, C.K. (2021). RNA G-quadruplexes (rG4s): genomics and biological functions. Nucleic

Acids Res 49, 5426–5450.

Ma, B., Zhang, L., Gao, Q., Wang, J., Li, X., Wang, H., Liu, Y., Lin, H., Liu, J., Wang, X., et al. (2021). A plasma membrane transporter coordinates phosphate reallocation and grain filling in cereals. Nat Genet 53, 906–915.

Macdonald, P.M., Kerr, K., Smith, J.L., and Leask, A. (1993). RNA regulatory element BLE1 directs the early steps of bicoid mRNA localization. Development 118, 1233–1243.

Mahieu, E., and Gabel, F. (2018). Biological small-angle neutron scattering: recent results and development. Acta Crystlogr D Struct Biol 74, 715–726.

Mailler, E., Paillart, J.C., Marquet, R., Smyth, R.P., and Vivet-Boudou, V. (2019). The evolution of RNA structural probing methods: from gels to next-generation sequencing. WIREs RNA 10, e1518.

Manfredonia, I., Nithin, C., Ponce-Salvatierra, A., Ghosh, P., Wirecki, T.K., Marinus, T., Ogando, N.S., Snijder, E.J., van Hemert, M.J., Bujnicki, J. M., et al. (2020). Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. Nucleic Acids Res 48, 12436–12452.

Manzano, M., Reichert, E.D., Polo, S., Falgout, B., Kasprzak, W., Shapiro, B.A., and Padmanabhan, R. (2011). Identification of *cis*-acting elements in the 3′-untranslated region of the dengue virus type 2 RNA that modulate translation and replication. J Biol Chem 286, 22521–22534.

Marin, M., Feeney, D.S., Brown, L.K., Naveed, M., Ruiz, S., Koebernick, N., Bengough, A.G., Hallett, P.D., Roose, T., Puértolas, J., et al. (2021). Significance of root hairs for plant performance under contrasting field conditions and water deficit. Ann Bot 128, 1–16.

Marinus, T., Fessler, A.B., Ogle, C.A., and Incarnato, D. (2021). A novel SHAPE reagent enables the analysis of RNA structure in living cells with unprecedented accuracy. Nucleic Acids Res 49, e34.

Martin, K.C., and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. Cell 136, 719–730.

Martinez, H.M., Maizel Jr, J.V., and Shapiro, B.A. (2008). RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. J Biomol Struct Dyn 25, 669–683.

Marz, M., Beerenwinkel, N., Drosten, C., Fricke, M., Frishman, D., Hofacker, I.L., Hoffmann, D., Middendorf, M., Rattei, T., Stadler, P.F., et al. (2014). Challenges in RNA virus bioinformatics. Bioinformatics 30, 1793–1799.

Mathew-Fenn, R.S., Das, R., and Harbury, P.A.B. (2008). Remeasuring the double helix. Science 322, 446–449.

Mathews, D.H. (2014). Using the RNA structure software package to predict conserved RNA structures. Curr Protoc Bioinformatics 46.

Mauger, D.M., Golden, M., Yamane, D., Williford, S., Lemon, S.M., Martin, D.P., and Weeks, K.M. (2015). Functionally conserved architecture of hepatitis C virus RNA genomes. Proc Natl Acad Sci USA 112, 3692–3697.

McCown, P.J., Corbino, K.A., Stav, S., Sherlock, M.E., and Breaker, R.R. (2017). Riboswitch diversity and distribution. RNA 23, 995–1011.

McCown, P.J., Wang, M.C., Jaeger, L., and Brown, J.A. (2019). Secondary structural model of human MALAT1 reveals multiple structure-function relationships. Int J Mol Sci 20, 5610.

Mehrshahi, P., Nguyen, G.T.D.T., Gorchs Rovira, A., Sayer, A., Llavero-Pasquina, M., Lim Huei Sin, M., Medcalf, E.J., Mendoza-Ochoa, G.I., Scaife, M.A., and Smith, A.G. (2020). Development of novel riboswitches for synthetic biology in the green alga chlamydomonas. ACS Synth Biol 9, 1406–1417.

Metkar, M., Ozadam, H., Lajoie, B.R., Imakaev, M., Mirny, L.A., Dekker, J., and Moore, M.J. (2018). Higher-order organization principles of pre-translational mRNPs. Mol Cell 72, 715–726.e3.

Meydan, S., Klepacki, D., Karthikeyan, S., Margus, T., Thomas, P., Jones, J.E., Khan, Y., Briggs, J., Dinman, J.D., Vázquez-Laslop, N., et al. (2017). Programmed ribosomal frameshifting generates a copper transporter and a copper chaperone from the same gene. Mol Cell 65, 207–219.

Mickelbart, M.V., Hasegawa, P.M., and Bailey-Serres, J. (2015). Genetic mechanisms of abiotic stress tolerance that translate to crop yield

stability. Nat Rev Genet 16, 237–251.

Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A., and Nudler, E. (2002). Sensing small molecules by nascent RNA. Cell 111, 747–756.

Moore, M.J.B., Schultes, C.M., Cuesta, J., Cuenca, F., Gunaratnam, M., Tanious, F.A., Wilson, W.D., and Neidle, S. (2006). Trisubstituted acridines as G-quadruplex telomere targeting agents. Effects of extensions of the 3,6- and 9-side chains on quadruplex binding, telomerase activity, and cell proliferation. J Med Chem 49, 582–599.

Morandi, E., Manfredonia, I., Simon, L.M., Anselmi, F., van Hemert, M.J., Oliviero, S., and Incarnato, D. (2021). Genome-scale deconvolution of RNA structure ensembles. Nat Methods 18, 249–252.

Moreno, M., Vázquez, L., López-Carrasco, A., Martín-Gago, J.A., Flores, R., and Briones, C. (2019). Direct visualization of the native structure of viroid RNAs at single-molecule resolution by atomic force microscopy. RNA Biol 16, 295–308.

Motte, H., Vanneste, S., and Beeckman, T. (2019). Molecular and environmental regulation of root development. Annu Rev Plant Biol 70, 465–488.

Mustoe, A.M., Busan, S., Rice, G.M., Hajdin, C.E., Peterson, B.K., Ruda, V.M., Kubica, N., Nutiu, R., Baryza, J.L., and Weeks, K.M. (2018). Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. Cell 173, 181–195.e18.

Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., and Breaker, R.R. (2002). Genetic control by a metabolite binding mRNA. Chem Biol 9, 1043–1049.

Nechooshtan, G., Elgrably-Weiss, M., and Altuvia, S. (2014). Changes in transcriptional pausing modify the folding dynamics of the pH-responsive RNA element. Nucleic Acids Res 42, 622–630.

Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E., and Altuvia, S. (2009). A pH-responsive riboregulator. Genes Dev 23, 2650–2662.

Neidle, S., and Balasubramanian, S. (2006). Quadruplex Nucleic Acids. Cambridge: Royal Society of Chemistry.

Nguyen, T.C., Cao, X., Yu, P., Xiao, S., Lu, J., Biase, F.H., Sridhar, B., Huang, N., Zhang, K., and Zhong, S. (2016). Mapping RNA-RNA interactome and RNA structure *in vivo* by MARIO. Nat Commun 7, 12023.

Niesters, H.G., and Strauss, J.H. (1990). Defined mutations in the 5′ nontranslated sequence of Sindbis virus RNA. J Virol 64, 4162–4168.

Niina, T., Fuchigami, S., and Takada, S. (2020). Flexible fitting of biomolecular structures to atomic force microscopy images via biased molecular simulations. J Chem Theor Comput 16, 1349–1358.

Niu, X., Sun, R., Chen, Z., Yao, Y., Zuo, X., Chen, C., and Fang, X. (2021). Pseudoknot length modulates the folding, conformational dynamics, and robustness of Xrn1 resistance of flaviviral xrRNAs. Nat Commun 12, 6417.

Nogales, E. (2016). The development of cryo-EM into a mainstream structural biology technique. Nat Methods 13, 24–27.

Ochsenreiter, R., Hofacker, I.L., and Wolfinger, M.T. (2019). Functional RNA structures in the 3′UTR of tick-borne, insect-specific and no-known-vector flaviviruses. Viruses 11, 298.

Oldroyd, G.E.D., and Leyser, O. (2020). A plant's diet, surviving in a variable nutrient environment. Science 368, eaba0196.

Olson, S.W., Turner, A.M.W., Arney, J.W., Saleem, I., Weidmann, C.A., Margolis, D.M., Weeks, K.M., and Mustoe, A.M. (2022). Discovery of a large-scale, cell-state-responsive allosteric switch in the 7SK RNA using DANCE-MaP. Mol Cell doi: 10.1016/j.molcel.2022.02.009.

Pallesen, J., Dong, M., Besenbacher, F., and Kjems, J. (2009). Structure of the HIV-1 Rev response element alone and in complex with regulator of virion (Rev) studied by atomic force microscopy. FEBS J 276, 4223–4232.

Paredes, E., Evans, M., and Das, S.R. (2011). RNA labeling, conjugation and ligation. Methods 54, 251–259.

Pennarun, G., Granotier, C., Gauthier, L.R., Gomez, D., Hoffschir, F., Mandine, E., Riou, J.F., Mergny, J.L., Mailliet, P., and Boussin, F.D. (2005). Apoptosis related to telomere instability and cell cycle

alterations in human glioma cells treated by new highly selective G-quadruplex ligands. Oncogene 24, 2917–2928.

Pijlman, G.P., Funk, A., Kondratieva, N., Leung, J., Torres, S., van der Aa, L., Liu, W.J., Palmenberg, A.C., Shi, P.Y., Hall, R.A., et al. (2008). A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. Cell Host Microbe 4, 579–591.

Pirakitikulr, N., Kohlway, A., Lindenbach, B.D., and Pyle, A.M. (2016). The coding region of the HCV genome contains a network of regulatory RNA structures. Mol Cell 62, 111–120.

Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K.J., Lukasiak, P., Bartol, N., Blazewicz, J., and Adamiak, R.W. (2012). Automated 3D structure composition for large RNAs. Nucleic Acids Res 40, e112.

Ramani, V., Qiu, R., and Shendure, J. (2015). High-throughput determination of RNA structure by proximity ligation. Nat Biotechnol 33, 980–984.

Ray, P.S., Jia, J., Yao, P., Majumder, M., Hatzoglou, M., and Fox, P.L. (2009). A stress-responsive RNA switch regulates VEGFA expression. Nature 457, 915–919.

Reis, R.S., Deforges, J., Schmidt, R.R., Schippers, J.H.M., and Poirier, Y. (2021). An antisense noncoding RNA enhances translation via localized structural rearrangements of its cognate mRNA. Plant Cell 33, 1381–1397.

Reyes, F.E., Garst, A.D., and Batey, R.T. (2009). Strategies in RNA crystallography. In: Methods in Enzymology. New York: Academic Press. 119–139.

Rice, G.M., Shivashankar, V., Ma, E.J., Baryza, J.L., and Nutiu, R. (2020). Functional atlas of primary miRNA maturation by the microprocessor. Mol Cell 80, 892–902.e4.

Rieder, E., Paul, A.V., Kim, D.W., van Boom, J.H., and Wimmer, E. (2000). Genetic and biochemical studies of poliovirus cis-acting replication element cre in relation to VPg uridylylation. J Virol 74, 10371–10380.

Ritchey, L.E., Su, Z., Tang, Y., Tack, D.C., Assmann, S.M., and Bevilacqua, P.C. (2017). Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure in vivo. Nucleic Acids Res 45, e135.

Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B. F.C., and Klug, A. (1974). Structure of yeast phenylalanine tRNA at 3 Å resolution. Nature 250, 546–551.

Romilly, C., Lippegaus, A., and Wagner, E.G.H. (2020). An RNA pseudoknot is essential for standby-mediated translation of the tisB toxin mRNA in Escherichia coli. Nucleic Acids Res 48, 12336–12347.

Rother, M., Rother, K., Puton, T., and Bujnicki, J.M. (2011). ModeRNA: a tool for comparative modeling of RNA 3D structure. Nucleic Acids Res 39, 4007–4022.

Rouleau, S.G., Garant, J.M., Bolduc, F., Bisaillon, M., and Perreault, J.P. (2018). G-Quadruplexes influence pri-microRNA processing. RNA Biol 15, 198–206.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature 505, 701–705.

Roy, R., Hohng, S., and Ha, T. (2008). A practical guide to single-molecule FRET. Nat Methods 5, 507–516.

Ruonala, R., Ko, D., and Helariutta, Y. (2017). Genetic networks in plant vascular development. Annu Rev Genet 51, 335–359.

Russel, D., Lasker, K., Webb, B., Velázquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., and Sali, A. (2012). Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol 10, e1001244.

Salvati, E., Leonetti, C., Rizzo, A., Scarsella, M., Mottolese, M., Galati, R., Sperduti, I., Stevens, M.F.G., D'Incalci, M., Blasco, M., et al. (2007). Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. J Clin Invest 117, 3236–3247.

Sasaki, J., Kusuhara, Y., Maeno, Y., Kobayashi, N., Yamashita, T., Sakae, K., Takeda, N., and Taniguchi, K. (2001). Construction of an infectious cDNA clone of aichi virus (a new member of the family

Picornaviridae) and mutational analysis of a stem-loop structure at the 5′ end of the genome. J Virol 75, 8021–8030.

Sato, K., Akiyama, M., and Sakakibara, Y. (2021). RNA secondary structure prediction using deep learning with thermodynamic integration. Nat Commun 12, 941.

Sauer, M., Juranek, S.A., Marks, J., De Magis, A., Kazemier, H.G., Hilbig, D., Benhalevy, D., Wang, X., Hafner, M., and Paeschke, K. (2019). DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. Nat Commun 10, 2421.

Schlüter, U., and Weber, A.P.M. (2020). Regulation and evolution of C_4 photosynthesis. Annu Rev Plant Biol 71, 183–215.

Schön, P. (2016). Imaging and force probing RNA by atomic force microscopy. Methods 103, 25–33.

Schön, P. (2018). Atomic force microscopy of RNA: state of the art and recent advancements. Semin Cell Dev Biol 73, 209–219.

Schwieters, C.D., Bermejo, G.A., and Clore, G.M. (2018). Xplor-NIH for molecular structure determination from NMR and other data sources. Protein Sci 27, 26–40.

Serganov, A., Huang, L., and Patel, D.J. (2009). Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. Nature 458, 233–237.

Serganov, A., and Nudler, E. (2013). A decade of riboswitches. Cell 152, 17–24.

Shan, M., Ji, X., Janssen, K., Silverman, I.M., Humenik, J., Garcia, B.A., Liebhaber, S.A., and Gregory, B.D. (2021). Dynamic changes in RNA-protein interactions and RNA secondary structure in mammalian erythropoiesis. Life Sci Alliance 4, e202000659.

Shao, X., Zhang, W., Umar, M.I., Wong, H.Y., Seng, Z., Xie, Y., Zhang, Y., Yang, L., Kwok, C.K., and Deng, X. (2020). RNA G-quadruplex structures mediate gene regulation in bacteria. mBio 11, 1–5.

Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B.J. (2016). Global mapping of human RNA-RNA interactions. Mol Cell 62, 618–626.

Sharma, S., Ding, F., and Dokholyan, N.V. (2008). iFoldRNA: three-dimensional RNA structure prediction and folding. Bioinformatics 24, 1951–1952.

Shen, E.Z., Chen, H., Ozturk, A.R., Tu, S., Shirayama, M., Tang, W., Ding, Y.H., Dai, S.Y., Weng, Z., and Mello, C.C. (2018). Identification of piRNA binding sites reveals the argonaute regulatory landscape of the C. elegans germline. Cell 172, 937–951.e18.

Sherlock, M.E., Sudarsan, N., and Breaker, R.R. (2018). Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. Proc Natl Acad Sci USA 115, 6052–6057.

Sherwood, A.V., and Henkin, T.M. (2016). Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. Annu Rev Microbiol 70, 361–374.

Shi, B., Zhang, J., Heng, J., Gong, J., Zhang, T., Li, P., Sun, B.F., Yang, Y., Zhang, N., Zhao, Y.L., et al. (2020). RNA structural dynamics regulate early embryogenesis through controlling transcriptome fate and function. Genome Biol 21, 120.

Shi, X., Bonilla, S., Herschlag, D., and Harbury, P. (2015). Quantifying nucleic acid ensembles with X-ray scattering interferometry. In: Methods in Enzymology. New York: Academic Press. 75–97.

Shi, X., Huang, L., Lilley, D.M.J., Harbury, P.B., and Herschlag, D. (2016). The solution structural ensembles of RNA kink-turn motifs and their protein complexes. Nat Chem Biol 12, 146–152.

Shi, X., Walker, P., Harbury, P.B., and Herschlag, D. (2017). Determination of the conformational ensemble of the TAR RNA by X-ray scattering interferometry. Nucleic Acids Res 45, gkw1352.

Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A.E., and Weeks, K.M. (2014). RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nat Methods 11, 959–965.

Silverman, I.M., Li, F., Alexander, A., Goff, L., Trapnell, C., Rinn, J.L., and Gregory, B.D. (2014). RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. Genome Biol 15, R3.

Simon, M.D. (2013). Capture hybridization analysis of RNA targets

(CHART). Curr Protoc Mol Biol 21, 101: 21.25.1–21.25.16.

Singh, J., Paliwal, K., Zhang, T., Singh, J., Litfin, T., and Zhou, Y. (2021). Improved RNA secondary structure and tertiary base-pairing prediction using evolutionary profile, mutational coupling and two-dimensional transfer learning. Bioinformatics 37, 2589–2600.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell 42, 794–805.

Somarowthu, S. (2016). Progress and current challenges in modeling large RNAs. J Mol Biol 428, 736–747.

Somarowthu, S., Legiewicz, M., Chillón, I., Marcia, M., Liu, F., and Pyle, A.M. (2015). HOTAIR forms an intricate and modular secondary structure. Mol Cell 58, 353–361.

Song, Y., Yang, W., Fu, Q., Wu, L., Zhao, X., Zhang, Y., and Zhang, R. (2020). irCLASH reveals RNA substrates recognized by human ADARs. Nat Struct Mol Biol 27, 351–362.

Spitale, R.C., Crisalli, P., Flynn, R.A., Torre, E.A., Kool, E.T., and Chang, H.Y. (2013). RNA SHAPE analysis in living cells. Nat Chem Biol 9, 18–20.

Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., et al. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature 519, 486–490.

Spokoini-Stern, R., Stamov, D., Jessel, H., Aharoni, L., Haschke, H., Giron, J., Unger, R., Segal, E., Abu-Horowitz, A., and Bachelet, I. (2020). Visualizing the structure and motion of the long noncoding RNA HOTAIR. RNA 26, 629–636.

Statello, L., Guo, C.J., Chen, L.L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 22, 96–118.

Stephenson, J.D., Kenyon, J.C., Symmons, M.F., and Lever, A.M.L. (2016). Characterizing 3D RNA structure by single molecule FRET. Methods 103, 57–67.

Stephenson, J.D., Li, H., Kenyon, J.C., Symmons, M., Klenerman, D., and Lever, A.M.L. (2013). Three-dimensional RNA structure of the major HIV-1 packaging signal region. Structure 21, 951–962.

Su, Z., Tang, Y., Ritchey, L.E., Tack, D.C., Zhu, M., Bevilacqua, P.C., and Assmann, S.M. (2018). Genome-wide RNA structurome reprogramming by acute heat shock globally regulates mRNA abundance. Proc Natl Acad Sci USA 115, 12170–12175.

Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J.L., and Moine, H. (2011). G-quadruplex RNA structure as a signal for neurite mRNA targeting. EMBO Rep 12, 697–704.

Sugimoto, Y., Chakrabarti, A.M., Luscombe, N.M., and Ule, J. (2017). Using hiCLIP to identify RNA duplexes that interact with a specific RNA-binding protein. Nat Protoc 12, 611–637.

Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D'Ambrogio, A., Luscombe, N.M., and Ule, J. (2015). hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. Nature 519, 491–494.

Sun, L., Fazal, F.M., Li, P., Broughton, J.P., Lee, B., Tang, L., Huang, W., Kool, E.T., Chang, H.Y., and Zhang, Q.C. (2019). RNA structure maps across mammalian cellular compartments. Nat Struct Mol Biol 26, 322–330.

Sun, L., Li, P., Ju, X., Rao, J., Huang, W., Ren, L., Zhang, S., Xiong, T., Xu, K., Zhou, X., et al. (2021a). In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs. Cell 184, 1865–1883.e20.

Sun, L., Xu, K., Huang, W., Yang, Y.T., Li, P., Tang, L., Xiong, T., and Zhang, Q.C. (2021b). Predicting dynamic cellular protein-RNA interactions by deep learning using in vivo RNA structures. Cell Res 31, 495–516.

Tack, D.C., Su, Z., Yu, Y., Bevilacqua, P.C., and Assmann, S.M. (2020). Tissue-specific changes in the RNA structurome mediate salinity response in Arabidopsis. RNA 26, 492–511.

Thapa, R.J., Ingram, J.P., Ragan, K.B., Nogusa, S., Boyd, D.F., Benitez, A. A., Sridharan, H., Kosoff, R., Shubina, M., Landsteiner, V.J., et al.

(2016). DAI senses influenza A virus genomic RNA and activates RIPK3-dependent cell death. Cell Host Microbe 20, 674–681.

Tomezsko, P.J., Corbin, V.D.A., Gupta, P., Swaminathan, H., Glasgow, M., Persad, S., Edwards, M.D., Mcintosh, L., Papenfuss, A.T., Emery, A., et al. (2020). Determination of RNA structural diversity and its role in HIV-1 RNA splicing. Nature 582, 438–442.

Townshend, R.J.L., Eismann, S., Watkins, A.M., Rangan, R., Karelina, M., Das, R., and Dror, R.O. (2021). Geometric deep learning of RNA structure. Science 373, 1047–1051.

Trinh, M.H., Odorico, M., Pique, M.E., Teulon, J.M., Roberts, V.A., Ten Eyck, L.F., Getzoff, E.D., Parot, P., Chen, S.W.W., and Pellequer, J.L. (2012). Computational reconstruction of multidomain proteins using atomic force microscopy data. Structure 20, 113–120.

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. Nature 596, 590–596.

Tuplin, A., Wood, J., Evans, D.J., Patel, A.H., and Simmonds, P. (2002). Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. RNA 8, 824–841.

Umar, M.I., and Kwok, C.K. (2020). Specific suppression of D-RNA G-quadruplex-protein interaction with an L-RNA aptamer. Nucleic Acids Res 48, 10125–10141.

Underwood, J.G., Uzilov, A.V., Katzman, S., Onodera, C.S., Mainzer, J.E., Mathews, D.H., Lowe, T.M., Salama, S.R., and Haussler, D. (2010). FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. Nat Methods 7, 995–1001.

Wachter, A., Tunc-Ozdemir, M., Grove, B.C., Green, P.J., Shintani, D.K., and Breaker, R.R. (2007). Riboswitch control of gene expression in plants by splicing and alternative 3′ end processing of mRNAs. Plant Cell 19, 3437–3450.

Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E., et al. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. Nature 505, 706–709.

Wang, J., Zuo, X., Yu, P., Xu, H., Starich, M.R., Tiede, D.M., Shapiro, B. A., Schwieters, C.D., and Wang, Y.X. (2009). A method for helical RNA global structure determination in solution using small-angle X-ray scattering and NMR measurements. J Mol Biol 393, 717–734.

Wang, S., Alseekh, S., Fernie, A.R., and Luo, J. (2019a). The structure and function of major plant metabolite modifications. Mol Plant 12, 899–919.

Wang, X., Alnabati, E., Aderinwale, T.W., Maddhuri Venkata Subramaniya, S.R., Terashi, G., and Kihara, D. (2021a). Detecting protein and DNA/RNA structures in cryo-EM maps of intermediate resolution using deep learning. Nat Commun 12, 2302.

Wang, X.W., Liu, C.X., Chen, L.L., and Zhang, Q.C. (2021b). RNA structure probing uncovers RNA structure-dependent biological functions. Nat Chem Biol 17, 755–766.

Wang, Y., Chen, Y., Hu, Y., and Fang, X. (2020a). Site-specific covalent labeling of large RNAs with nanoparticles empowered by expanded genetic alphabet transcription. Proc Natl Acad Sci USA 117, 22823–22832.

Wang, Y., Kathiresan, V., Chen, Y., Hu, Y., Jiang, W., Bai, G., Liu, G., Qin, P.Z., and Fang, X. (2020b). Posttranscriptional site-directed spin labeling of large RNAs with an unnatural base pair system under non-denaturing conditions. Chem Sci 11, 9655–9664.

Wang, Z., Ma, Z., Castillo-González, C., Sun, D., Li, Y., Yu, B., Zhao, B., Li, P., and Zhang, X. (2018). SWI2/SNF2 ATPase CHR2 remodels pri-miRNAs via Serrate to impede miRNA production. Nature 557, 516–521.

Wang, Z., Wang, M., Wang, T., Zhang, Y., and Zhang, X. (2019b). Genome-wide probing RNA structure with the modified DMS-MaPseq in Arabidopsis. Methods 155, 30–40.

Ward, A.B., Sali, A., and Wilson, I.A. (2013). Integrative structural biology. Science 339, 913–915.

Warner, K.D., Chen, M.C., Song, W., Strack, R.L., Thorn, A., Jaffrey, S.R.,

and Ferré-D'Amaré, A.R. (2014). Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. Nat Struct Mol Biol 21, 658–663.

Watkins, A.M., Rangan, R., and Das, R. (2019). Using Rosetta for RNA homology modeling. In: Methods in Enzymology. New York: Academic Press. 177–207.

Watkins, A.M., Rangan, R., and Das, R. (2020). FARFAR2: improved *de novo* Rosetta prediction of complex global RNA folds. Structure 28, 963–976.e6.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess Julian W., J., Swanstrom, R., Burch, C.L., and Weeks, K.M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460, 711–716.

Weinrich, T., Jaumann, E.A., Scheffer, U., Prisner, T.F., and Göbel, M.W. (2018). A cytidine phosphoramidite with protected nitroxide spin label: synthesis of a full-length TAR RNA and investigation by in-line probing and EPR spectroscopy. Chem Eur J 24, 6202–6207.

Weng, X., Gong, J., Chen, Y., Wu, T., Wang, F., Yang, S., Yuan, Y., Luo, G., Chen, K., Hu, L., et al. (2020). Keth-seq for transcriptome-wide RNA structure mapping. Nat Chem Biol 16, 489–492.

Westerhout, E.M., Ooms, M., Vink, M., Das, A.T., and Berkhout, B. (2005). HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome. Nucleic Acids Res 33, 796–804.

Wilkinson, M.E., Charenton, C., and Nagai, K. (2020). RNA splicing by the spliceosome. Annu Rev Biochem 89, 359–388.

Winkler, W., Nahvi, A., and Breaker, R.R. (2002a). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419, 952–956.

Winkler, W.C., Cohen-Chalamish, S., and Breaker, R.R. (2002b). An mRNA structure that controls gene expression by binding FMN. Proc Natl Acad Sci USA 99, 15908–15913.

Wu, B., Grigull, J., Ore, M.O., Morin, S., and White, K.A. (2013). Global organization of a positive-strand RNA virus genome. PLoS Pathog 9, e1003363.

Wu, X., and Bartel, D.P. (2017). Widespread influence of 3′-end structures on mammalian mRNA processing and stability. Cell 169, 905–917.e11.

Xiao, C.D., Shibata, T., Yamamoto, Y., and Xu, Y. (2018). An intramolecular antiparallel G-quadruplex formed by human telomere RNA. Chem Commun 54, 3944–3946.

Xu, X., Zhao, P., and Chen, S.J. (2014). Vfold: a web server for RNA structure and folding thermodynamics prediction. PLoS ONE 9, e107504.

Xue, Y., Chen, R., Qu, L., and Cao, X. (2020). Noncoding RNA: from dark matter to bright star. Sci China Life Sci 63, 463–468.

Xue, Y., Ouyang, K., Huang, J., Zhou, Y., Ouyang, H., Li, H., Wang, G., Wu, Q., Wei, C., Bi, Y., et al. (2013). Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. Cell 152, 82–96.

Xue, Z., Hennelly, S., Doyle, B., Gulati, A.A., Novikova, I.V., Sanbonmatsu, K.Y., and Boyer, L.A. (2016). A G-rich motif in the lncRNA braveheart interacts with a zinc-finger transcription factor to specify the cardiovascular lineage. Mol Cell 64, 37–50.

Yang, M., Woolfenden, H.C., Zhang, Y., Fang, X., Liu, Q., Vigh, M.L., Cheema, J., Yang, X., Norris, M., Yu, S., et al. (2020a). Intact RNA structurome reveals mRNA structure-mediated regulation of miRNA cleavage *in vivo*. Nucleic Acids Res 48, 8767–8781.

Yang, S.Y., Lejault, P., Chevrier, S., Boidot, R., Robertson, A.G., Wong, J. M.Y., and Monchaud, D. (2018). Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. Nat Commun 9, 4730.

Yang, X., Cheema, J., Zhang, Y., Deng, H., Duncan, S., Umar, M.I., Zhao, J., Liu, Q., Cao, X., Kwok, C.K., et al. (2020b). RNA G-quadruplex structures exist and function *in vivo* in plants. Genome Biol 21, 226.

Yang, X., Yu, H., Sun, W., Ding, L., Li, J., Cheema, J., Ramirez-Gonzalez, R., Zhao, X., Martín, A.C., Lu, F., et al. (2021). Wheat *in vivo* RNA structure landscape reveals a prevalent role of RNA structure in modulating translational subgenome expression asymmetry. Genome

Biol 22, 326.

Yang, Y., Liu, S., Egloff, S., Eichhorn, C.D., Hadjian, T., Zhen, J., Kiss, T., Zhou, Z.H., and Feigon, J. (2022). Structural basis of RNA conformational switching in the transcriptional regulator 7SK RNP. Mol Cell doi: 10.1016/j.molcel.2022.03.001.

Yang, Y., Zhan, L., Zhang, W., Sun, F., Wang, W., Tian, N., Bi, J., Wang, H., Shi, D., Jiang, Y., et al. (2011). RNA secondary structure in mutually exclusive splicing. Nat Struct Mol Biol 18, 159–168.

Yeung, P.Y., Zhao, J., Chow, E.Y.C., Mou, X., Hong, H.Q., Chen, L., Chan, T.F., and Kwok, C.K. (2019). Systematic evaluation and optimization of the experimental steps in RNA G-quadruplex structure sequencing. Sci Rep 9, 8091.

Yu, Y., Jia, T., and Chen, X. (2017). The 'how' and 'where' of plant microRNAs. New Phytol 216, 1002–1017.

Yue, Y., Yang, Y., Dai, L., Cao, G., Chen, R., Hong, W., Liu, B., Shi, Y., Meng, Y., Shi, F., et al. (2016). Long-range RNA pairings contribute to mutually exclusive splicing. RNA 22, 96–110.

Zalfa, F., Eleuteri, B., Dickson, K.S., Mercaldo, V., De Rubeis, S., di Penta, A., Tabolacci, E., Chiurazzi, P., Neri, G., Grant, S.G.N., et al. (2007). A new function for the fragile X mental retardation protein in regulation of PSD-95 mRNA stability. Nat Neurosci 10, 578–587.

Zettl, T., Das, R., Harbury, P.A.B., Herschlag, D., Lipfert, J., Mathew, R.S., and Shi, X. (2018). Recording and analyzing nucleic acid distance distributions with X-Ray scattering interferometry (XSI). Curr Protoc Nucleic Acid Chem 73, e54.

Zettl, T., Mathew, R.S., Seifert, S., Doniach, S., Harbury, P.A.B., and Lipfert, J. (2016). Absolute intramolecular distance measurements with angstrom-resolution using anomalous small-angle X-ray scattering. Nano Lett 16, 5353–5357.

Zhang, B., Zhang, X., Pearce, R., Shen, H.B., and Zhang, Y. (2020a). A new protocol for atomic-level protein structure modeling and refinement using low-to-medium resolution Cryo-EM density maps. J Mol Biol 432, 5365–5377.

Zhang, J., and Ferré-D'Amaré, A.R. (2016). The tRNA elbow in structure, recognition and evolution. Life 6, 3.

Zhang, K., Li, S., Kappel, K., Pintilie, G., Su, Z., Mou, T.C., Schmid, M.F., Das, R., and Chiu, W. (2019a). Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. Nat Commun 10, 5511.

Zhang, T., Yin, C., Boyd, D.F., Quarato, G., Ingram, J.P., Shubina, M., Ragan, K.B., Ishizuka, T., Crawford, J.C., Tummers, B., et al. (2020b). Influenza virus Z-RNAs induce ZBP1-mediated necroptosis. Cell 180, 1115–1129.e13.

Zhang, W., Thieme, C.J., Kollwig, G., Apelt, F., Yang, L., Winter, N., Andresen, N., Walther, D., and Kragler, F. (2016). tRNA-related sequences trigger systemic mRNA transport in plants. Plant Cell 28, 1237–1249.

Zhang, X., Spiegel, J., Martínez Cuesta, S., Adhikari, S., and Balasubramanian, S. (2021). Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. Nat Chem 13, 626–633.

Zhang, X., Yu, L., Ye, S., Xie, J., Huang, X., Zheng, K., and Sun, B. (2019b). MOV10L1 binds RNA G-quadruplex in a structure-specific manner and resolves it more efficiently than MOV10. iScience 17, 36–48.

Zhang, X.O., Wang, H.B., Zhang, Y., Lu, X., Chen, L.L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. Cell 159, 134–147.

Zhang, Y., Burkhardt, D.H., Rouskin, S., Li, G.W., Weissman, J.S., and Gross, C.A. (2018). A stress response that monitors and regulates mRNA structure is central to cold shock adaptation. Mol Cell 70, 274–286.e7.

Zhang, Y., Wang, J., and Xiao, Y. (2020c). 3dRNA: building RNA 3D structure with improved template library. Comput Struct Biotechnol J 18, 2416–2423.

Zhang, Y., Yang, M., Duncan, S., Yang, X., Abdelhamid, M.A.S., Huang, L., Zhang, H., Benfey, P.N., Waller, Z.A.E., and Ding, Y. (2019c). G-quadruplex structures trigger RNA phase separation. Nucleic Acids Res 47, 11746–11754.

Zhang, Y., Zhang, Y., Liu, Z.Y., Cheng, M.L., Ma, J., Wang, Y., Qin, C.F., and Fang, X. (2019d). Long non-coding subgenomic flavivirus RNAs have extended 3D structures and are flexible in solution. EMBO Rep 20.

Zhao, Y., Huang, Y., Gong, Z., Wang, Y., Man, J., and Xiao, Y. (2012). Automated and fast building of three-dimensional RNA structures. Sci Rep 2, 734.

Zheng, K.W., Zhang, J.Y., He, Y., Gong, J.Y., Wen, C.J., Chen, J.N., Hao, Y.H., Zhao, Y., and Tan, Z. (2020). Detection of genomic G-quadruplexes in living cells using a small artificial protein. Nucleic Acids Res 48, 11706–11720.

Zheng, Q., Ryvkin, P., Li, F., Dragomir, I., Valladares, O., Yang, J., Cao, K., Wang, L.S., and Gregory, B.D. (2010). Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. PLoS Genet 6, e1001141.

Zhou, X., Sunkar, R., Jin, H., Zhu, J.K., and Zhang, W. (2009). Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in *Oryza sativa*. Genome Res 19, 70–78.

Zhu, J., Li, C., Peng, X., and Zhang, X. (2021). RNA architecture influences plant biology. J Exp Bot 72, 4144–4160.

Ziv, O., Gabryelska, M.M., Lun, A.T.L., Gebert, L.F.R., Sheu-Gruttadauria, J., Meredith, L.W., Liu, Z.Y., Kwok, C.K., Qin, C.F., MacRae, I.J., et al. (2018). COMRADES determines *in vivo* RNA structures and interactions. Nat Methods 15, 785–788.

Ziv, O., Price, J., Shalamova, L., Kamenova, T., Goodfellow, I., Weber, F., and Miska, E.A. (2020). The short- and long-range RNA-RNA interactome of SARS-CoV-2. Mol Cell 80, 1067–1077.e5.

Zubradt, M., Gupta, P., Persad, S., Lambowitz, A.M., Weissman, J.S., and Rouskin, S. (2017). DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo*. Nat Methods 14, 75–82.