Genome-Wide Analysis of RNA Secondary Structure

Philip C. Bevilacqua,[1,2,3] Laura E. Ritchey,[1,3] Zhao Su,[4] and Sarah M. Assmann[4]

[1]Department of Chemistry, [2]Department of Biochemistry and Molecular Biology, [3]Center for RNA Molecular Biology, [4]Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802; email: sma3@psu.edu

Annu. Rev. Genet. 2016. 50:235–66

First published online as a Review in Advance on September 14, 2016

The *Annual Review of Genetics* is online at genet.annualreviews.org

This article's doi: 10.1146/annurev-genet-120215-035034

Keywords

DMS-seq, genome-wide, in vivo RNA folding, RNA structurome, SHAPE, Structure-seq

Abstract

Single-stranded RNA molecules fold into extraordinarily complicated secondary and tertiary structures as a result of intramolecular base pairing. In vivo, these RNA structures are not static. Instead, they are remodeled in response to changes in the prevailing physicochemical environment of the cell and as a result of intermolecular base pairing and interactions with RNA-binding proteins. Remarkable technical advances now allow us to probe RNA secondary structure at single-nucleotide resolution and genome-wide, both in vitro and in vivo. These data sets provide new glimpses into the RNA universe. Analyses of RNA structuromes in HIV, yeast, *Arabidopsis*, and mammalian cells and tissues have revealed regulatory effects of RNA structure on messenger RNA (mRNA) polyadenylation, splicing, translation, and turnover. Application of new methods for genome-wide identification of mRNA modifications, particularly methylation and pseudouridylation, has shown that the RNA "epitranscriptome" both influences and is influenced by RNA structure. In this review, we describe newly developed genome-wide RNA structure-probing methods and synthesize the information emerging from their application.

INTRODUCTION

RNA serves a multitude of functional roles in the cell as both a coding and a noncoding molecule. Unlike the elegant yet simple double helix of DNA, RNA can take on a vast array of complicated secondary and tertiary structures that in part account for its accentuated functionality. The primary structure of RNA folds into a complex secondary structure with various helical defects such as loops and bulges (**Figure 1**). Protein binding to RNA further contributes to a complex three-dimensional shape. One well-known, largely prokaryotic example of the importance of RNA structure to RNA function is the riboswitch, in which ligand binding alters untranslated region (UTR) structure, thereby either sequestering or releasing motifs that regulate transcription, stability, or translation. In eukaryotes, perhaps the most familiar examples of the importance of RNA structure come from microRNAs and other small RNAs in which intramolecular or intermolecular base pairing plays essential roles in small-RNA processing, leading to recognition and subsequent regulation of target messenger RNAs (mRNAs).

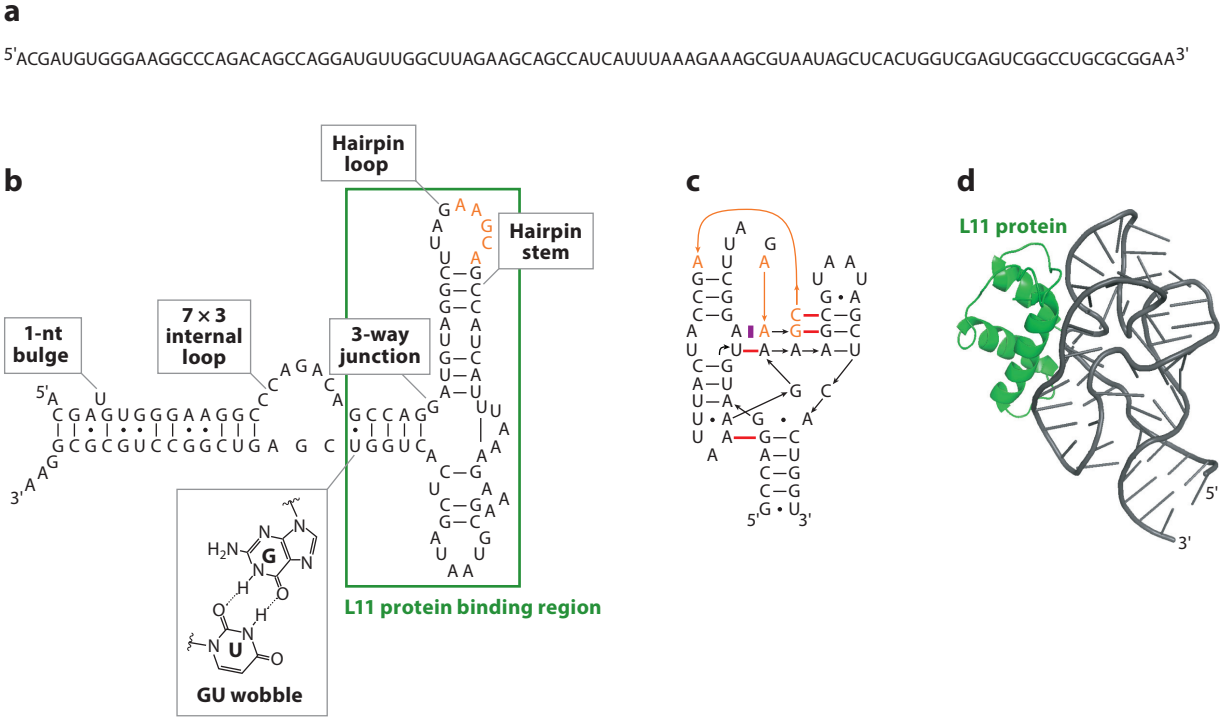


Figure 1

Structural hierarchy and complexity of RNA illustrated by the GTPase center RNA from *Escherichia coli*. (*a*) Primary structure of GTPase center RNA, a 99-nucleotide (nt) region of large subunit ribosomal RNA. (*b*) Secondary structure of GTPase center RNA, including a U1061A mutation (also present in panels *c* and *d*) in the hairpin stem that stabilizes tertiary interactions. Representative secondary structural elements are highlighted to demonstrate the diversity of RNA structure, which includes hairpin stems, hairpin loops, bulges, internal loops, 3-way junctions, and GU wobbles. The L11 protein binding region is in the green box. Redrawn with permission from Reference 8. (*c*) Supersecondary structure of the 58-nt L11 binding region highlighting tertiary interactions (*red*) and base stacking (*purple*). Orange coloring in panels *b* and *c* provides an aid in following the complex turns of the supersecondary structure. Redrawn with permission from Reference 16. (*d*) Crystal structure of the L11 binding region bound to L11 protein (*green*). Protein Data Bank ID: 1QA6 (16).

However, it is now becoming evident that RNA structure plays important regulatory roles not just for specialized classes of RNAs such as riboswitches, small RNAs, and enzymatic RNAs (ribozymes) but for essentially all RNAs, including mRNAs and long noncoding RNAs (lncRNAs). Base-pairing changes play a simple but powerful role in gene regulation in part by sequestering or exposing nucleotides that can interact with other RNAs, proteins, or other cellular constituents. RNA structure thus can be considered as another, previously hidden, layer of the genetic code that we are only just beginning to understand.

Developed over the past half-decade, experimental methods that couple RNA structure probing with high-throughput sequencing techniques have revealed RNA secondary structures genome-wide. Application of these methods has also revealed general principles of how structure informs function for all classes of RNAs and particularly for mRNAs. Whereas the first generation of such methods was restricted to in vitro applications (118), new methods developed starting in 2013 enabled genome-wide structure probing in vivo (22). This review summarizes genome-wide methods of RNA structure analysis and current findings based on them.

PARAMETERS THAT AFFECT RNA STRUCTURE

At the most basic level, RNA secondary structure consists of base-paired and unpaired nucleotides from which arise the stem and loop structure of hairpins, bulges, and pseudoknots in which *trans*-loop base pairing occurs (**Figure 1**). Base pairs are either canonical (for example, A-U and G-C) or noncanonical (among which the G•U wobble is the most common example). Each base pair has an associated contextual free energy. From this fundamental knowledge, along with extensive experimental measurements on melting of short oligonucleotide structures, a set of nearest-neighbor parameters, so-called Turner rules (76, 128), was developed. These rules can be applied to a given RNA sequence to predict its secondary structure via free energy minimization, and they form the basis of in silico methods of RNA structure prediction, as provided in several widely used programs, e.g., RNAstructure (87), the ViennaRNA package (66), and Mfold (73). Elevated temperatures tend to unfold RNA; thus, the RNAstructure program (69) and ViennaRNA (66) have incorporated temperature as a component of structure prediction.

However, in silico predictions fall short of predicting the actual structures of complex RNA polymers (18). Although the most prevalent in silico methods employ free energy minimization, this approach is not fully parameterized by experimental data. Indeed, such parameterization is an essentially impossible task given the number of variables that affect RNA structure. For instance, cations strongly influence RNA structure, especially Mg^{2+}, which tends to promote secondary and tertiary folding, and heavy metals, which tend to unfold RNA. In addition, pH, compatible solutes, and crowding all have various and complex effects on RNA structure. Even though these variables can be explicitly defined in vitro, their concentrations in living systems often remain poorly defined and/or highly modulated, such that exact reproduction in the test tube of in vivo conditions remains impossible.

The living cell adds still more layers of complexity. Interactions with other molecules such as binding of RNA by proteins and small molecules (e.g., those operating in the aforementioned riboswitches) can have profound effects on RNA structure. Enzymatic reactions can affect RNA structure via covalent modification of the RNA, which can alter RNA stability (38, 90), or via unwinding in the case of RNA helicases. Moreover, all living systems expend energy to exist in a non-lowest free energy state, yet free energy minimization approaches assume that all RNAs in a system have reached equilibrium. In fact, there is extensive evidence that some RNAs fold to biologically relevant structures that are not the minimum free energy configuration (15, 127). Finally, absent very specific experimental conditions in the test tube, such as high Mg^{2+} concentrations,

even highly structured RNAs tend to exist across a population of states rather than populating a single explicit structure. Modeling such biologically relevant population heterogeneity is a current major challenge even for purely in silico structure prediction.

For all the above reasons, improved prediction of RNA structure in living systems has required the development of wet bench methods to query RNA structure, and the need to derive general structure-function principles has required that such methods be applicable genome-wide. Further, because it is impossible to faithfully recapitulate in vitro all the physicochemical and enzymatic aspects of living systems, the most biologically informative structure-probing methods will be those that can be applied in vivo. As stated above, most RNAs cannot be well predicted using only thermodynamic parameters (so-called in silico methods); rather, experimental data are incorporated to restrain in silico prediction algorithms and yield-predicted structures (75, 101). **Figure 2** illustrates the dramatic differences that can arise when an RNA structure is predicted purely in silico versus when it is restrained by experimental information on folding derived from Structure-seq (21, 22), one of the new in vivo structure-probing methods discussed in this article.

METHODS OF GENOME-WIDE EXPERIMENTAL RNA STRUCTURE PROBING

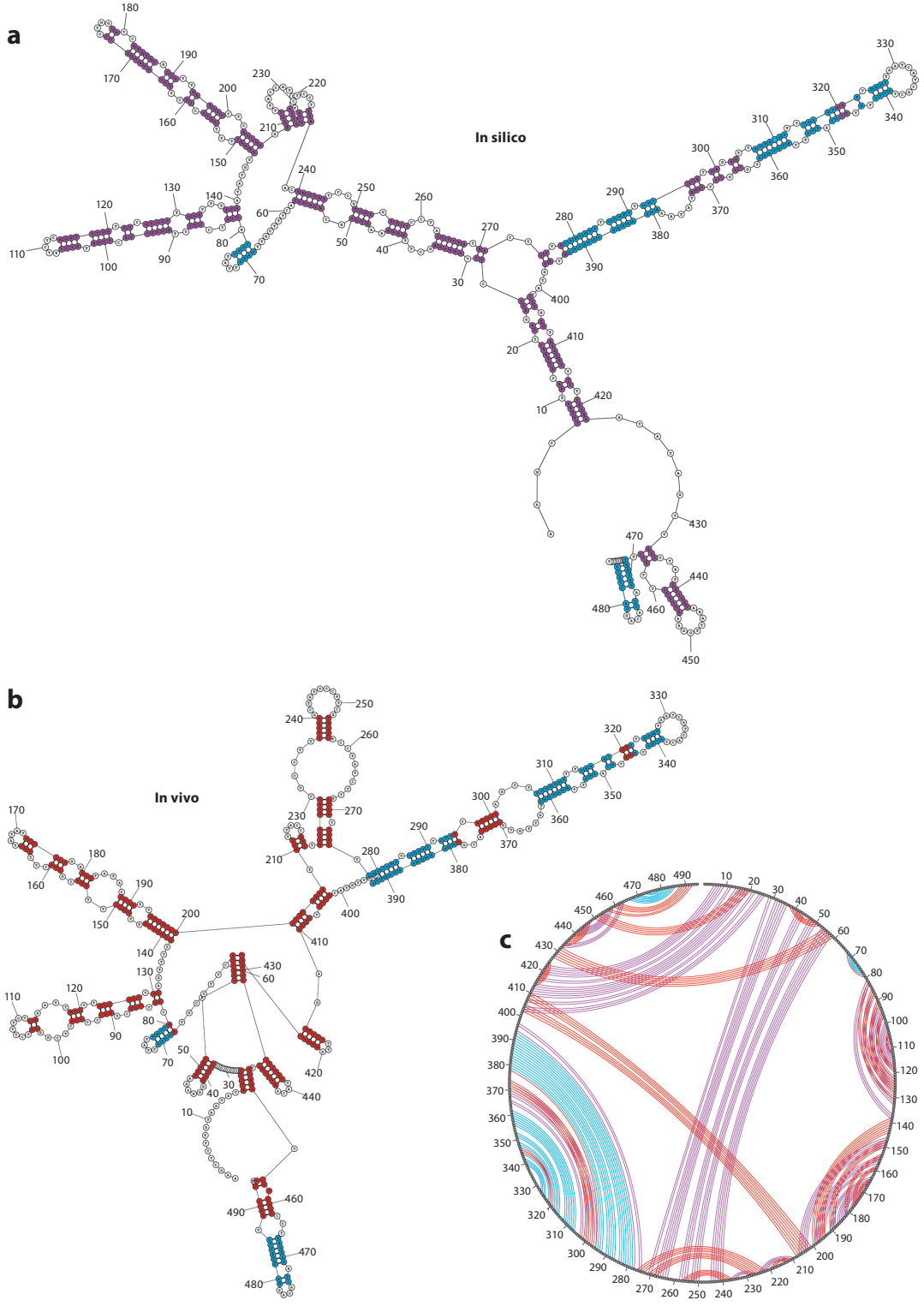
Technique development has played critical roles in many landmark discoveries in RNA biology and chemistry. For instance, sequencing of RNA and DNA was made possible by advances in RNA chemistry such as methods for chemical modification of the bases and dideoxy sequencing, and revolutions in cloning and genetics resulted from major developments in molecular biology such as the polymerase chain reaction. Methods for genome-wide probing of RNA are also advancing rapidly. In just the past 2–3 years, techniques have been developed to probe RNA structure in vivo across the genome. This has been achieved by combining methods to modify RNA using chemicals and enzymes developed within the past 20–40 years with methods using deep sequencing developed within the past 10 years. Continued improvement of these methods is critical, as it will open new doors into genomics and RNA biology and their applications.

Short Historical Perspective on Non-Genome-Wide Methods to Probe RNA Structure

It is instructive to consider briefly the development of non-genome-wide methods to probe RNA structure. When the structure of double-stranded DNA was determined in 1953, RNA also seemed likely to base pair, even though it is single stranded. In 1956, Rich & Davies (88) solved a fiber X-ray diffraction structure for double-stranded RNA comprising the abiological polymers poly(rA) and poly(rU) that was a right-handed double helix with base pairing and stacking akin to that of double-stranded DNA. Additional insights into RNA structure were obtained in the 1960s when researchers found that single-stranded RNA could stack through local van der Waals' interactions.

Figure 2

RNA folds differently in silico and in vivo. Messenger RNA *RCI2A* (*Rare-Cold-Inducible 2A*; AT3G05880) from *Arabidopsis* folded using RNAstructure (110) (*a*) without and (*b*) with in vivo restraints from Structure-seq (22). Redrawn with permission from Reference 22. (*c*) Circle-compare plot of structures shown in panels *a* and *b*. The plot shows the nucleotides that are base paired in both structures (*cyan*), those that are base paired only in the in silico–predicted structure (*purple*), and those that are base paired only in the in vivo–predicted structure (*red*). The same color coding is used in the structures in panels *a* and *b*.


a
In silico
b
In vivo
c

Stacking of RNA was first described for simple systems, homoribopolymers and dinucleotides, using spectroscopic methods of optical rotary dispersion (13) and nuclear magnetic resonance (4).

The structure of transfer RNA (tRNA) was the first to be determined for a biological RNA. The impact of this structure on RNA biology and chemistry cannot be overstated. The structure was determined through a variety of methods that built quickly upon each other. In 1965, the sequence of $tRNA^{Ala}$ was determined and manually folded to form A/U and G/C base pairs (36), and three possible structures were deduced. Soon thereafter, other tRNA sequences were determined and folded, and upon comparison among them only one common fold, the now familiar cloverleaf, was found. This was one of the first comparative analyses of RNA structure, and such comparative, sometimes termed "phylogenetic," methods remain powerful approaches for defining the base-paired structure of an RNA in vivo.

To assess experimentally the secondary structure of the cloverleaf model, a method of oligonucleotide hybridization was developed in 1972 in which tRNA was challenged with short tetramers. If the tetramers hybridized, then the target region of tRNA was deemed single stranded (113). A year later, researchers solved a crystal X-ray diffraction structure of tRNA (47) and found it to be consistent with this secondary structure. This crystal structure provided a paradigm shift in the understanding of RNA structure and function. For the first time, secondary and tertiary elements such as loops, bulges, and junctions were visible. This structure demonstrated that RNA can have a complex three-dimensional fold, not unlike a protein, resulting in myriad implications for function of tRNA specifically and RNA generally. For instance, appreciating how tRNA could link to amino acids as an adaptor for translation presaged envisioning how a molecule of the complexity of RNA could recognize small molecules in riboswitches and catalyze reactions in ribozymes. The molecular complexity of RNA, revealed in the seminal structure of tRNA, anticipated the biological complexity of RNA we now know.

The methods used in these early studies—fiber and crystal X-ray diffraction as well as optical rotary dispersion and nuclear magnetic resonance spectroscopy—are not readily applicable to studying RNA in a high-throughput manner or in vivo. Phylogenetic analysis of RNA, although powerful and based on biological comparisons, requires extensive well-annotated sequence information across genomes from distant life forms. Thus, new experimental approaches had to be developed to probe the structure of RNA directly.

Secondary Structure Probing

Two general classes of reagents are used to probe RNA structure: enzymes and chemicals (**Table 1**). Both provide specific information on various aspects of RNA structure (26). RNases used to probe RNA sequence cleave the RNA sequence at defined structural locations. For instance, RNase V1 cleaves RNA in base-paired regions, although the structural specificity is not absolute (68). By contrast, RNases S1 and P1 generally cleave RNA in unstructured regions, although these enzymes may miss small bulges or loops (114). By combining the results of various enzymatic treatments, the base-paired and unstructured regions of a given RNA can be largely determined. However, this method is not amenable to in vivo probing of RNA structure: These large enzymes are not membrane permeant, and some of the enzymes, e.g., RNase V1, require non-physiological Mg^{2+} concentrations for activity (49). Because Mg^{2+} also promotes RNA folding, this can be problematic if attempting to equate in vitro with in vivo structures.

In general, nucleic acids are chemically unreactive, which provides robustness to the genetic material. Nonetheless, chemists have devised reagents that modify the functional groups on the Watson-Crick (WC) face of the base, the top of the base (Hoogsteen face), or the sugar-phosphate backbone. On the WC face, adenosine and cytidine react with alkylating reagents such

Table 1 Summary of genome-wide methods of structure probing, organized by readout method

Method	Reagent	Region probed[c]	Readout[d]	In vivo, in vitro, or both[e]	Physical conditions					Reference(s)
					pH	Monovalent ions	Mg^{2+}	Temperature	Other	
FragSeq	RNase P1[a]	Single-stranded regions	Cleavage	In vitro only	7.5	150 mM NaCl	5 mM	37°C	0.01 mM $Zn(OAc)_2$	114
PARS, PARTE	RNase V1 + RNase S1[a]	Double-stranded and single-stranded regions	Cleavage	In vitro only	7.0–7.4	100–150 mM NaCl	10 mM	37°C (varied for PARTE)	NA	43, 84, 119, 120
PIP-seq	RNase V1 + RNase ONE[a]	Double-stranded and single-stranded regions and protein-binding sites	Cleavage	In vitro only	V1: 7.0 ONE: 7.5	V1: 100 mM KCl ONE: 200 mM $NaCH_3COO$	V1: 10 mM ONE: none	V1: RT, 37°C ONE: 37°C, RT	V1: NA ONE: 5 mM EDTA 200 μg/ml BSA	99, 100, 116
CIRS-seq	DMS + CMCT[b]	WC face of A, C, G, and U	RT stop	In vitro[f]	DMS: 7.9 CMCT: 8.0	100 mM KCl	10 mM	30°C	0.5% NP-40	40
DMS-seq	DMS[b]	WC face of A and C	RT stop	In vitro	8.0	100 mM NaCl	6 mM	Denatured control: 95°C In vitro: 30°C	NA	91
				In vivo	In vivo					91
icSHAPE	SHAPE (NAI-N_3)[b]	2′-hydroxyl	RT stop	In vitro	7.5–8.0	100 mM NaCl	6 mM	37°C	NA	28, 106
				In vivo	In vivo					106

(*Continued*)

Table 1 (*Continued*)

Method	Reagent	Region probed[c]	Readout[d]	In vivo, in vitro, or both[e]	Physical conditions					Reference(s)
					pH	Monovalent ions	Mg^{2+}	Temperature	Other	
Mod-seq	DMS[b]	WC face of A and C	RT stop	In vivo[g]	In vivo					61, 108
Structure-seq	DMS[b]	WC face of A and C	RT stop	In vivo[g]	In vivo					21, 22, 110
SHAPE-MaP	SHAPE (1M6, 1M7, NMIA)[b]	2′-hydroxyl	RT mutation	In vitro[f]	8.0	200 mM KOAc	3 mM	37°C	NA	98, 103

[a]Enzymatic method.
[b]Chemical method. DMS modifies the WC face of A and C; CMCT modifies the WC face of G and U; SHAPE modifies the 2′OH of the ribose sugar.
[c]Specificities for the enzymatic probes are not absolute, as described in References 64, 68, and 114.
[d]In cleavage, the probing method cleaves the RNA strands, which allows for library preparation. In reverse transcription (RT) stop, the probing method modifies the RNA, which prevents RT extension past the site of modification, thereby allowing library preparation. Note that RT stops one nucleotide before the modification. In RT mutation, the probing method involves modifying the RT method such that a mutation is inserted at the site of reactivity. The RNA is then fragmented to allow for library preparation.
[e]We divide the studies into in vivo, in vitro, and in vitro only, where in vitro encompasses a variety of protocols, only some of which renature the RNA. In vitro only indicates that this method is not amenable to in vivo approaches.
[f]Potential to be modified for in vivo treatment.
[g]To date, in vivo but could be done in vitro with no change to reagent.
Abbreviations: CIRS-seq, chemical inference of RNA structures followed by massive parallel sequencing; CMCT, 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfonate; DMS, dimethyl sulfate; EDTA, ethylenediaminetetraacetic acid; FragSeq, fragmentation sequencing; icSHAPE, in vivo click selective 2′-hydroxyl acylation analyzed by primer extension; NA, not applicable; PARS, parallel analysis of RNA structure; PARTE, parallel analysis of RNA structure with temperature elevation; PIP-seq, protein interaction profile sequencing; RT, reverse transcription; SHAPE-MaP, selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling; WC, Watson-Crick.

as DMS (dimethyl sulfate), whereas guanosine and uridine react with the carbodiimide modifying reagent CMCT [1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfonate]. Other chemicals react with the sugar-phosphate backbone and are thus not specific to an individual nucleotide. These include the SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) reagents that acylate the 2′-OH of the sugar (125). To date, most studies have used only one of these chemical reagents. However, each reagent has its advantages, and carrying out parallel studies with different reagents will likely give the most comprehensive view of RNA folding.

Once RNA is modified or cleaved, whether by a chemical or an enzyme (**Table 1**), the sites of reactivity must be determined. Different probing techniques require different readout methods. As RNases cleave the RNA backbone, RNA structure can be determined by analyzing the cleavage sites either using end-labeled RNAs, separated by PAGE (polyacrylamide gel electrophoresis) or capillary electrophoresis, or using reverse transcription where the primers run off at the sites of cleavage; the latter approach is amenable to genome-wide methods. Chemicals such as DMS and CMCT that react on the WC face of the nucleobase block reverse transcription by preventing base pairing. SHAPE reagents that react with the sugar-phosphate backbone add bulky groups that also block reverse transcription. Both cleavage reactions and reverse transcription–blocking reactions are directly transferable to creating genome-wide libraries; as such, historically gel-based methods can be readily adapted to next-generation sequencing methods to determine the structure of RNAs in a high-throughput manner.

In Vitro Genome-Wide Methods

The first genome-wide RNA structure-probing methods involved enzymatic probes, which can be used only in vitro. FragSeq (fragmentation sequencing) (114) probes the RNA structure with RNase P1, which cleaves single-stranded regions (**Table 1**). The scoring reported in FragSeq uses ratios of nuclease to control treatments to account for any natural degradation. PARS (parallel analysis of RNA structure) uses RNase V1 and nuclease S1 to cleave the RNA at double-stranded and single-stranded regions, respectively (**Table 1**). After next-generation sequencing, a PARS score is generated, which is a ratio of V1 to S1 reads and effectively reports on the double-strandedness of RNA (43). One advantage of PARS is that it can reveal regions of the RNA that adopt more than one fold in different copies of the same molecule, as both V1 and S1 will cleave at the same region. PARS has also been adapted to include PARTE (parallel analysis of RNA structure with temperature elevation), which allows calculation of RNA folding energies by measuring RNA structures at different temperatures (119). These three methods have been applied using a biological pH (pH 7–7.5) but at typical in vitro structure-probing concentrations of Mg^{2+} (5–10 mM) and monovalent salt (100–150 mM NaCl), which are well above the physiological levels of Mg^{2+} of ∼0.5–1 mM in eukaryotes (65) and 2 mM in prokaryotes (71, 111).

Although many interesting findings arise from observations of RNA structure, one key challenge is to understand the role proteins play in affecting RNA structure. PIP-seq (protein interaction profile sequencing) was developed to partially address this challenge. PIP-seq (99, 100) uses RNase V1 and a single-stranded RNA-specific RNase (RNase ONE) to cleave double- and single-stranded regions of RNA on two pools of RNA, one of which is untreated and one of which is treated with Proteinase K to remove all proteins (**Table 1**). Although conceptually this method provides information on both protein binding and RNA structure, this type of analysis may have inherent problems because it assumes the RNA structure remains static upon removal of proteins.

RNA structure also can be affected by the solution environment chosen for in vitro studies. In PIP-seq, the buffers used for single-stranded and double-stranded nuclease probing are not identical (32) (**Table 1**), which could be problematic. In some studies, free Mg^{2+}, in particular,

which promotes RNA folding, is absent (removed by the addition of EDTA) in the RNase ONE treatment but not in the V1 treatment. Thus, a given RNA species could have assumed a different structure during RNase ONE probing versus V1 probing. In sum, although PIP-seq may provide information regarding protein-binding sites, there are significant cautions in its application to deduce RNA structure.

Chemical-based methods provide an alternative approach to probe RNA structure genome-wide in vitro. For example, CIRS-seq (chemical inference of RNA structures followed by massive parallel sequencing) (40) uses DMS and CMCT to react with the WC faces of A/C and G/U, respectively. CIRS-seq thus provides nucleotide resolution on all four bases. Prior to CIRS-seq, genome-wide analyses had used only DMS, but those applications were in vivo (see next section). To date, CIRS-seq probing has been performed after protease-based removal of the proteins associated with the RNA and at pH 8.0 with 10 mM Mg^{2+} and the addition of 0.5% NP-40, which could alter the folding of the RNA (**Table 1**).

SHAPE chemicals modify the sugar of the sugar-phosphate backbone and provide information on the flexibility of all four nucleotides, wherein flexibility is a general indicator of the absence of base pairing (125). These chemicals have been used in a variety of genome-wide methods. Typically, SHAPE modification provides a stop in reverse transcription. A variation on this method, SHAPE-MaP (98, 103), uses mutational profiling to analyze the position of a SHAPE reaction under in vitro biological-like conditions, chosen by these authors to be pH 8.0, 200 mM monovalent salt, and 3 mM Mg^{2+} (**Table 1**). This method modifies the solution conditions of a typical reverse-transcription reaction to encourage a nucleotide mutation at the SHAPE-modified nucleotide (**Table 1**). These mutations are then identified by next-generation sequencing, wherein the site of mutation is inferred to be single stranded in the original fold. The major technical advantage of this method is that the polymerase reads through the mutation, thereby allowing multiple mutations to be detected in a single RNA molecule. Mutational profiling analysis is particularly useful for identifying correlations between two positions within a single RNA but has the drawback that mutational profiling analysis requires much deeper sequencing to confidently call a nucleotide single stranded because there may be mutations incorporated during library preparation unrelated to a SHAPE adduct. This method has the potential to become an in vivo probing method by incorporating one of the cell-permeable SHAPE reagents (102, 105).

In Vivo Genome-Wide Methods

Prevailing structures often differ in vitro versus in vivo (129), owing to the influence of the cellular environment on RNA structure. Thus, although some information can be gained from in vitro genome-wide RNA structure analysis, to answer biological questions in vivo RNA structure probing is important. Currently, the only chemical that reacts with the WC face of nucleotides and that has been used in in vivo genome-wide methods is DMS. Although DMS reacts with only A and C, this reactivity performs as well as does SHAPE reactivity for RNA structure prediction (17). Structure-seq (22), DMS-seq (91), and Mod-seq (108), which were all developed at almost the same time, use DMS to react with RNA in vivo. After RNA extraction, these methods use reverse transcription to identify the sites of reaction identified by a stop in the extension (**Table 1**) (for a more comprehensive analysis of the differences between these three methods, see Reference 54). Of these methods, Structure-seq has certain advantages: It converts the RNA to DNA early in the experimental pipeline, thus requiring fewer manipulations with the more labile ribonucleic acid, and has a powerful and user-friendly computational pipeline (110) for which open access is provided in the Galaxy environment (**https://usegalaxy.org/**).

As mentioned above, certain SHAPE chemicals can also be used to map RNA structure in vivo (105). One method using an icSHAPE (in vivo click SHAPE) reaction (106) enriches for SHAPE-modified RNA segments via biotin-streptavidin isolation (**Table 1**). This isolation is possible because a copper-free click reaction produces a modified SHAPE chemical attached to a biotinylated moiety. Because SHAPE methods probe the sugar-phosphate backbone, icSHAPE provides information on the flexibility of all four nucleotides.

Studying RNA structure in vivo is important, but there are a few limitations to the methods currently available. Particularly, there are no chemicals known to modify double-stranded regions; thus, direct information is available only on single-stranded regions from studies based on chemical reactivity, both in vitro and in vivo. The protection from reactivity offered by protein binding further compounds this issue. Therefore, lack of reactivity in chemical probing studies could be due either to the RNA section being double stranded or to a protein binding the region. Nevertheless, the presence of reactivity is a highly reliable indicator of single-strandedness (22), and this information alone can greatly restrain RNA structure prediction. As such, although some areas in genome-wide methodologies need further development (discussed in Challenges and Future Directions, below), current technologies are very valuable because they make possible acquisition of substantial knowledge about a system; for instance, reactivity changes at single-stranded regions provide information about RNA switching as well as the binding and unbinding of proteins. In addition, much can be learned concerning changes in RNA structure and protein binding by comparing DMS reactivity readouts across, e.g., environmental conditions or developmental stages.

Methods to Identify Natural RNA Modifications Genome-Wide

Whereas natural modifications present in tRNA and ribosomal RNA (rRNA) have been known for many years (34), only since 2014 have genome-wide modifications been described. Recently, five RNA modifications—N^6-methyladenosine (m^6A) (23, 62), pseudouridine (ψ) (14, 93), N^1-methyladenosine (m^1A) (24, 59), 5-methylcytosine (44, 45, 107), and 2′-O-ribose methylation (7)—were identified genome-wide; the first three have also been associated with aspects of mRNA structure. As these modifications are covalent changes, they are stable; thus, in vitro studies can accurately report on their location. m^6A, which does not modify the WC face, was initially identified using antibody-based techniques (23) to pull down regions of RNA with modifications, whereas ψ was identified by a chemical reaction with CMC (N-cyclohexyl-N9-(2-morpholinoethyl)-carbodiimide metho-p-toluenesulfonate), which modifies the ψ and provides reverse-transcription stops (14, 93). m^1A has been identified in toto in mRNA by liquid chromatography–tandem mass spectrometry, whereas antibody-based pulldowns coupled with next-generation sequencing have been used to map m^1A-containing sequences genome-wide (24, 59). m^1A modifies the WC face; thus, it can also be identified by its induction of a combined stop and mismatch profile in reverse transcription (112). Biological implications of structural signatures associated with ψ, m^6A, and m^1A are described below.

INSIGHTS FROM EXPERIMENTAL GENOME-WIDE STUDIES OF THE RNA STRUCTUROME

Results from single-gene studies on RNA structure-function relationships, particularly for mRNAs, are being reassessed and extended using data from transcriptome-wide analyses. These studies are also uncovering new meta-properties—phenomena that would not be evident from studies on individual transcripts but that emerge upon collective analysis of large-scale data sets.

Next, we summarize meta-properties of the RNA structurome that have clear statistical support from genome-wide studies but whose functional impact and molecular bases often remain poorly understood. Subsequent sections then summarize meta-properties that are correlated with RNA modification, processing, stability, and translation.

This field is at a nascent stage. To date, experimental structuromes have been generated for only a few systems. The first RNA structurome to be described was that of the HIV-1 RNA genome, which was probed in vitro by Weeks and colleagues using their SHAPE method (125). Since that seminal publication in 2009, RNA structure has been probed genome-wide in yeast, *Escherichia coli*, *Arabidopsis*, *Drosophila*, *Caenorhabditis elegans*, and mouse and human tissues and cell lines (**Table 2**). As more species, systems, and conditions are analyzed, it will be important to ascertain which results reinforce the principles described in the following sections and summarized in **Table 2** and which may reveal heretofore unenvisaged mechanisms of RNA structure control. An overview of phenomena related to RNA structure and revealed via genome-wide studies is provided in **Figure 3**.

Meta-Properties of RNA Structuromes

The complete RNA structurome includes structural data on all classes of RNAs, both coding and noncoding. A few genome-wide studies have compared the extent of structure as a characteristic of noncoding RNAs versus mRNAs. Through an analysis of the temperature dependence of RNA unfolding in vitro using PARTE, Chang and colleagues (119) found that yeast noncoding RNAs (rRNA, tRNA, small nucleolar RNA, small nuclear RNA) melted at a higher temperature than did mRNAs. mRNAs showed the greatest variance in T_m, consistent with their highly diverse nature. Compared with mRNA, more double-strandedness, indicative of greater secondary structure, was found in rRNAs, tRNAs, microRNAs, small nucleolar RNAs, and small nuclear RNAs of an *Arabidopsis* in vitro structurome (130). Robust structure in noncoding RNAs is expected given their biological roles and mechanism of action.

Genome-wide structure-probing methods that include a poly(A) selection step result in underrepresentation of small RNAs, and most genome-wide structure analyses have focused on mRNAs. One robust meta-property of mRNA structuromes is a triplet repeat pattern of predicted structure and/or reactivity that is present along the coding sequences (CDSs) of mRNAs but absent from UTRs. This phenomenon is recurrent in structuromes in silico (95), in vitro (19, 40, 43, 119), and in vivo (22, 106) and has been observed in multiple organisms, including *E. coli* (19), yeast (43, 119), *Arabidopsis* (22), and mouse (40, 106) and human (120) cell lines. The underlying basis of the triplet repeat—whether completely inherent in sequence, imposed by experimental or cellular conditions that affect RNA folding and reactivity, and/or somehow emergent from biases in library construction—has yet to be resolved, nor is it known why natural selection may have favored such a pattern. We hypothesize that functionally the triplet repeat may provide a register that minimizes ribosome slipping during translation (i.e., ribosomes advancing two or four nucleotides instead of three), but this is only speculation at present. The triplet repeat is not simply imposed by ribosome association because it is seen in deproteinized in vitro samples (**Table 2**).

Another meta-property observed in multiple in vivo mRNA structuromes but with unknown functional consequences is an overall difference in average reactivity/RNA structure between the CDS and UTRs. The relative extent of structure between the CDS and UTRs appears to vary in an organism-specific manner (see **Table 2**). Moreover, this relationship may vary in different cellular compartments. For example, within *Arabidopsis*, both in vitro (57) and in vivo (22) analyses of total mRNA structure showed the CDS was more structured than were the UTRs. By contrast, in vitro analysis (32) of nuclear-localized transcripts, presumably enriched in pre-mRNAs, showed

Table 2 Phenomena observed through experimental genome-wide probing of RNA structure

Organism	Accession number, cell line, or strain	Growth phase or stage	Tissue or material	RNA of focus	Method	Condition	Phenomena observed	Reference(s)
Arabidopsis thaliana	Columbia (Col-0)	Five days	Seedlings	mRNA	Structure-seq (chemical based)	In vivo	More structure in CDS than in UTRs Low structure near start and stop codons Triplet periodicity in reactivity in CDS and not in UTRs High structure before alternative polyadenylation sites and low structure immediately after High structure in donor exons upstream of alternative splice sites Greater implied structural flexibility for stress-related genes than for housekeeping genes Inverse correlation between mRNA secondary structure and protein structure	22, 109
		Not provided	Unopened flower buds	mRNA	Single-stranded cleavage (RNase based)	In vitro	Double-stranded RNA library reads enriched in rRNA, tRNA, miRNA, snoRNA, and snRNA, suggesting these ncRNAs are more structured than are mRNAs	130
					Double- or single-stranded cleavage (RNase based)	In vitro	More structure in CDS than in UTRs Less structure in CDSs than in transposons Less structure in miRNA-binding sites than in flanking regions Positive correlation between mRNA secondary structure and smRNA levels Inverse correlation between mRNA secondary structure and transcript abundance Correlations between mRNA secondary structure and different types of histone modification	57
	UBQ10p:NTF/*ACT2p*:BirA transgenic line in Col-0 background	10 days	Nuclei	pre-mRNA	PIP-seq (RNase based)	In vitro	Less structure in CDS than in UTRs Inverse correlation between mRNA secondary structure and RBP binding Higher protein-protected site density in CDS than in UTRs Complex relationships between mRNA secondary structure and alternative splicing No relationship between mRNA secondary structure and alternative polyadenylation	32

(*Continued*)

Table 2 (*Continued*)

Organism	Accession number, cell line, or strain	Growth phase or stage	Tissue or material	RNA of focus	Method	Condition	Phenomena observed	Reference(s)
Caenorhabditis elegans	N2 worms	Mixed stages	Nematodes	mRNA	Double- or single-stranded cleavage (RNase based)	In vitro	Less structure in CDS than in UTRs Low structure near start and stop codons Less structure in miRNA-binding sites than in flanking regions Correlations between mRNA secondary structure and different types of histone modification	56
Drosophila melanogaster	DL1 culture cells	Not provided	Cells	mRNA	Double- or single-stranded cleavage (RNase based)	In vitro	Less structure in CDS than in UTRs Low structure near start and stop codons More structure in miRNA-binding sites than in flanking regions Correlations between mRNA secondary structure and different types of histone modification	56
Escherichia coli	Strain MC4100	Mid-log phase (OD600 ~ 0.4)	Cells	mRNA	Double- or single-stranded cleavage (RNase based)	In vitro	Equally structured in CDS and UTRs Triplet periodicity in reactivity in CDS and not in UTRs Low structure near start codon More structure upstream of UAA stop codon Direct correlation between CDS secondary structure and mRNA abundance	19
HIV-1	Strain NL4-3, group M, subtype B	Virions purified from non-Hodgkin's T cell lymphoma cell line	Virus	RNA genome	SHAPE (chemical based)	In vitro	~10 structured and ~7 unstructured regions identified Lack of structure at consensus sequences for constitutive splicing More structure at ribosome pause sites Sequences encoding hypervariable domains are internally unstructured and bordered by stable RNA structures More structure in mRNA sequences encoding inter- and intraprotein peptide linkers than in random sequences	125
					SHAPE-MaP (chemical based)	In vitro	Improved pseudoknot identification Disruption of U3 pseudoknot reduces viral fitness	98

Homo sapiens	K562 cells	Log phase	Cells	mRNA	DMS-seq (chemical based)	In vivo and In vitro	Less structure in vivo than in vitro	91
	Foreskin fibroblasts	80% confluence	Cells	mRNA	DMS-seq (chemical based)	In vivo and In vitro	Less structure in vivo than in vitro	91
	HeLa cells	90% confluence	Cells	mRNA	PIP-seq (RNase based)	In vitro	Uncovered many putative RBP binding motifs Overlap of some RBP binding sites with disease-linked SNPs	100
	Lymphoblastoid cell lines GM12878, GM12891, and GM12892	Not provided	Cells	mRNA	PARS (RNase based)	In vitro	Less structure in CDS than in UTRs Triplet periodicity in reactivity in CDS and not in UTRs Inverse correlation between structure at the translation start site and translation efficiency At exon-exon junctions, low structure at the AG dinucleotide at the 3′ end of the 5′ exon, with more structure at the initial nucleotide of the 3′ exon Structure-altering SNPs promote alternative splicing	120
Mus musculus	Embryonic stem cells v6.5 line	Not provided	Cells	mRNA	icSHAPE (chemical based)	In vivo and In vitro	Less structure in vivo than in vitro Less structure in mRNAs than in ncRNAs (except snoRNAs) in vivo Low structure upstream of canonical and 5′-UTR noncanonical AUG start sites High reactivity at ribosome A sites and lower reactivity at ribosome peptidyltransferase and exit sites in vivo and in vitro owing to sequence Triplet periodicity in reactivity in CDS and not in UTRs both vivo and in vitro Low structure at GGm^{6}ACU m^{6}A motifs near stop codons in vivo and in vitro	106
	KH2 embryonic stem cells	Not provided	Cells	ncRNA	FragSeq (RNase based)	In vitro	Cleavage correlated with single-stranded RNA regions in known structures of snRNAs and snoRNAs High correlation of cleavage between libraries from differentiated and undifferentiated embryonic stem cells High correlation between enzymatic cleavage and chemical probing reactivity on snRNAs and snoRNAs	114
	E14 embryonic stem cells	Not provided	Cells	mRNA	CIRS-seq (chemical based)	In vitro	Less structure in CDS than in UTRs More structure in snoRNAs, snRNAs, tRNAs, and lincRNAs than in mRNAs Low structure at Kozak sequence and stop codon Triplet periodicity in CDS and not in UTRs Lin28a RBP binding motif is preferentially single stranded	40

(*Continued*)

Table 2 (*Continued*)

Organism	Accession number, cell line, or strain	Growth phase or stage	Tissue or material	RNA of focus	Method	Condition	Phenomena observed	Reference(s)
Saccharomyces cerevisiae	Strain BY4741	Not provided	Cells	mRNA	DMS-seq (chemical based)	In vivo and In vitro	Less structure in vivo than in vitro Low (1 mM) Mg^{2+} unfolded RNAs in vitro In vivo structure increased to a level similar to in vitro upon ATP depletion High UTR structure affected production of a Venus reporter protein	91
	Strain BY4741 and *rpl26Δ*	Mid-log phase	Cells	rRNA and other ncRNA	Mod-seq (chemical based)	In vivo	More DMS reactivity at the L26 ribosomal protein binding site in 5.8S and 25S rRNAs in an L26 deletion strain	108
	Strain S288C	Log phase	Cells	mRNA	PARS (RNase based)	In vitro	More structure in CDS than in UTRs Triplet periodicity in reactivity in CDS and not in UTRs Inverse correlation between structure at the translation start site and translation efficiency	43
		Log phase	Cells	mRNA	PARTE (RNase based)	In vitro	More structure in CDS than in UTRs 5′-UTR structure less stable than both CDS and 3′ UTR Triplet periodicity in reactivity in CDS and not in UTRs Inverse correlation between structural stability at the translation start site and translation efficiency Low structural stability near start codon Less structural stability in mRNA than in ncRNAs mRNA structural stability correlates directly with transcript abundance and inversely with exosome degradation during heat shock	119
Satellite tobacco mosaic virus	STMV	Virions from infected tobacco leaves	Virus	RNA genome	SHAPE (chemical based)	In vivo and in vitro	Long-range base-pairing interactions revealed by SHAPE probing Less structure in vivo than in vitro High overall structure in the STMV RNA genome	2

Abbreviations: CDS, coding sequence; CIRS-seq, chemical inference of RNA structures followed by massive parallel sequencing; DMS, dimethyl sulfate; FragSeq, fragmentation sequencing; icSHAPE, in vivo click selective 2′-hydroxyl acylation analyzed by primer extension; lincRNA, long intergenic noncoding RNA; mRNA, messenger RNA; miRNA, microRNA; ncRNA, noncoding RNA; PARS, parallel analysis of RNA structure; PARTE, parallel analysis of RNA structure with temperature elevation; PIP-seq, protein interaction profile sequencing; pre-mRNA, precursor messenger RNA; RBP, RNA-binding protein; rRNA, ribosomal RNA; SHAPE-MaP, selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling; smRNA, small RNA; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA; STMV, satellite tobacco mosaic virus; tRNA, transfer RNA; UTR, untranslated region.

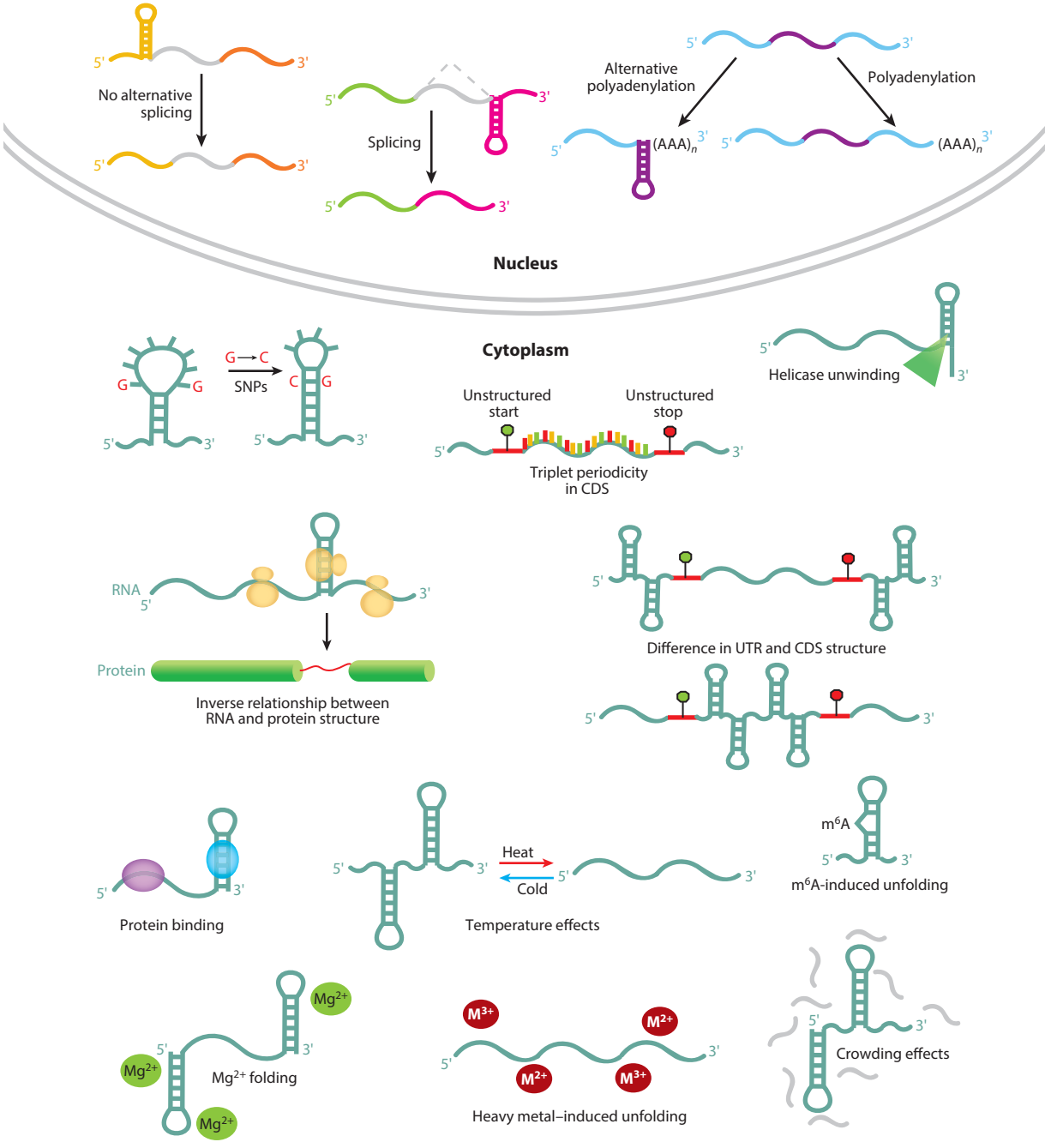


Figure 3

Relationships between RNA structure and cellular conditions and processes. Effects of various cellular conditions and processes on RNA structure, and vice versa, are illustrated. Various hairpins indicate RNA structure, although numerous additional forms of structure will exist. Abbreviations: CDS, coding sequence; SNP, single-nucleotide polymorphism; UTR, untranslated region.

the CDS was less structured than were the UTRs. Taken at face value, these results suggest that, compared with pre-mRNAs, RNA-processing events increase average structure within the processed CDS or decrease average structure in the UTRs. Why evolution may have selected for varying structural patterns in different organisms or cellular compartments remains an intriguing question.

Another type of meta-analysis involves comparison of in vivo versus in vitro structuromes. Meta-analyses from the Weissman lab suggest caution should be exercised in drawing conclusions from studies of in vitro structuromes (91). They performed a direct comparison of in vitro and in vivo structuromes of yeast and found the in vivo structurome was less structured overall than was the in vitro structurome (91). Although this result was initially unexpected, in retrospect it is perhaps less surprising because most in vitro analyses are typified by high Mg^{2+} concentrations, which tend to promote folding. Indeed, Rouskin et al. (91) found that as buffer concentrations of Mg^{2+} were lowered to 1 mM, significant unfolding of the in vitro structurome occurred. They further demonstrated that experimental ATP depletion in vivo resulted in a more in vitro–like structurome. This result suggests that energy-dependent processes, e.g., helicase unwinding of RNA, play important roles in reducing overall RNA structure in vivo. Spitale et al. (106) subsequently demonstrated that mouse embryonic stem cell lines also show less RNA structure in vivo than they do in vitro, particularly in mRNAs. A similar phenomenon was observed for the RNA genome of satellite tobacco mosaic virus when probed within the virion or in vitro (2).

Meta-Properties Relating RNA Structure and mRNA Processing

Nascent mRNAs are subjected to numerous processing steps, including capping, splicing, and polyadenylation. During and after export from the nucleus, still more levels of regulation occur, most prominently those associated with subcellular localization, translation, and turnover. Association with proteins (30, 41) can be fleeting or, as in the case of the ribosome, more long-lived. As revealed by previous single-gene studies and increasingly by RNA structurome analyses, RNA structure affects all stages of RNA processing. Conversely, throughout the lifetime of an individual RNA molecule, changes in the subcellular environment and in protein association can affect RNA structure in ways that are only beginning to be revealed by high-throughput RNA structure studies.

Alternative polyadenylation. Alternative polyadenylation occurs by cleavage and polyadenylation at alternative sites within the 3′ UTR or, infrequently, within introns or exons. Resultant differences in the sequence can affect all aspects of mRNA processing and utilization, including splicing, subcellular localization, translation, and degradation (27, 77). Alternative polyadenylation is common in both mammals and plants. Whereas sequence signatures have been identified for alternative polyadenylation sites (27) and secondary structure has been proposed to influence polyadenylation efficiency (48) and site selection (33), a possible role for RNA structure has been assessed in only two experimental genome-wide structure studies to date (22, 32). In plants, 50–80% of pre-mRNAs can undergo alternative polyadenylation (39, 97). Using in vivo Structure-seq analysis on *Arabidopsis* seedlings, Ding et al. (22) found that sites known to undergo alternative polyadenylation exhibit a region of low DMS reactivity at nucleotides −22 to −15 (i.e., upstream of the cleavage site) and a region of high DMS reactivity, indicative of low structure, at nucleotides −1 to 5 (i.e., crossing and downstream of the cleavage site). They also found that this pattern was not simply due to nucleotide composition. In vitro analysis of the nuclear *Arabidopsis* structurome, however, failed to reveal a structure signature associated with alternative polyadenylation sites (32).

Splicing. By combinatorial mathematics alone, alternative splicing, in which inclusion or exclusion of introns or exons allows a single gene to encode multiple protein isoforms, vastly increases the information content of the genome. Approximately half the genes in the model plant species *Arabidopsis* and rice and essentially all human genes (79, 83) are subject to alternative splicing, and investigators have proffered the idea that the extent of this phenomenon may functionally correlate with organism complexity (79, 83). RNA structure can regulate splicing (10, 79) in at least three ways: (*a*) by affecting assembly of the actual spliceosome; (*b*) as a component of motifs recognized by splicing factors, i.e., proteins that regulate splicing; and (*c*) by altering base pairing in ways that affect splicing choice or efficiency, independent of protein binding. As described below, transcriptome-wide information on RNA structure is starting to illuminate all three aspects and should help the field to move from case-by-case analyses (for reviews, see 10, 79, 83, 124) to general principles.

Spliceosome assembly. The spliceosome is a large ribonucleoprotein complex that assembles on and interacts with the 5′ splice site, the 3′ splice site, and a branchpoint region within the intron. Structure at any of these sites can affect splicing (42). In human lymphoblastoid cell lines from a parent-offspring trio subjected to in vitro structurome analysis with nucleases (PARS), a structure signature was observed at exon-exon junctions wherein the final AG dinucleotide at the end of the donor (5′) exon tended to be less structured, whereas the first nucleotide of the acceptor (3′) exon tended to be more structured (120), indicating that these structural aspects could constitute part of a structural signature for efficient spliceosome recognition. Single-nucleotide polymorphisms (SNPs) in the trio could also have large or minimal effects on structure; those in the first category promoted alternative splicing to a greater extent than did those that had minimal effect on structure (120).

The first in vivo genome-wide analysis of structure at alternative splice sites was performed in *Arabidopsis*. When DMS reactivity was characterized transcriptome-wide for *Arabidopsis* mRNAs susceptible to alternative splicing, significantly lower DMS reactivity was found at the 3′ end of the donor exon (5′ splice site) for transcripts in which alternative splicing had not occurred; no signature of reactivity was found at the 5′ end of the acceptor exon (3′ splice site) (22). In contrast to in vitro results from human lymphoblastoid cells (120), no pattern in donor or acceptor exon DMS reactivity was found for spliced events in *Arabidopsis*. As such, RNA structure at the 5′ end of the acceptor exon may suppress alternative splicing of these genes at the developmental stage assayed (whole etiolated seedlings), perhaps by interfering with spliceosome binding. A nucleotide composition control ruled out the possibility that the observed low reactivity in the donor exon of unspliced events arose from sequence bias. However, protein protection cannot be ruled out as an alternative or contributing factor to low DMS reactivity.

When Gosai, Gregory, and colleagues performed a transcriptome-wide study of protein binding to *Arabidopsis* nuclear mRNAs, ~40% of which were unspliced pre-mRNAs, they observed a protein-binding pattern that was a partial match to the above-mentioned DMS reactivity pattern (32). For sites subject to intron retention rather than constitutive splicing events, protein protection was greater in the 40 nucleotides upstream of the 3′ end of the donor exon, consistent with the low DMS reactivity in this region observed by Ding et al. (22). However, Gosai et al. (32) also observed elevated protein protection within the intron and at the 5′ end of the acceptor exon, regions where Ding et al. (22) had not reported low DMS reactivity. Gosai et al. (32) proposed that within-intron protein binding inhibited spliceosome assembly, thus promoting intron retention. They also reported an overall inverse relationship in *Arabidopsis* between protein binding and the extent of in vitro structure at sites of alternative splicing, although cause-effect relationships remain to be determined.

RNA sequence and structure motifs of splicing factors. Many dozens of proteins regulate splicing (29). Two prominent families of these proteins are the SR factors, which (with some exceptions) promote splicing, and the heterogeneous nuclear ribonucleoproteins (hnRNPs), which (with some exceptions) inhibit splicing (11). Computational approaches or experimental methods such as PAR-CLIP (photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation) and iCLIP (individual-nucleotide cross-linking immunoprecipitation) have identified sequence motifs for many splicing factors. However, different methods yield different results, and not all enriched motifs can be associated with a splice factor (29). These discrepancies suggest that sequence alone is an insufficient predictor of splice-factor binding sites, a premise supported by two recent genome-wide studies.

The first study investigated hnRNPC, an hnRNP that contains two RNA recognition motifs that favor binding to polyU tracts (50). Parallel transcriptome-wide analysis of m^6A covalent modifications by m^6A immunoprecipitation with anti-m^6A antibody and of hnRNPC binding sites by PAR-CLIP revealed overrepresentation of m^6A sites in RNA regions associated with hnRNPC-binding targets (63). On the basis of biochemical studies with specific hnRNPC targets, the authors (63) proposed that m^6A in the complementary strands of U-rich hairpins promotes hnRNPC binding by weakening the hairpin secondary structure. Support for m^6A regulation of RNA structure and hnRNPC binding as a genome-wide mechanism was provided by the observation that knockdown of two m^6A methyltransferases reduced hnRNPC binding at more than 2,500 sites and also affected alternative splicing, promoting intron exclusion as deduced from the reduced presence of intron-exon junctions in RNA-Seq data (63).

The second study focused on the "feminizing locus on X" (Fox) family of RNA-binding proteins, which function in tissue-specific control of RNA splicing. Spitale and colleagues (106) combined in vivo and in vitro icSHAPE structure data with the known UGCAUG sequence targets of the RNA-binding protein Rbfox2 (fox-1 homolog in mouse) identified from iCLIP experiments. Their study showed significantly improved predictions of Rbfox2-binding sites in mouse embryonic stem cells compared with those from sequence alone. It is a virtual certainty that the binding of essentially all splicing regulators is similarly regulated by combined patterns of sequence and structure. Transcriptome-wide RNA structure analysis combined with computational and experimental sequence searches provides a powerful tool to elucidate these signatures with greater confidence.

Regulation of splicing by non-protein-bound RNA structures. Alternative splicing can also be affected by RNA structural motifs that do not directly interact with splicing factors or with the spliceosome. Given computational predictions of RNA structure, stem-loop structures were hypothesized to form downstream of Rbfox-binding sites to bring distally bound Rbfox proteins into proximity with their exon targets. In support of this hypothesis, targeted analysis on exon 11a of the enabled homolog (*ENAH*) gene showed that experimental disruption of the predicted stem structure impaired splicing, whereas mutations that strengthened base pairing in the stem improved splicing efficiency (67). Improved structure prediction afforded by in vivo probing will facilitate identification of such novel structure-based mechanisms across the entire RNA structurome.

One of the most dramatic examples of long-distance regulation of alternative splicing by RNA structures that interact with neither the spliceosome nor splicing factors is regulation by the eukaryotic thiamine pyrophosphate (TPP) riboswitch (79), which is present in the 3′ UTR of the *TPP* gene of fungi, algae, and plants (117). In the fungus *Neurospora crassa*, TPP binding favors a long-distance (>500 nucleotides distant) base-pairing interaction that loops out internal 5′ splice sites, thereby favoring a specific alternative splicing configuration (58). In contrast to the approximately two dozen experimentally validated classes of riboswitches in prokaryotes (9), the

TPP riboswitch is the only known eukaryotic riboswitch identified to date, but high-throughput screening for metabolite effects on RNA structuromes may identify eukaryotic RNA sequences that perform as functional riboswitches, perhaps with the involvement of proteins.

Stability and degradation. The stability of an RNA molecule within the living cell can be considered in terms of both its structural dynamics and its lifetime, i.e., its susceptibility to degradation. A study of the yeast structurome (119) found evidence that these two variables are interrelated. Analysis of in vitro structural stability using PARTE, coupled with transcriptome analysis, showed that mRNAs with low average melting temperatures (i.e., less structure) decreased most rapidly in abundance following heat shock; thus, structural stability and transcript stability are directly related. The exosome provides the major 3′-to-5′ exonucleolytic processing machinery in yeast, and mutational inactivation of the cytoplasmic exosome stabilized levels of these mRNAs following a heat-shock event, suggesting that their unfolding targets them for degradation (119).

Another way to assess the relationship between mRNA structure and transcript longevity is through parallel analysis of the structurome and degradome, the latter defined as those transcripts that lack a 5′ cap but retain a 5′-monophosphate and thus have arisen from small-RNA–directed cleavage. Such an analysis has been reported by Gregory and coworkers for *Arabidopsis* using an in vitro structurome and previously published degradome data, along with small-RNA-Seq analysis (57). An inverse relationship between the extent of in vitro structure and mRNA abundance was found, which is opposite to the relationship found for heat-shocked yeast (119) as well as for a combined in vitro structurome and RNA-Seq analysis in *E. coli* (19). Degradome and small-RNA-Seq analysis implied that the inverse relationship in *Arabidopsis* was due at least in part to the greater susceptibility of structured RNAs to degradation by small-RNA–based mechanisms (57), which is somewhat unexpected given that structure is theoretically expected to sequester sequence within the target mRNA away from base pairing with small RNAs. This study (57) also compared in vitro structurome data with previous genome-wide histone modification data sets and found that, within the CDS, structured mRNA regions were correlated with heterochromatic histone modifications, while less-structured regions of mRNAs were correlated with euchromatin-associated histone modifications (57); similar correlations were found in *Drosophila* and *C. elegans* (56). These results suggest an intriguing mechanistic connection between RNA structure and the epigenetic pathways of small-interfering RNA, an interesting topic to address with in vivo data sets.

Epigenetic pathways involve covalent modifications to chromatin. In recent years, modifications of mRNAs, resulting in the so-called epitranscriptome, have become increasingly recognized as important regulatory events. In particular, ψ and m^6A modifications have apparent implications for RNA structural stability and half-life. With regard to structural stability, ψ can base pair noncanonically with all four nucleotides. Experiments on oligonucleotides show this base pairing results in a significant increase in RNA stability compared with that for U in the same position, owing in part to an extra hydrogen bond donor in ψ (38, 46). RNA structurome analyses are required to determine whether this effect prevails in longer RNAs and in living cells. With regard to RNA lifetime, when heat-shock–induced pseudouridylation in yeast was abrogated via knockout of one of the pseudouridine synthases, target mRNA levels were reduced compared to wild-type levels. The authors speculated that increased structural stability imposed by pseudouridylation may extend mRNA lifetime, perhaps by marking these transcripts for spatial sequestration away from degradatory machinery (93).

Whereas ψ is relatively rare in mRNAs, reported at only several hundred mRNA sites in studies on yeast and human cells (14, 93), m^6A is a common covalent modification of mRNAs, with many thousands of mRNA sites reported for mouse and human cell lines (23) as well as mouse (81)

and human (23) brain tissue. m^6A sites are conserved between mouse and human. They are also enriched near stop codons and within long exons as well as upstream of microRNA-binding sites within 3′ UTRs, while they are rarely found in poly(A) tails (81).

Consensus motifs for m^6A identified by classical biochemical approaches include G(m^6A)C and A(m^6A)C (20), and more recent genome-wide anti-m^6A probing has revealed additional information, including a significant representation of G[A or G]ACU sites (81). Genome-wide assessment of G[A or G]ACU m^6A sites using icSHAPE combined with anti-m^6A probing (106) showed that m^6A unfolds RNA structure locally, as indicated by stronger icSHAPE reactivity at m^6A-modified GGACU sites than at unmodified GGACU sites. This effect was seen in both mRNAs and lncRNAs, suggesting it is a general effect (90). In *Mettl3* methyltransferase knockout cells, reaction of canonical m^6A motifs with icSHAPE reagent was significantly reduced, suggesting that loss of m^6A conferred increased structure (106). Thus, lack of RNA structure arises from m^6A modification rather than the converse, i.e., m^6A methyltransferases favoring unstructured sites. As described in more detail above, in U-rich stems, m^6A weakens RNA structure and promotes splicing-factor binding to the opposite strand.

Researchers have also studied stability in the sense of RNA turnover (121). In a human HeLa cell line, m^6A modification reduces stability via interaction with the m^6A reader protein YTH domain family 2 (YTHDF2) (121), which associates transcripts with mRNA processing bodies (P-bodies) that are known sites of RNA degradation. RNA lifetime profiling using the transcriptional inhibitor actinomycin D, of a HeLa YTHDF2 knockdown line (121) as well as a mouse embryonic stem cell methyltransferase knockdown line (123), confirmed increased lifetime of YTHDF2/m^6A target mRNAs relative to control lines.

Translation. Protein translation rates vary across the translatome. Single-gene studies over the past several decades have provided strong evidence that the initiation step commonly limits translation rates (85) and that RNA structure around the start site impedes translation initiation (51, 53, 85). A genome-wide comparison of the in silico RNA structurome with experimental polysome profiling for allelic variants in hybrid mouse fibroblasts supports these assertions, as alleles with lower in silico–predicted mRNA secondary structure around the transcription start site had greater translation efficiency (defined as the ratio between ribosome-associated mRNA and total mRNA abundance) (37). Ribosome profiling in yeast (86), coupled with a transcriptome-wide sliding window analysis of either in vivo or in vitro structure along transcripts (91), identified low structure in the first window as the strongest regulator of translational efficiency (defined as total amount of protein produced per mRNA), consistent with previous genome-wide in vitro (32, 43, 57) and in silico structural analyses (5, 25, 37). In vitro and in vivo structurome studies in systems as disparate as *Arabidopsis* (22), yeast (91, 119), *Drosophila*, *C. elegans* (56), mouse and human cell lines (40, 106, 120), and *E. coli* (19) have revealed that low structure upstream of the translation start site in comparison with flanking regions is a conserved meta-property (**Table 2**). Low structure is also typically observed near the stop codon in the eukaryotic systems examined to date (22, 40, 43, 56, 106) (**Table 2**). The observation of these properties in in silico, in vitro, and in vivo structuromes suggests that this structural functionality is encoded in the primary sequence of the RNA.

Structure is affected not only by primary sequence and environmental conditions, but also by covalent modifications of the RNA. In both human and mouse cells, m^1A modification is clustered around canonical and noncanonical start sites, as well as around the first splice site. m^1A modification is underrepresented in 3′ UTRs relative to 5′ UTRs and CDSs (24, 59). Intriguingly, no such preferential patterns are seen in either budding (*Saccharomyces cerevisiae*) or fission yeast (*Saccharomyces pombe*) (24). Unlike m^6A, sites of m^1A modification are not characterized by a specific

sequence motif but instead occur in regions of high GC content. These regions also show stronger structure as deduced from PARS (24), which is expected given the strong base pairing of GC. On the basis of a comparison of the m^1A methylome with previously published ribosome profiling and proteomics data sets, m^1A modification promotes ribosome association and protein production. However, whether translation is promoted via m^1A effects on structure, e.g., via disruption of WC base-pairing or charge-charge interactions (m^1A is cationic), or via m^1A-induced alterations in binding of translation initiation factors or other proteins, awaits further investigation.

The initiation of translation for a given protein is not uniform across an organism's lifetime but is modulated by cell type, developmental stage, and environmental conditions. A role of another covalent modification of RNA, m^6A, in stress regulation of translation initiation was recently uncovered. m^6A modifications are recognized by reader proteins, including YTHDF1 and YTHDF2. YTHDF1 also interacts with the translation machinery, thereby enhancing translation of mRNAs with m^6A-modified 3′ UTRs (122). Although m^6A is rarely found in 5′ UTRs, heat shock enriches its presence there (80, 122) through a mechanism in which the reader protein YTHDF2 protects m^6A modification from erasure by the fat-mass and obesity-associated (FTO) demethylase (131). Intriguingly, the presence of m^6A in 5′ UTRs allows cap-independent translation (80, 131) by providing a binding site for eukaryotic initiation factor 3 (eIF3) (80), thus providing an alternative mechanism for accelerated initiation of translation under stress. Whether the local unfolding typically induced by m^6A is important for eIF3 binding and is enhanced by these stresses has yet to be explicitly elucidated.

m^6A and ψ also appear to play a role in regulating translation initiation by the innate immune response. Upon RNA binding, the kinase PKR inhibits translation by phosphorylating the translation initiation factor eIF-2α. In vitro experiments have shown that PKR activation is abrogated if m^6A or ψ is present in the RNA (1, 82). m^6A has no effect and ψ has only a moderate effect on RNA-binding affinity to PKR, suggesting that reduced activation of PKR by these modified RNAs may be due primarily to effects at the level of RNA secondary structure. Because most bacteria and some viruses lack RNA modifications, the presence of these modifications in "self" RNA may help to distinguish self from nonself and thus to maintain translation initiation in the absence of infection.

Downstream of translation initiation, translation rates along the transcript are not uniform (89). Studies on model oligonucleotides have shown that the presence of RNA secondary structure promotes ribosome pausing (126). However, in vivo, the situation is considerably more complicated, reflecting interplay between multiple parameters, including RNA structure, tRNA abundance, and codon choice (72). One study using in vitro structurome data from yeast (43) and in silico–predicted structures in *E. coli* proposed that rapid translation of codons associated with highly abundant tRNAs is offset by greater RNA structure associated with these codons, leading to a smoothing of translation rate (31). In vivo structurome data in *Arabidopsis* have not shown any relationship between codon usage and the extent of RNA structure (Y. Tang, S.M. Assmann, P.C. Bevilacqua, unpublished results), which may reflect a dichotomy between codon usage and codon optimality as reflected in tRNA abundance, a dichotomy between in vivo and in vitro/in silico structuromes, or species-specific effects.

A dichotomy between conclusions based on in vitro and those based on in vivo structurome data was also evident when the relationship between RNA secondary structure and structure of the encoded protein was evaluated in *Arabidopsis*. A direct correlation between regions of mRNA structure and protein structure was reported using *Arabidopsis* in vitro structurome data (115). However, in vivo structurome data revealed the opposite: a correlation between high DMS reactivity (indicative of reduced RNA secondary structure) and folded domains within the encoded protein (109). Further, analysis of the subset of proteins with intrinsically disordered regions

revealed that the disordered regions tend to be encoded by regions of the mRNA with low DMS reactivity. Decreased reactivity of RNA regions that encode protein-domain junctions or intrinsically disordered regions suggests increased RNA structure that may slow translation and allow time for the nascent upstream protein domain or ordered region of the protein to fold, thereby reducing protein misfolding (109). This idea was first put forth for the HIV-1 structurome/proteome, where a similar inverse relationship was observed (125).

Even though analysis of meta-properties is invaluable, analyses focused solely on the average behavior of the structurome can obscure important aspects of structural diversity. For example, when the structures of ~10,000 *Arabidopsis* mRNAs were predicted either with (in vivo) or without (in silico) experimental restraints, an approximately bell-shaped distribution in positive predictive value (PPV), i.e., the percentage of base pairs present identically in silico and in vivo, was observed, revealing a wide range of behaviors (22). Analysis of the upper 5% tail of this distribution revealed that high-PPV mRNAs tend to be enriched in mRNAs coding for proteins involved in housekeeping functions such as gene regulation, suggesting that these mRNAs may have evolved highly resilient structures allowing their stable presence in the transcriptome and, by inference, in the encoded proteome. By contrast, analysis of the lowest 5% of the PPV distribution revealed an enrichment of mRNAs encoding proteins involved in stress sensing and response, including many stresses that affect RNA structure, such as temperature and heavy metals. These mRNAs had features indicative of greater structural flexibility, such as larger maximum loop length and higher average free energy per nucleotide, leading us to hypothesize (22) that the dynamic structural nature of these mRNAs enables their production, processing, and translation to be highly regulated in response to environmental factors. However, using in vitro structurome analysis, the Gregory lab reached a different conclusion: They found that the set of mRNAs with low structure in vitro was enriched in basic biological processes (57), again highlighting how methodological approaches can influence RNA structuromes and ensuing biological interpretations.

CHALLENGES AND FUTURE DIRECTIONS

This is an exciting time for the RNA structure field. A proliferation of new methods has allowed us to begin to address previously unanswerable questions, and insights arising from analyses of large-scale data sets have enabled us to pose entirely new questions. Future directions comprise areas in which the necessary tools are essentially available but have not yet been applied as well as those for which additional technical innovations are required. Examples in the former category include more extensive genome-wide structure probing of additional classes of RNAs (such as all classes of small RNAs as well as organellar RNAs), genome-wide structure probing in a cell-specific and tissue-specific manner, and additional genome-wide structure-probing experiments to study RNA structure responses to abiotic and biotic stresses and signals. More studies in which the RNA structurome is assessed in genetic backgrounds wherein hypothesized regulatory proteins of RNA structure have been knocked out or overexpressed are also needed. It is also important to expand genome-wide structure probing to new organisms beyond the few model species and cultured cell lines that have been assessed to date. Such studies may lead to identification of new functional classes of RNAs: For example, although RNA-structure–based thermometers have been identified in prokaryotes, their presence in eukaryotes remains elusive but might be revealed by genome-wide analyses, particularly in nonhomeothermic organisms such as poikilothermic animals, fungi, and plants (78).

Research also needs to move from describing correlations to uncovering causative mechanisms. Although genome-wide studies have revealed striking correlations between RNA structure and numerous aspects of transcriptional and translational control (discussed above), in many cases we

still do not know why such correlations have been selected for, i.e., the underlying mechanistic underpinnings at both the evolutionary and molecular levels. Understanding these mechanisms will require the concerted and creative application of molecular biological and chemical methods far beyond structure probing and prediction alone. Such information will also be relevant to efforts to develop new drugs to combat infectious diseases affecting both humans and our food sources (livestock as well as crop plants). In the fight against pathogens, secondary-structure information can be used to optimize the screening or design of small-molecule RNA ligands as candidate pharmaceuticals (6) and to identify single-stranded regions as optimal targets for antisense oligonucleotide technologies. Such information will also aid efforts to treat genetic diseases that arise from RNA misediting or mispairing events, such as Huntington's disease and myotonic dystrophy type 1 (52), or that are associated with covalent modifications of RNA, such as obesity related to polymorphisms in the *FTO* gene, which encodes an eraser of m^6A modification (81).

SNPs that directly affect RNA structure are another emerging topic of great interest. Those that change structure have been dubbed riboSNitches by the Laederach group (35). riboSNitches first predicted computationally by Laederach and colleagues (35) are associated with half a dozen genetic diseases, and several of the implicated structural changes were subsequently verified experimentally (104). A genome-wide study from the Chang lab identified almost 2,000 candidate riboSNitches within ~12,000 SNPs in the in vitro structuromes of lymphoblast cell lines from a mother-father-offspring trio (120). This result suggests that the majority of SNPs will not affect RNA structure. However, 22 of the candidate riboSNitches were in regions that had been associated with known human etiologies via genome-wide association study. SNP-induced structural alterations can be rescued by nearby second-site mutations, and evidence of natural selection for such compensatory genetic changes has been found via computational analysis that predicted more than 400 of such paired SNPs in UTRs of the human genome (74), providing a rich source of information for follow-up via wet-bench analyses.

RNA modifications may also modulate structure, either directly or indirectly, via recruitment of RNA-binding proteins. There are currently tools for genome-wide identification of five modifications—m^6A, m^1A, ψ, 5-methylcytosine, and 2′-O-ribose methylation (7, 44, 45, 107). However, we are only starting to uncover the effects of the resultant "epitranscriptomes" on RNA structure and gene regulation. Moreover, these modifications may be just the tip of the iceberg. tRNAs undergo ~90 different modifications, approximately a dozen of which have been detected in mRNAs (12, 60, 92). New methodologies that can reveal these modifications genome-wide must be developed before we can determine whether all the modifications found in tRNAs also occur in mRNAs and lncRNAs and the extent to which such modifications may impact structure-function relationships.

Development of new computational methods is also of paramount concern to address the aforementioned challenge of improving algorithms to predict ensembles of structures wherein RNAs adopt more than one structure. This topic is worthy of a review in its own right, and the reader is referred to References 55 and 94 for more discussion. In the experimental realm, development of an in vivo probe that would specifically identify base-paired regions would be greatly beneficial. Likewise, a reagent that could distinguish between protections arising from protein binding and those arising from base pairing would yield a quantum advance by providing comprehensive data on base pairing along transcripts and across transcriptomes. Some progress has been made very recently in this direction (3, 70, 96). Such probes would help determine the extent to which the above-mentioned discrepancies regarding in vivo versus in vitro protections that have arisen in structure-probing studies are the result of protein binding in vivo or of RNA refolding in nonbiological test-tube conditions. Finally, although this review and the field of RNA

structure determination are currently focused on RNA secondary structure, development of new computational and empirical methods for genome-wide determination of RNA tertiary structure (**Figure 1*d***) in vivo is a grand challenge that will occupy the field for many years to come.

SUMMARY POINTS

1. RNA folds into complex secondary and tertiary structures.
2. RNA folding in living cells is influenced by myriad environmental and physical factors. Modifications of RNA and RNA-protein interactions also influence RNA structure.
3. In vitro methods based on nucleases and chemical probing to investigate genome-wide RNA secondary structure are available.
4. In vivo methods for genome-wide chemical probing of RNA secondary structure have recently been developed.
5. Methods to probe both nucleobases and the sugar-phosphate backbone exist.
6. Genome-wide structure-probing methods are revealing new meta-properties of RNA structure related to all aspects of mRNA processing, including polyadenylation, splicing, covalent modification, translation, and turnover.
7. Changes in RNA structure constitute a fundamental aspect of gene regulation in response to changing environmental conditions that we are only beginning to understand.

FUTURE ISSUES

1. Which features of in vitro RNA structuromes faithfully reflect in vivo structuromes, and which do not?
2. What new information will be revealed from class-, tissue-, and developmental-stage–specific in vivo RNA structuromes?
3. How does the plasticity of the RNA structurome help organisms cope with biotic and abiotic environmental stresses?
4. How does the complete RNA epitranscriptome influence RNA structure and function?
5. How can we identify additional RNA modifications and their impact on RNA structure genome-wide?
6. How can we further improve in vivo structural probing of RNA? How can we better distinguish, both in vivo and genome-wide, RNA base pairing and protein protection?
7. Can we develop methods to probe all key atoms in an RNA molecule in vivo?
8. How can we best predict the ensemble of structures for a given RNA in the living cell?
9. How can we best design RNA-based therapies against infectious diseases?
10. How can we best design drugs to repair aberrations in RNA structure that lead to genetic disease?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Research on genome-wide aspects of RNA structure in the Assmann and Bevilacqua laboratories is supported by National Science Foundation grant NSF-IOS-1339282 and by an Innovation Award from Penn State University.

LITERATURE CITED

1. Anderson BR, Muramatsu H, Nallagatla SR, Bevilacqua PC, Sansing LH, et al. 2010. Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. *Nucleic Acids Res.* 38:5884–92
2. Archer EJ, Simpson MA, Watts NJ, O'Kane R, Wang B, et al. 2013. Long-range architecture in a viral RNA genome. *Biochemistry* 52:3182–90
3. Aw JGA, Shen Y, Wilm A, Sun M, Lim XN, et al. 2016. In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell* 62:603–17
4. Bangerte BW, Chan SI. 1969. Proton magnetic resonance studies of ribose dinucleoside monophosphates in aqueous solution. 2. Nature of base-stacking interaction in adenylyl-(3′ → 5′)-cytidine and cytidylyl-(3′ → 5′)-adenosine. *J. Am. Chem. Soc.* 91:3910–21
5. Bentele K, Saffert P, Rauscher R, Ignatova Z, Bluthgen N. 2013. Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* 9:675
6. Bernat V, Disney MD. 2015. RNA structures as mediators of neurological diseases and as drug targets. *Neuron* 87:28–46
7. Birkedal U, Christensen-Dalsgaard M, Krogh N, Sabarinathan R, Gorodkin J, Nielsen H. 2015. Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem. Int. Ed.* 54:451–55
8. Blyn LB, Risen LM, Griffey RH, Draper DE. 2000. The RNA-binding domain of ribosomal protein L11 recognizes an rRNA tertiary structure stabilized by both thiostrepton and magnesium ion. *Nucleic Acids Res.* 28:1778–84
9. Breaker RR. 2011. Prospects for riboswitch discovery and analysis. *Mol. Cell* 43:867–79
10. Buratti E, Baralle FE. 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol.* 24:10505–14
11. Busch A, Hertel KJ. 2012. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip. Rev. RNA* 3:1–12
12. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, et al. 2011. The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* 39:D195–201
13. Cantor CR, Jaskunas SR, Tinoco I. 1966. Optical properties of ribonucleic acids predicted from oligomers. *J. Mol. Biol.* 20:39–62
14. Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV. 2014. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515:143–46
15. Chadalavada DM, Cerrone-Szakal AL, Bevilacqua PC. 2007. Wild-type is the optimal sequence of the HDV ribozyme under cotranscriptional conditions. *RNA* 13:2189–201
16. Conn GL, Draper DE, Lattman EE, Gittis AG. 1999. Crystal structure of a conserved ribosomal protein-RNA complex. *Science* 284:1171–74
17. Cordero P, Kladwang W, VanLang CC, Das R. 2012. Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry* 51:7037–39
18. Cruz JA, Westhof E. 2009. The dynamic landscapes of RNA architecture. *Cell* 136:604–9
19. Del Campo C, Bartholomaus A, Fedyunin I, Ignatova Z. 2015. Secondary structure across the bacterial transcriptome reveals versatile roles in mRNA regulation and function. *PLOS Genet.* 11:e1005613

20. Dimock K, Stoltzfus CM. 1977. Sequence specificity of internal methylation in B77 avian sarcoma virus RNA subunits. *Biochemistry* 16:471–78
21. Ding Y, Kwok CK, Tang Y, Bevilacqua PC, Assmann SM. 2015. Genome-wide profiling of *in vivo* RNA structure at single-nucleotide resolution using structure-seq. *Nat. Protoc.* 10:1050–66
22. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. 2014. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505:696–700
23. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, et al. 2012. Topology of the human and mouse m^6A RNA methylomes revealed by m^6A-seq. *Nature* 485:201–6
24. Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, Peer E, Kol N, et al. 2016. The dynamic N^1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* 530:441–46
25. Dvir S, Velten L, Sharon E, Zeevi D, Carey LB, et al. 2013. Deciphering the rules by which 5′-UTR sequences affect protein expression in yeast. *PNAS* 110:E2792–801
26. Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res.* 15:9109–28
27. Elkon R, Ugalde AP, Agami R. 2013. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* 14:496–506
28. Flynn RA, Zhang QC, Spitale RC, Lee B, Mumbach MR, Chang HY. 2016. Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nat. Protoc.* 11:273–90
29. Fu XD, Ares M Jr. 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15:689–701
30. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. 2008. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 582:1977–86
31. Gorochowski TE, Ignatova Z, Bovenberg RA, Roubos JA. 2015. Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res.* 43:3022–32
32. Gosai SJ, Foley SW, Wang D, Silverman IM, Selamoglu N, et al. 2015. Global analysis of the RNA–protein interaction and RNA secondary structure landscapes of the *Arabidopsis* nucleus. *Mol. Cell* 57:376–88
33. Graveley BR, Fleming ES, Gilmartin GM. 1996. RNA structure is a critical determinant of poly(A) site recognition by cleavage and polyadenylation specificity factor. *Mol. Cell. Biol.* 16:4942–51
34. Grosjean H. 2005. Modification and editing of RNA: historical overview and important facts to remember. *Topics Curr. Genet.* 12:1–22
35. Halvorsen M, Martin JS, Broadaway S, Laederach A. 2010. Disease-associated mutations that alter the RNA structural ensemble. *PLOS Genet.* 6:e1001074
36. Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, et al. 1965. Structure of a ribonucleic acid. *Science* 147:1462–65
37. Hou J, Wang X, McShane E, Zauber H, Sun W, et al. 2015. Extensive allele-specific translational regulation in hybrid mice. *Mol. Syst. Biol.* 11:825
38. Hudson GA, Bloomingdale RJ, Znosko BM. 2013. Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides. *RNA* 19:1474–82
39. Hunt AG. 2011. RNA regulatory elements and polyadenylation in plants. *Front. Plant Sci.* 2:109
40. Incarnato D, Neri F, Anselmi F, Oliviero S. 2014. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* 15:491
41. Jackson RJ, Hellen CU, Pestova TV. 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11:113–27
42. Jin YF, Yang Y, Zhang P. 2011. New insights into RNA secondary structure in the alternative splicing of pre-mRNAs. *RNA Biol.* 8:450–57
43. Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, et al. 2010. Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467:103–7
44. Khoddami V, Cairns BR. 2013. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* 31:458–64
45. Khoddami V, Yerra A, Cairns BR. 2015. Experimental approaches for target profiling of RNA cytosine methyltransferases. *Methods Enzymol.* 560:273–96

46. Kierzek E, Malgowska M, Lisowiec J, Turner DH, Gdaniec Z, Kierzek R. 2014. The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.* 42:3492–501
47. Kim SH, Quigley GJ, Suddath FL, McPherson A, Sneden D, et al. 1973. Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science* 179:285–88
48. Klasens BI, Das AT, Berkhout B. 1998. Inhibition of polyadenylation by stable RNA secondary structure. *Nucleic Acids Res.* 26:1870–76
49. Knapp G. 1989. Enzymatic approaches to probing of RNA secondary and tertiary structure. *Method Enzymol.* 180:192–212
50. Konig J, Zarnack K, Rot G, Curk T, Kayikci M, et al. 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* 17:909–15
51. Kozak M. 2005. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 361:13–37
52. Krzyzosiak WJ, Sobczak K, Wojciechowska M, Fiszer A, Mykowska A, Kozlowski P. 2012. Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.* 40:11–26
53. Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324:255–58
54. Kwok CK, Tang Y, Assmann SM, Bevilacqua PC. 2015. The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. *Trends Biochem. Sci.* 40:221–32
55. Leamy KA, Assmann SM, Mathews DH, Bevilacqua PB. 2016. Bridging the gap between in vitro and in vivo RNA folding. *Q. Rev. Biol.* In press
56. Li F, Zheng Q, Ryvkin P, Dragomir I, Desai Y, et al. 2012. Global analysis of RNA secondary structure in two metazoans. *Cell Rep.* 1:69–82
57. Li F, Zheng Q, Vandivier LE, Willmann MR, Chen Y, Gregory BD. 2012. Regulatory impact of RNA secondary structure across the *Arabidopsis* transcriptome. *Plant Cell* 24:4346–59
58. Li S, Breaker RR. 2013. Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic Acids Res.* 41:3022–31
59. Li X, Xiong X, Wang K, Wang L, Shu X, et al. 2016. Transcriptome-wide mapping reveals reversible and dynamic N^1-methyladenosine methylome. *Nat. Chem. Biol.* 12:311–16
60. Limbach PA, Crain PF, Mccloskey JA. 1994. The modified nucleosides of RNA: summary. *Nucleic Acids Res.* 22:2183–96
61. Lin Y, May GE, Joel McManus C. 2015. Mod-seq: a high-throughput method for probing RNA secondary structure. *Methods Enzymol.* 558:125–52
62. Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR. 2015. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* 12:767–72
63. Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T. 2015. N^6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature* 518:560–64
64. Lockard RE, Kumar, A. 1981. Mapping tRNA structure in solution using double-strand-specific ribonuclease V1 from cobra venom. *Nucleic Acids Res.* 9:5125–40
65. London RE. 1991. Methods for measurement of intracellular magnesium: NMR and fluorescence. *Annu. Rev. Physiol.* 53:241–58
66. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, et al. 2011. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6:26
67. Lovci MT, Ghanem D, Marr H, Arnold J, Gee S, et al. 2013. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.* 20:1434–42
68. Lowman HB, Draper DE. 1986. On the recognition of helical RNA by cobra venom V1 nuclease. *J. Biol. Chem.* 261:5396–403
69. Lu ZJ, Turner DH, Mathews DH. 2006. A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation. *Nucleic Acids Res.* 34:4912–24
70. Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, et al. 2016. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165:1267–79
71. Lusk JE, Williams RJ, Kennedy EP. 1968. Magnesium and the growth of *Escherichia coli*. *J. Biol. Chem.* 243:2618–24

72. Mao YH, Liu HL, Liu YL, Tao SH. 2014. Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 42:4813–22
73. Markham NR, Zuker M. 2008. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.* 453:3–31
74. Martin JS, Halvorsen M, Davis-Neulander L, Ritz J, Gopinath C, et al. 2012. Structural effects of linkage disequilibrium on the transcriptome. *RNA* 18:77–87
75. Mathews DH. 2004. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA* 10:1178–90
76. Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288:911–40
77. Mayr C. 2016. Evolution and biological roles of alternative 3′ UTRs. *Trends Cell Biol.* 26:227–37
78. McClung CR, Davis SJ. 2010. Ambient thermometers in plants: from physiological outputs towards mechanisms of thermal sensing. *Curr. Biol.* 20:R1086–92
79. McManus CJ, Graveley BR. 2011. RNA structure and the mechanisms of alternative splicing. *Curr. Opin. Genet. Dev.* 21:373–79
80. Meyer KD, Patil DP, Zhou J, Zinoviev A, Skabkin MA, et al. 2015. 5′ UTR m^6A promotes cap-independent translation. *Cell* 163:999–1010
81. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. 2012. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* 149:1635–46
82. Nallagatla SR, Bevilacqua PC. 2008. Nucleoside modifications modulate activation of the protein kinase PKR in an RNA structure-specific manner. *RNA* 14:1201–13
83. Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463:457–63
84. Ouyang ZQ, Snyder MP, Chang HY. 2013. SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res.* 23:377–87
85. Pelletier J, Sonenberg N. 1987. The involvement of mRNA secondary structure in protein synthesis. *Biochem. Cell Biol.* 65:576–81
86. Pop C, Rouskin S, Ingolia NT, Han L, Phizicky EM, et al. 2014. Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol. Syst. Biol.* 10:770
87. Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinform.* 11:129
88. Rich A, Davies DR. 1956. A new two-stranded helical structure: polyadenylic acid and polyuridylic acid. *J. Am. Chem. Soc.* 78:3548–49
89. Richter JD, Coller J. 2015. Pausing on polyribosomes: Make way for elongation in translational control. *Cell* 163:292–300
90. Roost C, Lynch SR, Batista PJ, Qu K, Chang HY, Kool ET. 2015. Structure and thermodynamics of N^6-methyladenosine in RNA: a spring-loaded base modification. *J. Am. Chem. Soc.* 137:2107–15
91. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505:701–5
92. Rozenski J, Crain PF, McCloskey JA. 1999. The RNA modification database: 1999 update. *Nucleic Acids Res.* 27:196–97
93. Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, et al. 2014. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159:148–62
94. Seetin MG, Mathews DH. 2012. RNA structure prediction: an overview of methods. *Methods Mol. Biol.* 905:99–122
95. Shabalina SA, Ogurtsov AY, Spiridonov NA. 2006. A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* 34:2428–37
96. Sharma E, Sterne-Weiler T, O'Hanlon D, Blencowe BJ. 2016. Global mapping of human RNA-RNA interactions. *Mol. Cell* 62:618–26
97. Shen YJ, Venu RC, Nobuta K, Wu XH, Notibala V, et al. 2011. Transcriptome dynamics through alternative polyadenylation in developmental and environmental responses in plants revealed by deep sequencing. *Genome Res.* 21:1478–86

98. Siegfried NA, Busan S, Rice GM, Nelson JAE, Weeks KM. 2014. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* 11:959–65
99. Silverman IM, Gregory BD. 2015. Transcriptome-wide ribonuclease-mediated protein footprinting to identify RNA–protein interaction sites. *Methods* 72:76–85
100. Silverman IM, Li F, Alexander A, Goff L, Trapnell C, et al. 2014. RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. *Genome Biol.* 15:R3
101. Sloma MF, Mathews DH. 2015. Improving RNA secondary structure prediction with structure mapping data. *Methods Enzymol.* 553:91–114
102. Smola MJ, Calabrese JM, Weeks KM. 2015. Detection of RNA–protein interactions in living cells with SHAPE. *Biochemistry* 54:6867–75
103. Smola MJ, Rice GM, Busan S, Siegfried NA, Weeks KM. 2015. Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* 10:1643–69
104. Solem AC, Halvorsen M, Ramos SB, Laederach A. 2015. The potential of the riboSNitch in personalized medicine. *Wiley Interdiscip. Rev. RNA* 6:517–32
105. Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. 2013. RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* 9:18–20
106. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, et al. 2015. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* 519:486–90
107. Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, et al. 2012. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* 40:5023–33
108. Talkish J, May G, Lin Y, Woolford JL Jr., McManus CJ. 2014. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* 20:713–20
109. Tang Y, Assmann SM, Bevilacqua PC. 2016. Protein structure is related to RNA structural reactivity in vivo. *J. Mol. Biol.* 428:758–66
110. Tang Y, Bouvier E, Kwok CK, Ding Y, Nekrutenko A, et al. 2015. StructureFold: genome-wide RNA secondary structure mapping and reconstruction *in vivo*. *Bioinformatics* 31:2668–75
111. Truong DM, Sidote DJ, Russell R, Lambowitz AM. 2013. Enhanced group II intron retrohoming in magnesium-deficient *Escherichia coli* via selection of mutations in the ribozyme core. *PNAS* 110:E3800–9
112. Tserovski L, Marchand V, Hauenschild R, Blanloeil-Oillo F, Helm M, Motorin Y. 2016. High-throughput sequencing for 1-methyladenosine (m^1A) mapping in RNA. *Methods*. In press
113. Uhlenbeck OC. 1972. Complementary oligonucleotide binding to transfer RNA. *J. Mol. Biol.* 65:25–41
114. Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, et al. 2010. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* 7:995–1001
115. Vandivier L, Li F, Zheng Q, Willmann M, Chen Y, Gregory B. 2013. *Arabidopsis* mRNA secondary structure correlates with protein function and domains. *Plant Signal. Behav.* 8:e24301
116. Vandivier LE, Li F, Gregory BD. 2015. High-throughput nuclease-mediated probing of RNA secondary structure in plant transcriptomes. *Methods Mol. Biol.* 1284:41–70
117. Wachter A, Tunc-Ozdemir M, Grove BC, Green PJ, Shintani DK, Breaker RR. 2007. Riboswitch control of gene expression in plants by splicing and alternative 3′ end processing of mRNAs. *Plant Cell* 19:3437–50
118. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. 2011. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* 12:641–55
119. Wan Y, Qu K, Ouyang ZQ, Kertesz M, Li J, et al. 2012. Genome-wide measurement of RNA folding energies. *Mol. Cell* 48:169–81
120. Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, et al. 2014. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505:706–9
121. Wang X, Lu Z, Gomez A, Hon GC, Yue Y, et al. 2014. N^6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505:117–20
122. Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, et al. 2015. N^6-methyladenosine modulates messenger RNA translation efficiency. *Cell* 161:1388–99
123. Wang Y, Li Y, Toth JI, Petroski MD, Zhang Z, Zhao JC. 2014. N^6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* 16:191–98

124. Warf MB, Berglund JA. 2010. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* 35:169–78
125. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr., et al. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460:711–16
126. Wen JD, Lancaster L, Hodges C, Zeri AC, Yoshimura SH, et al. 2008. Following translation by single ribosomes one codon at a time. *Nature* 452:598–603
127. Wickiser JK, Winkler WC, Breaker RR, Crothers DM. 2005. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell* 18:49–60
128. Xia TB, SantaLucia J, Burkard ME, Kierzek R, Schroeder SJ, et al. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37:14719–35
129. Zaug AJ, Cech TR. 1995. Analysis of the structure of *Tetrahymena* nuclear RNAs in vivo: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. *RNA* 1:363–74
130. Zheng Q, Ryvkin P, Li F, Dragomir I, Valladares O, et al. 2010. Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in *Arabidopsis*. *PLOS Genet.* 6:e1001141
131. Zhou J, Wan J, Gao X, Zhang X, Jaffrey SR, Qian SB. 2015. Dynamic m^6A mRNA methylation directs translational control of heat shock response. *Nature* 526:591–94

RELATED RESOURCES

Computational resources for RNA folding:

- MFold/UNAFold: **http://unafold.rna.albany.edu/**
- RNAstructure: **http://rna.urmc.rochester.edu/RNAstructure.html**
- StructureFold in Galaxy: **https://usegalaxy.org/** (select "NGS: RNA Structure")
- ViennaRNA Package: **https://www.tbi.univie.ac.at/RNA/**

RNA modification database: **http://mods.rna.albany.edu/mods/modifications/search**