



Getting to the bottom of lncRNA mechanism: structure–function relationships

Karissa Sanbonmatsu¹

Received: 2 July 2021 / Accepted: 28 September 2021 / Published online: 12 October 2021

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021

Abstract

While long non-coding RNAs are known to play key roles in disease and development, relatively few structural studies have been performed for this important class of RNAs. Here, we review functional studies of long non-coding RNAs and expose the need for high-resolution 3-D structural studies, discussing the roles of long non-coding RNAs in the cell and how structure–function relationships might be used to elucidate further understanding. We then describe structural studies of other classes of RNAs using chemical probing, nuclear magnetic resonance, small-angle X-ray scattering, X-ray crystallography, and cryogenic electron microscopy (cryo-EM). Next, we review early structural studies of long non-coding RNAs to date and describe the way forward for the structural biology of long non-coding RNAs in terms of cryo-EM.

Introduction

Long non-coding RNAs (lncRNAs) are typically defined as RNAs longer than 200 nucleotides in length without significant coding potential, often playing regulatory roles in mammalian systems (Winkle et al. 2021). Because this class of RNA molecules has been found to be important for processes in cancer, development, and brain function, there is keen interest in the pharmaceutical community (Kashi et al. 1859; Hon et al. 2017). However, the enormous size of these RNAs, which are often kilobases or tens of kilobases in length, makes the prospects of drugging them daunting. If the drug is a small molecule, then which of the 10,000 bases on a 10 kb lncRNA should be targeted? If the drug is an antisense oligo, which region of the RNA should be targeted? Which regions should not be targeted?

The steps of pre-clinical trials, clinical trials, and regulatory approval have been in the news lately regarding COVID-19 vaccine development. Similar steps are required for protein-based drugs, such as anti-viral therapeutics, cancer drugs, anti-depressants, antibiotics, and disease-related therapies (Matthews et al. 2016). However, before these steps can begin, target identification, lead generation, lead optimization, and drug candidate selection must take

place. Each of these stages requires considerable structural characterization. In the case of protein-based drugs, often a high-resolution 3-D structure of the target protein is solved by either X-ray crystallography or cryo-EM, followed by binding pocket characterization, hit identification, lead development, and lead optimization (Grey and Thompson 2010). Currently, there are no high-resolution 3-D structures of lncRNAs.

In addition to drug development, structural biology has been quite useful for understanding protein mechanism. Since a protein's mechanism and function are often determined by the other molecules that the protein interacts with, the 3-D structure of the protein can directly reveal its mechanism, as the structure provides the details of how the protein fits with its interaction partner, i.e., the details of how the protein works. Many actually define protein mechanism as the relationship between its structure and function. This is exemplified by the fact that, if, hypothetically, the positions of amino acids in a protein were to change drastically, then the function would likely also change drastically, if not ruined entirely.

Relative to the history of structural biology in mechanistic studies and drug development in the protein community, and the fact that high-resolution structures of lncRNAs have not yet been solved, lncRNA mechanism is not well understood at the molecular level of detail. Without a clear understanding of structure, structure–function relationships, and mechanism, lncRNA drug discovery is in its early stages. In the case of lncRNAs, we expect

✉ Karissa Sanbonmatsu
kys@lanl.gov

¹ Los Alamos National Laboratory, Los Alamos, NM 87545, USA

understanding mechanism will require determination of the structure–function relationship for the RNA, and determination of the structure–function relationship will require solving the lncRNA 3-D structure at high resolution, similar to how structure–function relationships and mechanisms were worked out for proteins. Thus, we anticipate that solving structures of lncRNAs will be an important stage for determination of lncRNA mechanism and for lncRNA drug discovery.

Although long non-coding RNAs have been shown to be important in development, epigenetics, stem cell biology, plant biology, RNA processing, hormone response, cancer, and brain function (Rinn and Chang 2012; Klattenhoff et al. 2013; Mercer and Mattick 2013; Swiezewski et al. 2009; Ulitsky and Bartel 2013; Gong and Maquat 2011; Kaneko et al. 2014; Heard et al. 1999; Rocha et al. 2014; Boumil and Lee 2001; Davidovich et al. 2013, 2015; Cech and Steitz 2014; Brown et al. 2014; Dharap et al. 2012; Ponting et al. 2009; Derrien et al. 2012), many researchers have avoided 3-D structural studies either because (i) they believe the RNAs are unstructured, (ii) they believe structural studies of lncRNAs are too difficult, or (iii) they are unaware of the success of structural biology techniques in other fields of RNA biology, such as RNAi, Crispr-Cas9, protein synthesis, splicing, and bacterial metabolism (Doherty and Doudna 2000; Wilson and Doudna 2013; Pyle 2016; Voorhees and Ramakrishnan 2013; Montange and Batey 2008; Frank and Gonzalez 2010; Hashem and Frank 2018). Because of the reluctance to study lncRNA structures, structure–function relations for these RNAs have lagged behind other sub-fields in RNA biology. The hesitancy, however, is not necessarily justified. Reason (i) is not necessarily true, in light of the physical properties of RNA: since the bases and backbone of RNA are polar, Watson–Crick and non-Watson–Crick base pairs form for almost any RNA sequence. This propensity to form base pairs combined with well-known non-specific backbone-to-base backbone-to-backbone (often ion-mediated) interactions results in the tendency of RNA to ‘stick to itself’ and form intricate secondary and tertiary structures. Reason (ii) has merit; however, breakthroughs in cryo-EM have proved the feasibility of 3-D studies of purely RNA systems, producing cryo-EM structures of riboswitch RNAs, frame shifting pseudoknot elements of mRNAs, and tRNA-like structures (Zhang et al. 2019, 2020; Kappel et al. 2020; Sherlock et al. 2021). As for reason (iii), x-ray crystallography, small-angle x-ray scattering, nuclear magnetic resonance imaging, and cryo-EM have enjoyed enormous success in determining 3-D structures of other RNA systems (Pyle 2016; Montange and Batey 2008; Zhang et al. 2019, 2020; Kappel et al. 2020; Sherlock et al. 2021; Liu et al. 2021; Roy et al. 2017a; Torabi et al. 2021; Pollack 2011). Here, we describe the lncRNA functional studies, review high-resolution structure–function relationships in other

RNA systems, and discuss early results and the prospects for higher-resolution structure–function studies in lncRNAs.

Long non-coding RNAs (lncRNAs)

Long non-coding RNAs (lncRNAs) are often found in mammalian epigenetic systems, exceed 200 nucleotides in length, polyadenylated, alternatively spliced, low in abundance, and display relatively low sequence conservation. A subset of the non-coding RNAs (K. Numata et al. 2003; P. Carninci et al. 2005), long non-coding RNAs have been shown to have specificity to tissue type and developmental stage (Ponjavic et al. 2007; Dinger et al. 2008; Rinn and Chang 2012). Many genome-wide studies have been performed to identify large classes of lncRNAs associated with environmental changes, tissues, and diseases (Rinn and Chang 2012). Loss-of-function studies have been performed to characterize functional roles of lncRNAs (Charles Richard and Eichhorn 2018). Biochemical and low-resolution methods have been used to obtain structural information yielding glimpses of lncRNA structure (Novikova et al. 2013a). High-resolution structural biology techniques have been instrumental in determining structure–function relationships in other classes of RNA (riboswitches, ribozymes, and ribosomes) (Westhof 2015; Reyes et al. 2009). These structure–function relationships enable more precise understanding of mechanism in terms of structural dynamics, thermodynamics, kinetics, and Mg^{2+} effects. Yet, few studies have examined lncRNA mechanism at the atomistic level of detail (Novikova et al. 2013a).

Structure–function relationships

Structure–function relationships have been critical in understanding biological systems in molecular detail. Since the inception of structural biology, 3-D structures of proteins have led to breakthroughs in understanding protein binding, protein complex formation, ligand binding, and self-assembly, all of which are important throughout biology and biomedicine. In biological systems, we often first know that a molecule is important, and even what it does, but not how it does it. The ‘how,’ in the case of a protein, is then worked out by solving the protein’s 3-D structure and relating it to its function. Once we know the ‘how,’ we can begin to understand the molecule in context and start thinking about drugging the molecule. In the case of protein molecules, their function almost always hinges on interaction with another molecule, such as another protein, RNA, or DNA molecule. Solving the structure of the protein in isolation and complexed with its target molecules produces invaluable information about its function and about the structure–function relationship. A wide variety of techniques have been developed to gain

information about the 3-D structures of proteins and protein complexes, including X-ray crystallography, X-ray free-electron laser crystallography, cryo-EM, NMR, and small-angle scattering (Adams et al. 2013; Sekhar and Kay 2019; Glaeser 2019; Rambo and Tainer 2013; Smith et al. 2018; Gruner and Lattman 2015). X-ray crystallography has been a leading technique for many decades. For example, the molecular basis of the biological functions of the lysozyme, ATP synthase, and ion channels was provided by their X-ray crystal structures (Blake et al. 1965; Doyle et al. 1998; Boyer 1997). In addition to producing the mechanism of a molecule, structural studies (Hunter 1997) have also led to new drugs, as in the case of Plexxikon (scaffold-based drug discovery) and Zelboraf (metastatic melanoma) (Gul and Zimmermann 2017). More recently, cryo-EM has taken center stage in protein structural biology. For example, cryo-EM structures of the COVID-19 spike protein in various states were used to optimize stable spike constructs for mRNA-based vaccines (Ma et al. 2021). In the case of nucleic acids, the DNA double helix structure immediately led to an understanding of the role of DNA in the cell as the carry of reproducible information (Watson and Crick 1953). More recently, cryo-EM structures of nucleosome complexes have produced new insights into chromatin organization and gene regulation (Han et al. 2020; Takizawa et al. 2020). On the whole, high-resolution 3-D structures have been instrumental in determining mechanism, discovering drugs, and identifying function in a large number of biomolecular systems.

Structural studies of RNA systems

As far fewer RNA systems have been studied relative to protein systems, RNA structural biology has lagged behind protein structural biology considerably. However, as described below, high-resolution structures have been obtained for several classes of RNAs, leading to important insights into their structure–function relationships.

Self-splicing introns

Some of the earliest RNA-only systems solved to high resolution are the group I and group II introns (Pyle 2016). Using X-ray crystallography, these structures revealed the overall 3-D architecture of the RNA, detailed local RNA–RNA interaction motifs connecting the RNA together, the role of Mg^{2+} ions in the structure, and how the 2-D secondary structure maps translate into 3-D structures. Importantly, the 3-D structures were critical in determining the mechanism of catalysis for splicing, answering questions that were difficult or impossible to solve using other methods.

Riboswitch RNAs

Riboswitch RNAs are regulatory stretches of RNA commonly residing in the 5'-UTR of mRNA in bacterial metabolism-related genes (Montange and Batey 2008; Breaker 2011). These RNAs control gene expression by detecting environmental molecules through ligand-binding 3-D folds that alter the regulatory behavior of the RNA. In a riboswitch, one sequence has two competing secondary structures (and two competing tertiary structures). The presence of ligand shifts the equilibrium to one structure, altering the gene expression ON/OFF state. The majority of riboswitches were discovered with cell-free, *in vitro* chemical probing studies revealing the ligand dependence of the secondary structure, supported by *in vivo* functional studies. These *in vitro* secondary structures were later validated by *in vitro* high-resolution X-ray crystallographic 3-D structures (Serganov and Patel 2012). The dynamics of these systems have been studied using small-angle X-ray scattering (SAXS) experiments and molecular dynamics simulations (Zhang et al. 1839). SAXS and biochemical studies have also revealed that ligand-free conformations tend to be extended and flexible, whereas ligand-bound conformations tend to be compact and ordered. Most recently, molecular dynamics simulations have been used to integrate crystallographic, biochemical, and SAXS data, elucidating the operational principles of riboswitches and their dependence on magnesium (Roy et al. 2017a, 2019, 2017b; Hayes et al. 2014, 2015; Hennelly et al. 2013).

Ribonucleoprotein complexes

Structural studies of several ribonucleoprotein complexes have been studied, including the ribosome, RNA processing complexes, and the spliceosome. The ribosome is perhaps the most extensively studied ribonucleoprotein complex (Jobe et al. 2019). Structural studies have been attempted since the 1980s, commencing with biochemical studies to determine the secondary structure of the small subunit ribosome RNA (16S) and large subunit ribosomal RNA (23S) (Rummel and Noller 1973; Woese et al. 1980; Noller et al. 1981; Noller and Woese 1981). Neutron scattering enabled the rough placement of proteins in 3-D space relative to the ribosome complex (Engelman and Moore 1976; Moore et al. 1975). Early cryo-EM studies yielded the morphologies of the two subunits, the tRNA and mRNA ligands, the ribosomal proteins, and various conformations of the ribosome (Frank and Gonzalez 2010). Details were filled in with X-ray crystallography structures (Voorhees and Ramakrishnan 2013). High-resolution cryo-EM enabled studies of ribosomes in a wide variety of functional states, for a variety of different species (Hashem and Frank 2018). With structures in hand, structural dynamics studies have been performed,

integrating cryo-EM, single-molecule FRET, and large-scale molecular dynamics simulations, providing a comprehensive picture of the molecular mechanism of the ribosome, characterizing the energy landscape and transition rates in the context of the detailed structures of beginning, ending and a plethora of intermediate states for various stages of protein synthesis (Sanbonmatsu 2012, 2019, 2006; Morse et al. 2020; Sanbonmatsu et al. 2005; Tung and Sanbonmatsu 2004; Girodat et al. 2020; Wasserman et al. 2016; Ferguson et al. 2015; Munro et al. 2009).

RNA processing

3-D structures of macromolecular complexes that process RNA molecules have yielded important insights. Passmore and co-workers used X-ray crystallography to obtain high-resolution structures of *Saccharomyces cerevisiae* Pan2 in complex with RNA to show that Pan2 recognizes the stacked, helical conformation of poly(A) RNA (Kumar et al. 2019). This complex was reconstituted in a cell-free, in vitro system (Tang et al. 2019). They also used a combination of crystallography and electron microscopy to obtain structures of CPF/CPSF, a multi-protein complex essential for formation of mRNA 3' ends, showing that the process requires incorporation of the Ysh1 endonuclease into an eight subunit core complex (Hill et al. 2019).

Spliceosome

The high-resolution structures of a large number of full spliceosome complexes have been solved using cryo-EM over the past five years in a wide variety of splicing states. These structures were the culmination of decades of biochemical and genetic work, as well as lower-resolution cryo-EM structures of complexes along with high-resolution crystallography structures of smaller sub-regions of the complex (Yan et al. 2019). The spliceosome complex assembles on the pre-mRNA through a variety of protein and RNA interactions that work together to recognize specific splicing sites. This is followed by RNA-based catalyzation of cleavage and ligation, removing the intron stretches of RNA and reconnecting the remaining RNA to form the mRNA. Like the ribosome, the spliceosome has a rich history in mechanism and structural studies and, in terms of structural studies, is one of the most important ribonucleoprotein complexes (Fica and Nagai 2017; Fica 2020; Wilkinson et al. 2020; Smathers and Robart 1862). Unlike the ribosome, spliceosome operation is significantly more complex: factors are continuously coming on and off the complex during the myriad of sub-steps required for splicing. It has been hypothesized that in humans, the composition of the complex may be transcript-specific. Furthermore, in addition to undergoing changes

in tertiary structure, the secondary structure of the RNA also changes, requiring major rearrangements of the RNA. Although, from an RNA structure standpoint, spliceosome operation is more complex than ribosome operation, the spliceosome may present a more apt analog to a lncRNA molecular machine, since the complex is more dynamic, both in terms of the composition of the complex and in terms of the conformational changes required for the RNA (Wilkinson et al. 2020).

3-D structural techniques used to study other classes of RNAs

High-resolution techniques have been used to determine structures for a number of other classes of RNA systems, such as riboswitches, ribozymes, introns, ribosomes, and spliceosomes. In terms of techniques, nuclear magnetic resonance imaging (NMR) can be used to study small systems. This method has the advantage of capturing precise information about the dynamics of the RNA, multiple configurations, and rates of transition between configurations (Liu et al. 2021). NMR has been used to obtain such information for a variety of riboswitches and regions of viral RNAs, as well as a small region of Xist RepA lncRNA (Duszczuk et al. 2008). X-ray crystallography is a traditional form of high-resolution structure determination used for small- and medium-sized RNA systems. High-resolution structures have been determined for riboswitches, ribozymes, introns, and ribosomes. Cryogenic electron microscopy (cryo-EM) can be used to determine high-resolution structures for medium-sized and large-sized protein systems and ribonucleoprotein systems. To date, this method has determined a wide variety of structures for ribonucleoprotein complexes, including many ribosome complexes and several spliceosome complexes. Quite recently, the method has been used to obtain medium-resolution structures of several RNA-only systems, including riboswitches and regions of viral RNAs (Zhang et al. 2019, 2020; Kappel et al. 2020; Sherlock et al. 2021).

Studies of long non-coding RNAs

Loss-of-function studies have identified important lncRNAs, in terms of their functional roles in the cell, including epigenetic sensing and recruitment, sponging, P-bodies, scaffolding, RNA processing (lncRNA_{bnb1/2}), and hormone response (Gong and Maquat 2011). *Knockdown studies* also improve understanding. Knockdowns of Braveheart showed that this lncRNA is critical for lineage commitment in cardiomyocytes (Klattenhoff et al. 2013). *CRISPR/Cas9 knockout studies* have expanding the number of clear causal roles of lncRNAs. CRISPR/Cas9 knockout of an 11-nucleotide r-turn RNA motif showed that this structural motif is critical for

the overall function of Braveheart (Xue et al. 2016). Knockouts had a major reduction in embryoid body beating assays, along with dramatic decreases in normal development. Protein binding studies offer some insight into mechanism. In pulldowns and SAXS analysis, Braveheart was shown to bind zinc finger protein CNBP (Kim et al. 2020). Several genome-wide studies have been performed to identify proteins that bind to Xist (Minajigi et al. 2015a).

Mechanisms of lncRNAs

One of the earliest discovered lncRNAs is Xist (X chromosome inactivation-stimulated transcript), responsible for inactivation of the X chromosome during development (Lee and Jaenisch 1997). More recently, several lncRNAs have been associated with HOX gene systems during development (Rinn and Chang 2012). The 1/2sbs-lncRNA controls mRNA decay by hybridizing with mRNA to form a platform for STAU1 protein binding, triggering degradation of mRNA (Gong and Maquat 2011). Other lncRNAs are required for p21 activation (Huarte et al. 2010), stem cell reprogramming (Guttman et al. 2011), and stress response (Kino et al. 2010).

lncRNAs with phenotypes

Although the physiological relevance of many of the reported lncRNAs has not been determined, many lncRNAs have been shown to possess important, visible phenotypes (Li and Chang 2014). In addition to Xist, required for dosage compensation, the Braveheart lncRNA has been shown to be required for lineage commitment in cardiomyocytes (Klattenhoff et al. 2013). FENDRR lncRNA is required for heart, lung, and gastrointestinal development (Sauvageau et al. 2013). Linc-brn1b is required for neocortex development (Sauvageau et al. 2013). The COOLAIR lncRNA is required in *A. thaliana* for cold-timed flowering (Swiezevski et al. 2009). Additionally, the NEAT1 lncRNA has the clear phenotype of being critical for paraspeckle formation (Naganuma et al. 2012; Nakagawa and Hirose 2012; Sasaki et al. 2009).

lncRNA–protein interactions

Many studies have been performed to determine the protein partners of lncRNAs and elucidate the functions of these RNA–protein interactions (Davidovich et al. 2015; Minajigi et al. 2015a). Lee and co-workers developed an RNA centric proteomic method (iDRIP) to determine the Xist lncRNA interactome, showing cohesin repulsion and an RNA-directed chromosome conformations (Chu et al. 2021; Minajigi et al. 2015b). The group also identified lncRNAs associated with Polycomb repressive complex PRC2 using RIP-seq (Zhao et al. 2010). Carninci and co-workers developed a

new technology to map genome-wide RNA–chromatin interactions in intact nuclei (RNA And DNA Interacting Complexes Ligated and sequenced, RADICL-seq) (Bonetti et al. 2020). This proximity ligation-based methodology identifies patterns of genome occupancy for different classes of transcripts (Bonetti et al. 2020).

2-D Structural studies of lncRNAs: lncRNA secondary structure studies using chemical probing

Genome-wide studies of secondary structure have revealed that lncRNAs are more structured than mRNAs, but less structured than ribosomal RNAs (Wan et al. 2014, 2013, 2012; Ouyang et al. 2013; Kertesz et al. 2010; Ding et al. 2014; Rouskin et al. 2014). Detailed secondary structure studies of complete, intact lncRNA systems show that some lncRNAs are hierarchically structured with sub-domains containing modular RNA secondary structure motifs (Novikova et al. 2012; Ilik et al. 2013; Somarowthu et al. 2015). Studies of Malat-1 and related lncRNAs show that the 3'-end forms a triple helix, protecting it from RNase degradation (Brown et al. 2014; Wilusz et al. 2012, 2008). Other studies have elucidated lncRNA–protein interactions, emphasizing the need for detailed structural studies and mechanistic studies at the molecular and atomistic level (Chu et al. 2015; Spitale et al. 2015).

lncRNAs tend to have relatively low sequence identity and are often described as non-conserved. Some non-coding RNAs (miRNAs and rRNAs) have very high sequence identity (> 78% in nucleic acid sequence identity) (Griffiths-Jones et al. 2003). In contrast, many other important classes of non-coding RNAs have relatively low sequence identity (nucleic acid sequence identity of ~50%–65%), but secondary structures that are conserved across thousands of sequences. For example, riboswitches, which regulate metabolism in bacteria, typically have sequence identities of only 50%–65%, but have secondary structures conserved across thousands of species (Griffiths-Jones et al. 2003). The U2 and U4 spliceosomal RNAs have sequence identities < 60% but secondary structures conserved for > 9000 sequences. The 5S ribosomal RNA has sequence identity of ~60% but secondary structure conserved over 229,000 sequences. The group I intron has decidedly low sequence identity (~36%) but structure conserved across 60,000 species (Griffiths-Jones et al. 2003).

RNAs with low sequence identity are difficult to find using conventional search algorithms such as BLAST. However, knowledge of secondary structure dramatically enhances the search success. A wide variety of computational techniques to predict RNA secondary structure exists, using either free-energy estimates, multiple sequence alignment and direct coupling analysis, machine learning, or a combination of these (Yao et al. 2017; Dallaire and Major

2016; Parisien and Major 2008; Mathews 2019; Spasic et al. 2018; Tan et al. 2017; Eggenhofer et al. 2016; Lorenz et al. 2016a, 2016b; Pucci et al. 2020). These can be highly effective for a range of RNAs. With the growing number of possibilities for long-range interactions, pseudoknots, and multiway junctions, the number of potential RNA secondary structure folds exponentiates as a function of sequence length, making the task of predicting long non-coding RNA secondary structure formidable. In many RNA systems, *in vitro* chemical probing experiments have produced highly accurate secondary structures, subsequently verified by X-ray crystallography. In the case of riboswitches, RNA secondary structures were determined experimentally for a single species using *in vitro* chemical probing of the RNA in cell-free reconstituted systems (Regulski and Breaker 2008; Winkler et al. 2002, 2004; Mandal et al. 2003, 2004; Sudarshan et al. 2006, 2008; Cheah et al. 2007). Next, this structure was used as a fingerprint to find the structure in thousands of other species, despite the low sequence identity (Weinberg et al. 2007). These secondary structures determined from cell-free systems by chemical probing were verified by X-ray crystallography (Montange and Batey 2008, 2006; Batey et al. 2004; Gilbert et al. 2008; Stoddard et al. 2010).

To determine the RNA secondary structure of lncRNA molecules, strategies similar to those used to determine the original 16S rRNA secondary structure (Woese et al. 1980; Noller et al. 1981; Noller and Woese 1981) and the riboswitches (Winkler et al. 2003) have been employed. Chemical probing experiments determine nucleotides that are highly mobile and likely to reside in looping regions, as well as those nucleotides with low mobility, likely to participate in Watson–Crick base pairs. To cope with the large RNA size, 3S (Shot-Gun Secondary Structure) can be used, which probes the entire RNA first and then probes shorter segments of the RNA in successive rounds of probing (Novikova et al. 2012, 2013b). By matching signals of short segments with full RNA experiments, modular sub-domains are identified, for which a secondary structure is often readily discernable. The resulting secondary structure can be used to improve existing phylogenetic sequence alignments and, in principle, can be used to find instances of the lncRNA not previously found in other species (Hawkes et al. 2016).

An interesting case is the 873 nt steroid receptor RNA activator lncRNA in humans (SRA-1). This lncRNA co-activates the hormone response in human T-47D cells and co-immunoprecipitates with a large number of important proteins, including several hormone receptors (estrogen receptor, progesterone receptor, androgen receptor, glucocorticoid receptor, and thyroid receptor) (Yao et al. 2010; Xu et al. 2009; Colley et al. 2008; Huet et al. 2014). Binding assays in *in vitro* cell-free reconstituted systems have shown strong binding to the pseudouridylase Pus1p, estrogen receptor, thyroid receptor, the sex reversal factor DAX-1, and

the epigenetic factor SHARP. While the primary function of SRA-1 is to co-activate the hormone response, a speculated secondary function involving the binding of SRA-1 to its cognate protein SRAP has recently been shown not to occur (SRA-1 does not bind to SRAP) (McKay et al. 2014).

A previous study demonstrated that SRA-1 contains four modular secondary structure sub-domains, each containing multiple secondary structure motifs. The secondary structure was consistent with four different probing techniques (SHAPE, DMS, *in-line*, and RNase V1). Binding studies have shown that SHARP binds to the helix 12/helix 13 (H12/13) domain (Arieti et al. 2014).

Because the probing signal *in vivo* may be obfuscated by multiple proteins binding to the RNA (Davidovich et al. 2013, 2015), *in vitro* studies establish an important *ab initio* structure. There are few known cases of high-resolution 3-D structures, where an *in vitro* structure of an intact, individual RNA has been shown to differ from its corresponding *in vivo* structure. For example, the vast majority of crystallographic structures of RNAs, which are determined *in vitro*, have either (i) been validated *in vivo* or (ii) not been disproven *in vivo*. In the case of riboswitch RNAs, crystallographic data strongly support initial secondary structures determined by chemical probing techniques discussed above.

On the whole, determination of the precise and detailed secondary structure of lncRNAs allows classification into (i) highly structured RNAs with sub-domains and complex structural motifs, such as multiway junctions; (ii) loosely structured RNAs with multiple stem-loops, but lacking hierarchical domain structure and complex motifs; and (iii) unstructured, disordered RNAs, which lack secondary structure.

3-D studies of long non-coding RNAs at low resolution

Studies of tertiary interactions in long non-coding RNAs. Pyle and co-workers used UV crosslinking to identify individual tertiary interactions in lncRNA systems (Liu et al. 2017).

Small-angle X-ray scattering (SAXS). Small-angle X-ray scattering studies have been used to characterize the 3-D structure of RNA systems that are too flexible to be studied with X-ray crystallography. Often, RNA molecules sample a multitude of conformations. SAXS can characterize the distribution of configurations samples. In addition, SAXS can be a first step toward higher-resolution structure determination as the requirements for sample preparation are much less stringent than for X-ray crystallography or for higher-resolution cryo-EM. Recently, low-resolution structures of the Braveheart lncRNA and Braveheart-CNBP ribonucleoprotein complex were determined using SAXS (Kim et al. 2020). The structures were consistent with 2-D secondary

structures determined via chemical probing, with secondary structure domains fairly well-separated in 3-D physical space. The molecule was found to be somewhat flexible, where multiple all-atom 3-D configurations were consistent with 3-D volume reconstructions consistent with the SAXS data. However, the SAXS data demonstrated compaction upon Mg^{2+} titration, which is clear evidence of well-defined tertiary structures in the RNA system. This is similar to riboswitch systems, which still sample well-defined 3-D structures, even in their ligand-free states, known to be extended and flexible. Additionally, Braveheart underwent significant reorganization upon protein binding, as evidenced by the substantial change in scattering profiles and corresponding 3-D volume reconstructions as a result of CNBP binding.

Atomic force microscopy (AFM) studies of lncRNAs. AFM has been used to characterize the 3-D structure of lncRNA systems without solution. In these experiments, MEG3 displayed tertiary structure consistent with 2-D secondary structures determined by chemical probing (Uroda et al. 2019). Bachelet and co-workers used fast AFM scanning to quantification of the motion of HOTAIR lncRNA, describing the anatomy and intrinsic properties of HOTAIR (Spokoini-Stern et al. 2020).

Fluorescence correlation spectroscopy (FCS). FCS has been used to characterize the size, in terms of extended vs. compact, of lncRNAs systems in 3-D. In one FCS study, lncRNAs (e.g., HOTAIR) were found to be more compact than mRNA transcripts, but less compact than ribosomes (Borodavka et al. 2016).

Expansion of structural tools to study long noncoding RNAs at high resolution

High-resolution structural studies of lncRNA systems will undoubtedly reveal new information about their mechanisms. As early studies present evidence for tertiary contacts, at minimum, cryo-EM studies of lncRNAs may reveal structured tertiary motifs surrounded by flexible regions or large swaths of RNA. At the other extreme, these studies may uncover highly structured ribonucleoprotein complexes, or even structured RNA-only systems. The past decade of lncRNA research has clearly shown that lncRNAs represent a highly diverse class of RNAs with a wide range of functional roles. Thus, a wide range of structural content may be observed, ranging from highly dynamic to highly structured. Higher-resolution structural studies will be able to shed light on structure–function relationships, in terms of specific protein binding partners, RNA binding partners, DNA binding partners, conformational changes, and roles in pathways. These studies may also offer insight into the evolution of lncRNAs. Since lncRNAs often have fairly low sequence identity, structure–function studies will enable analysis of

conservation in terms of more general measures, such as 2-D structure, 3-D structural RNA motifs, 3-D RNA–protein binding motifs, RNA dynamics, and RNA function (Hezroni et al. 2015; Ulitsky 2016).

Acknowledgements The author acknowledges generous support by the LANL LDRD program (20210082DR).

Declarations

Conflict of interests On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Adams PD, Baker D, Brunger AT, Das R, DiMaio F, Read RJ, Richardson DC, Richardson JS, Terwilliger TC (2013) Advances, interactions, and future developments in the CNS, Phenix, and Rosetta structural biology software systems. *Annu Rev Biophys* 42:265–287
- Arieti F, Gabus C, Tambalo M, Huet T, Round A, Thore S (2014) The crystal structure of the Split End protein SHARP adds a new layer of complexity to proteins containing RNA recognition motifs. *Nucleic Acids Res* 42:6742–6752
- Batey RT, Gilbert SD, Montange RK (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* 432:411–415
- Blake CC, Koenig DF, Mair GA, North AC, Phillips DC, Sarma VR (1965) Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Ångstrom resolution. *Nature* 206:757–761
- Bonetti A, Agostini F, Suzuki AM, Hashimoto K, Pascarella G, Gimenez J, Roos L, Nash AJ, Ghilotti M, Cameron CJF, Valentine M, Medvedeva YA, Noguchi S, Agirre E, Kashi K, Samudiyata J, Luginbuhl R, Cazzoli S, Agrawal NM, Luscombe M, Blanchette T, Kasukawa M, Hoon E, Arner B, Lenhard C, Plessy G, Castelo-Branco V, Orlando PC (2020) RADICL-seq identifies general and cell type-specific principles of genome-wide RNA-chromatin interactions. *Nat Commun* 11:e1018
- Borodavka A, Singaram SW, Stockley PG, Gelbart WM, Ben-Shaul A, Tuma R (2016) Sizes of long RNA molecules are determined by the branching patterns of their secondary structures. *Biophys J* 111:2077–2085
- Boumil RM, Lee JT (2001) Forty years of decoding the silence in X-chromosome inactivation. *Hum Mol Genet* 10:2225–2232
- Boyer PD (1997) The ATP synthase—a splendid molecular machine. *Annu Rev Biochem* 66:717–749
- Breaker RR (2011) Prospects for riboswitch discovery and analysis. *Mol Cell* 43:867–879
- Brown JA, Bulkley D, Wang J, Valenstein ML, Yario TA, Steitz TA, Steitz JA (2014) Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat Struct Mol Biol* 21:633–640
- Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest AR, Zavolan M, Davis MJ, Wilming LG, Aidinis V, Allen JE, Ambesi-Impiomato A, Apweiler R, Aturaliya RN, Bailey TL, Bansal M, Baxter L, Beisel KW, Bersano T, Bono H, Chalk AM, Chiu KP, Choudhary V, Christoffels A, Clutterbuck DR, Crowe ML, Dalla E, Dalrymple BP, de Bono B, Della Gatta G, di Bernardo D, Down T, Engstrom P, Fagiolini M, Faulkner G, Fletcher CF, Fukushima T,

- Furuno M, Futaki S, Gariboldi M, Georgii-Hemming P, Gingeras TR, Gojobori T, Green RE, Gustincich S, Harbers M, Hayashi Y, Hensch TK, Hirokawa N, Hill D, Huminiecki L, Iacono M, Ikeo K, Iwama A, Ishikawa T, Jakt M, Kanapin A, Katoh M, Kawasawa Y, Kelso J, Kitamura H, Kitano H, Kollias G, Krishnan SP, Kruger A, Kummerfeld SK, Kurochkin IV, Lareau LF, Lazarevic D, Lipovich L, Liu J, Liuni S, McWilliam S, Madan Babu M, Madera M, Marchionni L, Matsuda H, Matsuzawa S, Miki H, Mignone F, Miyake S, Morris K, Mottagui-Tabar S, Mulder N, Nakano N, Nakauchi H, Ng P, Nilsson R, Nishiguchi S, Nishikawa S, Nori F, Ohara O, Okazaki Y, Orlando V, Pang KC, Pavan WJ, Pavesi G, Pesole G, Petrovsky N, Piazza S, Reed J, Reid JF, Ring BZ, Ringwald M, Rost B, Ruan Y, Salzberg SL, Sandelin A, Schneider C, Schonbach C, Sekiguchi K, Sempke CA, Seno S, Sessa L, Sheng Y, Shibata Y, Shimada H, Shimada K, Silva D, Sinclair B, Sperling S, Stupka E, Sugiura K, Sultana R, Takenaka Y, Taki K, Tammouja K, Tan SL, Tang S, Taylor MS, Tegner J, Teichmann SA, Ueda HR, van Nimwegen E, Verardo R, Wei CL, Yagi K, Yamanishi H, Zabarovsky E, Zhu S, Zimmer A, Hide W, Bult C, Grimmond SM, Teasdale RD, Liu ET, Brusuc V, Quackenbush J, Wahlestedt C, Mattick JS, Hume DA, Kai C, Sasaki D, Tomaru Y, Fukuda S, Kanamori-Katayama M, Suzuki M, Aoki J, Arakawa T, Iida J, Imamura K, Itoh M, Kato T, Kawaji H, Kawagashira N, Kawashima T, Kojima M, Kondo S, Konno H, Nakano K, Ninomiya N, Nishio T, Okada M, Plessy C, Shibata K, Shiraki T, Suzuki S, Tagami M, Waki K, Watahiki A, Okamura-Oho Y, Suzuki H, Kawai J, Hayashizaki Y, F Consortium RGER Group G Genome Science (2005) The transcriptional landscape of the mammalian genome. *Science* 309:1559–1563
- Cech TR, Steitz JA (2014) The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* 157:77–94
- Charles Richard JL, Eichhorn PJA (2018) Platforms for Investigating LncRNA Functions. *SLAS Technol* 23:493–506
- Cheah MT, Wachter A, Sudarsan N, Breaker RR (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 447:497–500
- Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY (2015) Systematic discovery of Xist RNA binding proteins. *Cell* 161:404–416
- H.P. Chu, A. Minajigi, Y. Chen, R. Morris, C.Y. Guh, Y.H. Hsieh, M. Boukhali, W. Haas, J.T. Lee 2021 iDRiP for the systematic discovery of proteins bound directly to noncoding RNA. *Nat Protoc*
- Colley SM, Iyer KR, Leedman PJ (2008) The RNA coregulator SRA, its binding proteins and nuclear receptor signaling activity. *IUBMB Life* 60:159–164
- da Rocha ST, Boeva V, Escamilla-Del-Arenal M, Ancelin K, Granier C, Matias NR, Sanulli S, Chow J, Schulz E, Picard C, Kaneko S, Helin K, Reinberg D, Stewart AF, Wutz A, Margueron R, Heard E (2014) Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol Cell* 53:301–316
- Dallaire P, Major F (2016) Exploring alternative RNA structure sets using MC-flashfold and db2cm. *Methods Mol Biol* 1490:237–251
- Davidovich C, Zheng L, Goodrich KJ, Cech TR (2013) Promiscuous RNA binding by polycomb repressive complex 2. *Nat Struct Mol Biol* 20:1250–1257
- Davidovich C, Wang X, Cifuentes-Rojas C, Goodrich KJ, Gooding AR, Lee JT, Cech TR (2015) Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol Cell* 57:552–558
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhatar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigo R (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22:1775–1789
- Dharap A, Nakka VP, Vemuganti R (2012) Effect of focal ischemia on long noncoding RNAs. *Stroke* 43:2800–2802
- 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
- Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Solda G, Simons C, Sunkin SM, Crowe ML, Grimmond SM, Perkins AC, Mattick JS (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res* 18:1433–1445
- Doherty EA, Doudna JA (2000) Ribozyme structures and mechanisms. *Annu Rev Biochem* 69:597–615
- Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, Chait BT, MacKinnon R (1998) The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* 280:69–77
- Duszczyk MM, Zanier K, Sattler M (2008) A NMR strategy to unambiguously distinguish nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat. *Nucleic Acids Res* 36:7068–7077
- Eggenhofer F, Hofacker IL, C. Honer Zu Siederdisen, (2016) RNALien - Unsupervised RNA family model construction. *Nucleic Acids Res* 44:8433–8441
- Engelman DM, Moore PB (1976) Neutron-scattering studies of the ribosome. *Sci Am* 235:44–54
- Ferguson A, Wang L, Altman RB, Terry DS, Juette MF, Burnett BJ, Alejo JL, Dass RA, Parks MM, Vincent CT, Blanchard SC (2015) Functional dynamics within the human ribosome regulate the rate of active protein synthesis. *Mol Cell* 60:475–486
- Fica SM (2020) Cryo-EM snapshots of the human spliceosome reveal structural adaptations for splicing regulation. *Curr Opin Struct Biol* 65:139–148
- Fica SM, Nagai K (2017) Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat Struct Mol Biol* 24:791–799
- Frank J, Gonzalez RL Jr (2010) Structure and dynamics of a processive Brownian motor: the translating ribosome. *Annu Rev Biochem* 79:381–412
- Gilbert SD, Rambo RP, Van Tyne D, Batey RT (2008) Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat Struct Mol Biol* 15:177–182
- Girodat D, Blanchard SC, Wieden HJ, Sanbonmatsu KY (2020) Elongation factor Tu Switch I element is a gate for aminoacyl-tRNA selection. *J Mol Biol* 432:3064–3077
- Glaeser RM (2019) How good can single-particle cryo-EM become? what remains before it approaches its physical limits? *Annu Rev Biophys* 48:45–61
- Gong C, Maquat LE (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* 470:284–288
- Grey JL, Thompson DH (2010) Challenges and opportunities for new protein crystallization strategies in structure-based drug design. *Expert Opin Drug Discov* 5:1039–1045
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR (2003) Rfam: an RNA family database. *Nucleic Acids Res* 31:439–441
- Gruner SM, Lattman EE (2015) Biostructural Science Inspired by Next-Generation X-Ray Sources. *Annu Rev Biophys* 44:33–51
- S. Gul, S. Zimmermann 2017 Structure based drug discovery facilitated by crystallography. *Drug Target Review*
- Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477:295–300

- Han Y, Reyes AA, Malik S, He Y (2020) Cryo-EM structure of SWI/SNF complex bound to a nucleosome. *Nature* 579:452–455
- Hashem Y, Frank J (2018) The jigsaw puzzle of mRNA translation initiation in eukaryotes: a decade of structures unraveling the mechanics of the process. *Annu Rev Biophys* 47:125–151
- Hawkes EJ, Hennelly SP, Novikova IV, Irwin JA, Dean C, Sanbonmatsu KY (2016) COOLAIR antisense RNAs form evolutionarily conserved elaborate secondary structures. *Cell Rep* 16:3087–3096
- Hayes RL, Noel JK, Whitford PC, Mohanty U, Sanbonmatsu KY, Onuchic JN (2014) Reduced model captures Mg(2+)-RNA interaction free energy of riboswitches. *Biophys J* 106:1508–1519
- Hayes RL, Noel JK, Mandic A, Whitford PC, Sanbonmatsu KY, Mohanty U, Onuchic JN (2015) Generalized manning condensation model captures the RNA ion atmosphere. *Phys Rev Lett* 114:e58105
- Heard E, Mongelard F, Arnaud D, Chureau C, Vourc'h C, Avner P (1999) Human XIST yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* 96:6841–6846
- Hennelly SP, Novikova IV, Sanbonmatsu KY (2013) The expression platform and the aptamer: cooperativity between Mg²⁺ and ligand in the SAM-I riboswitch. *Nucleic Acids Res* 41:1922–1935
- Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I (2015) Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep* 11:1110–1122
- Hill CH, Boreikaite V, Kumar A, Casanal A, Kubik P, Degliesposti G, Maslen S, Mariani A, von Loeffelholz O, Girbig M, Skehel M, Passmore LA (2019) Activation of the endonuclease that defines mRNA 3' ends requires incorporation into an 8-subunit core cleavage and polyadenylation factor complex. *Mol Cell* 73:1217–1231
- Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJ, Gough J, Denisenko E, Schmeier S, Poulsen TM, Severin J, Lizio M, Kawaji H, Kasukawa T, Itoh M, Burroughs AM, Noma S, Djebali S, Alam T, Medvedeva YA, Testa AC, Lipovich L, Yip CW, Abugessaisa I, Mendez M, Hasegawa A, Tang D, Lassmann T, Heutink P, Babina M, Wells CA, Kojima S, Nakamura Y, Suzuki H, Daub CO, de Hoon MJ, Arner E, Hayashizaki Y, Carninci P, Forrest AR (2017) An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* 543:199–204
- Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL (2010) A large intergenic non-coding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142:409–419
- Huet T, Miannay FA, Patton JR, Thore S (2014) Steroid receptor RNA activator (SRA) modification by the human pseudouridine synthase 1 (hPus1p): RNA binding, activity, and atomic model. *PLoS ONE* 9:e94610
- Hunter WN (1997) A structure-based approach to drug discovery; crystallography and implications for the development of antiparasite drugs. *Parasitology* 114(Suppl):S17–29
- Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, Toscano S, Wan Y, Spitale RC, Luscombe N, Backofen R, Chang HY, Akhtar A (2013) Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol Cell* 51:156–173
- Jobe A, Liu Z, Gutierrez-Vargas C, Frank J (2019) New insights into ribosome structure and function. *Cold Spring Harb Perspect Biol* 11(1):e032615
- Kaneko S, Bonasio R, Saldana-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, Reinberg D (2014) Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell* 53:290–300
- Kappel K, Zhang K, Su Z, Watkins AM, Kladwang W, Li S, Pintilie G, Topkar VV, Rangan R, Zheludev IN, Yesselman JD, Chiu W, Das R (2020) Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat Methods* 17:699–707
- Kashi K, Henderson L, Bonetti A, Carninci P (1859) Discovery and functional analysis of lncRNAs: methodologies to investigate an uncharacterized transcriptome. *Biochim Biophys Acta* 2016:3–15
- Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467:103–107
- Kim DN, Thiel BC, Mrozowich T, Hennelly SP, Hofacker IL, Patel TR, Sanbonmatsu KY (2020) Zinc-finger protein CNBP alters the 3-D structure of lncRNA Braveheart in solution. *Nat Commun* 11:148
- Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal*. <https://doi.org/10.1126/scisignal.2000568>
- Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhilber ML, Ding H, Butty VL, Torrey L, Haas S, Abo R, Tabebordbar M, Lee RT, Burge CB, Boyer LA (2013) Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 152:570–583
- Kumar A, Clerici M, Muckenfuss LM, Passmore LA, Jinek M (2019) Mechanistic insights into mRNA 3'-end processing. *Curr Opin Struct Biol* 59:143–150
- Lee JT, Jaenisch R (1997) The (epi)genetic control of mammalian X-chromosome inactivation. *Curr Opin Genet Dev* 7:274–280
- Li L, Chang HY (2014) Physiological roles of long noncoding RNAs: insight from knockout mice. *Trends Cell Biol* 24:594–602
- Liu F, Somarowthu S, Pyle AM (2017) Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat Chem Biol* 13:282–289
- Liu B, Shi H, Al-Hashimi HM (2021) Developments in solution-state NMR yield broader and deeper views of the dynamic ensembles of nucleic acids. *Curr Opin Struct Biol* 70:16–25
- Lorenz R, Hofacker IL, Stadler PF (2016a) RNA folding with hard and soft constraints. *Algorithms Mol Biol* 11:8
- Lorenz R, Wolfinger MT, Tanzer A, Hofacker IL (2016b) Predicting RNA secondary structures from sequence and probing data. *Methods* 103:86–98
- Ma J, Su D, Sun Y, Huang X, Liang Y, Fang L, Ma Y, Li W, Liang P, Zheng S (2021) Cryo-EM structure of S-Trimer, a subunit vaccine candidate for COVID-19. *J Virol* 95(11):e00194
- Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR (2003) Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* 113:577–586
- Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 306:275–279
- Mathews DH (2019) How to benchmark RNA secondary structure prediction accuracy. *Methods* 162–163:60–67
- Matthews H, Hanison J, Nirmalan N (2016) “Omics”-informed drug and biomarker discovery: opportunities, challenges and future perspectives. *Proteomes* 4(3):e28
- McKay DB, Xi L, Barthel KK, Cech TR (2014) Structure and function of steroid receptor RNA activator protein, the proposed partner of SRA noncoding RNA. *J Mol Biol* 426:1766–1785
- Mercer TR, Mattick JS (2013) Structure and function of long non-coding RNAs in epigenetic regulation. *Nat Struct Mol Biol* 20:300–307
- Minajigi A, Froberg JE, Wei C, Sunwoo H, Kesner B, Colognori D, Lessing D, Payer B, Boukhali M, Haas W, Lee JT (2015) A comprehensive Xist interactome reveals cohesin repulsion and

- an RNA-directed chromosome conformation. *Science*. <https://doi.org/10.1126/science.aab2276>
- Minajigi A, Froberg J, Wei C, Sunwoo H, Kesner B, Colognori D, Lessing D, Payer B, Boukhali M, Haas W, Lee JT (2015) Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science*. <https://doi.org/10.1126/science.aab2276>
- Montange RK, Batey RT (2006) Structure of the S-adenosyl-methionine riboswitch regulatory mRNA element. *Nature* 441:1172–1175
- Montange RK, Batey RT (2008) Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys* 37:117–133
- Moore PB, Engelman DM, Schoenborn BP (1975) A neutron scattering study of the distribution of protein and RNA in the 30 S ribosomal subunit of *Escherichia coli*. *J Mol Biol* 91:101–120
- Morse JC, Girodat D, Burnett BJ, Holm M, Altman RB, Sanbonmatsu KY, Wieden HJ, Blanchard SC (2020) Elongation factor-Tu can repetitively engage aminoacyl-tRNA within the ribosome during the proofreading stage of tRNA selection. *Proc Natl Acad Sci U S A* 117:3610–3620
- Munro JB, Sanbonmatsu KY, Spahn CM, Blanchard SC (2009) Navigating the ribosome's metastable energy landscape. *Trends Biochem Sci* 34:390–400
- Naganuma T, Nakagawa S, Tanigawa A, Sasaki YF, Goshima N, Hirose T (2012) Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J* 31:4020–4034
- Nakagawa S, Hirose T (2012) Paraspeckle nuclear bodies—useful uselessness? *Cell Mol Life Sci* 69:3027–3036
- Noller HF, Woese CR (1981) Secondary structure of 16S ribosomal RNA. *Science* 212:403–411
- Noller HF, Kop J, Wheaton V, Brosius J, Gutell RR, Kopylov AM, Dohme F, Herr W, Stahl DA, Gupta R, Waese CR (1981) Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res* 9:6167–6189
- Novikova IV, Hennelly SP, Sanbonmatsu KY (2012) Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res* 40:5034–5051
- Novikova IV, Hennelly SP, Tung CS, Sanbonmatsu KY (2013a) Rise of the RNA machines: exploring the structure of long non-coding RNAs. *J Mol Biol* 425:3731–3746
- Novikova IV, Hennelly SP, Sanbonmatsu KY (2013) 3S: shotgun secondary structure determination for long non-coding RNAs. *Methods* 63(2):170–177
- Numata K, Kanai A, Saito R, Kondo S, Adachi J, Wilming LG, Hume DA, Hayashizaki Y, Tomita M, R.G. Group, G.S.L. Members (2003) Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res* 13:1301–1306
- Ouyang Z, Snyder MP, Chang HY (2013) SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res* 23:377–387
- Parisien M, Major F (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452:51–55
- Pollack L (2011) SAXS studies of ion-nucleic acid interactions. *Annu Rev Biophys* 40:225–242
- Ponjavic J, Ponting CP, Lunter G (2007) Functionality or transcriptional noise? Evidence for Selection within Long Noncoding RNAs. *Genome Res* 17:556–565
- Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* 136:629–641
- Pucci F, Zerihun MB, Peter EK, Schug A (2020) Evaluating DCA-based method performances for RNA contact prediction by a well-curated data set. *RNA* 26:794–802
- Pyle AM (2016) Group II Intron Self-Splicing. *Annu Rev Biophys* 45:183–205
- Rambo RP, Tainer JA (2013) Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annu Rev Biophys* 42:415–441
- Regulski EE, Breaker RR (2008) In-line probing analysis of riboswitches. *Methods Mol Biol* 419:53–67
- Reyes FE, Garst AD, Batey RT (2009) Strategies in RNA crystallography. *Methods Enzymol* 469:119–139
- Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81:145–166
- 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* 505:701–705
- Roy S, Onuchic JN, Sanbonmatsu KY (2017b) Cooperation between magnesium and metabolite controls collapse of the SAM-I riboswitch. *Biophys J* 113:348–359
- Roy S, Hennelly SP, Lammert H, Onuchic JN, Sanbonmatsu KY (2019) Magnesium controls aptamer-expression platform switching in the SAM-I riboswitch. *Nucleic Acids Res* 47:3158–3170
- Roy S, Lammert H, Hayes RL, Chen B, LeBlanc R, Dayie TK, Onuchic JN, Sanbonmatsu KY (2017a) A magnesium-induced triplex pre-organizes the SAM-II riboswitch. *PLoS Comput Biol* 13:e1005406
- Rummel DP, Noller HF (1973) Use of protection of 30S ribosomal proteins by tRNA for functional mapping of the *E. coli* ribosome. *Nat New Biol* 245:72–75
- Sanbonmatsu KY (2006) Alignment/misalignment hypothesis for tRNA selection by the ribosome. *Biochimie* 88:1075–1089
- Sanbonmatsu KY (2012) Computational studies of molecular machines: the ribosome. *Curr Opin Struct Biol* 22:168–174
- Sanbonmatsu KY (2019) Large-scale simulations of nucleoprotein complexes: ribosomes, nucleosomes, chromatin, chromosomes and CRISPR. *Curr Opin Struct Biol* 55:104–113
- Sanbonmatsu KY, Joseph S, Tung CS (2005) Simulating movement of tRNA into the ribosome during decoding. *Proc Natl Acad Sci U S A* 102:15854–15859
- Sasaki YT, Ideue T, Sano M, Mituyama T, Hirose T (2009) MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci USA* 106:2525–2530
- Sauvageau M, Goff L.A., Lodato S, Bonev B, Groff A.F, Gerhardinger C, Sanchez-Gomez D.B, Hacisuleyman E, Li E, Spence M, Liapis S.C, Mallard W, Morse M, Swerdel M.R, D'Ecclesiss M.F, Moore J.C, Lai V, Gong G, Yancopoulos G.D, Friendewey D, Kellis M, Hart R.P, Valenzuela D.M, Arlotta P, Rinn J.L 2013 Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* 2: e01749
- Sekhar A, Kay LE (2019) An NMR view of protein dynamics in health and disease. *Annu Rev Biophys* 48:297–319
- Serganov A, Patel DJ (2012) Metabolite recognition principles and molecular mechanisms underlying riboswitch function. *Annu Rev Biophys* 41:343–370
- Sherlock ME, Hartwick EW, MacFadden A, Kieft JS (2021) Structural diversity and phylogenetic distribution of valyl tRNA-like structures in viruses. *RNA* 27:27–39
- Smathers CM, Robart AR (2019) The mechanism of splicing as told by group II introns: ancestors of the spliceosome. *Biochim Biophys Acta Gene Regul Mech* 1862:e194390
- Smith JC, Tan P, Petridis L, Hong L (2018) Dynamic neutron scattering by biological systems. *Annu Rev Biophys* 47:335–354
- Somarowthu S, Legiewicz M, Chillon I, Marcia M, Liu F, Pyle AM (2015) HOTAIR forms an intricate and modular secondary structure. *Mol Cell* 58:353–361
- Spasic A, Assmann SM, Bevilacqua PC, Mathews DH (2018) Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic Acids Res* 46:314–323
- Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET, Chang HY

- (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* 519:486–490
- Spokoini-Stern R, Stamov D, Jessel H, Aharoni L, Haschke H, Giron J, Unger R, Segal E, Abu-Horowitz A, Bachelet I (2020) Visualizing the structure and motion of the long noncoding RNA HOTAIR. *RNA* 26:629–636
- Stoddard CD, Montange RK, Hennelly SP, Rambo RP, Sanbonmatsu KY, Batey RT (2010) Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure* 18:787–797
- Sudarsan N, Hammond MC, Block KF, Welz R, Barrick JE, Roth A, Breaker RR (2006) Tandem riboswitch architectures exhibit complex gene control functions. *Science* 314:300–304
- Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* 321:411–413
- Swiezewski S, Liu F, Magusin A, Dean C (2009) Cold-induced silencing by long antisense transcripts of an arabidopsis Polycomb target. *Nature* 462:799–802
- Takizawa Y, Ho CH, Tachiwana H, Matsunami H, Kobayashi W, Suzuki M, Arimura Y, Hori T, Fukagawa T, Ohi MD, Wolf M, Kurumizaka H (2020) Cryo-EM structures of centromeric tri-nucleosomes containing a central CENP-A nucleosome. *Structure* 28:44–53
- Tan Z, Fu Y, Sharma G, Mathews DH (2017) TurboFold II: RNA structural alignment and secondary structure prediction informed by multiple homologs. *Nucleic Acids Res* 45:11570–11581
- Tang TTL, Stowell JAW, Hill CH, Passmore LA (2019) The intrinsic structure of poly(A) RNA determines the specificity of Pan2 and Caf1 deadenylases. *Nat Struct Mol Biol* 26:433–442
- Torabi S.F, Chen Y.L, Zhang K, Wang J, DeGregorio S.J, Vaidya A.T, Su Z, Pabit S.A, Chiu W, Pollack L, Steitz J.A. 2021 Structural analyses of an RNA stability element interacting with poly(A). *Proc Natl Acad Sci U S A* 118
- Tung C.S, Sanbonmatsu K.Y 2004 Atomic model of the T thermophilus 70S ribosome developed in silico. *Biophys J Submitted*
- Ulitsky I (2016) Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet* 17:601–614
- Ulitsky I, Bartel DP (2013) lincRNAs: genomics, evolution, and mechanisms. *Cell* 154:26–46
- Uroda T, Anastasakou E, Rossi A, Teulon JM, Pellequer JL, Annibale P, Pessey O, Inga A, Chillon I, Marcia M (2019) Conserved pseudoknots in lncRNA MEG3 are essential for stimulation of the p53 pathway. *Mol Cell* 75:982–995
- Voorhees RM, Ramakrishnan V (2013) Structural basis of the translational elongation cycle. *Annu Rev Biochem* 82:203–236
- Wan Y, Qu K, Ouyang Z, Kertesz M, Li J, Tibshirani R, Makino DL, Nutter RC, Segal E, Chang HY (2012) Genome-wide measurement of RNA folding energies. *Mol Cell* 48:169–181
- Wan Y, Qu K, Ouyang Z, Chang HY (2013) Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat Protoc* 8:849–869
- Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E, Chang HY (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505:706–709
- Wasserman MR, Alejo JL, Altman RB, Blanchard SC (2016) Multiperspective smFRET reveals rate-determining late intermediates of ribosomal translocation. *Nat Struct Mol Biol* 23:333–341
- Watson JD, Crick FH (1953) The structure of DNA. *Cold Spring Harb Symp Quant Biol* 18:123–131
- Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, Sudarsan N, Neph S, Tompa M, Ruzzo WL, Breaker RR (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* 35:4809–4819
- Westhof E (2015) Twenty years of RNA crystallography. *RNA* 21:486–487
- Wilkinson ME, Charenton C, Nagai K (2020) RNA splicing by the spliceosome. *Annu Rev Biochem* 89:359–388
- Wilson RC, Doudna JA (2013) Molecular mechanisms of RNA interference. *Annu Rev Biophys* 42:217–239
- Wilusz JE, Freier SM, Spector DL (2008) 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 135:919–932
- Wilusz JE, JnBaptiste CK, Lu LY, Kuhn CD, Joshua-Tor L, Sharp PA (2012) A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev* 26:2392–2407
- Winkle M, El-Daly SM, Fabbri M, Calin GA (2021) Noncoding RNA therapeutics - challenges and potential solutions. *Nat Rev Drug Discov* 18:1–23
- Winkler W, Nahvi A, Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419:952–956
- Winkler WC, Nahvi A, Sudarsan N, Barrick JE, Breaker RR (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Biol* 10:701–707
- Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428:281–286
- Woese CR, Magrum LJ, Gupta R, Siegel RB, Stahl DA, Kop J, Crawford N, Brosius J, Gutell R, Hogan JJ, Noller HF (1980) Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res* 8:2275–2293
- Xu B, Yang WH, Gerin I, Hu CD, Hammer GD, Koenig RJ (2009) Dax-1 and steroid receptor RNA activator (SRA) function as transcriptional coactivators for steroidogenic factor 1 in steroidogenesis. *Mol Cell Biol* 29:1719–1734
- Xue Z, Hennelly S, Doyle B, Gulati AA, Novikova IV, Sanbonmatsu KY, Boyer LA (2016) A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. *Mol Cell* 64:37–50
- Yan C, Wan R, Shi Y (2019) Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome. *Cold Spring Harb Perspect Biol* 11(1):e032409
- Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, Felsenfeld G (2010) Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev* 24:2543–2555
- Yao J, Reinharz V, Major F, Waldispuhl J (2017) RNA-MoIP: prediction of RNA secondary structure and local 3D motifs from sequence data. *Nucleic Acids Res* 45:W440–W444
- Zhang J, Jones CP, Ferre-D'Amare AR (1839) Global analysis of riboswitches by small-angle X-ray scattering and calorimetry. *Biochim Biophys Acta* 2014:1020–1029
- Zhang K, Li S, Kappel K, Pintilie G, Su Z, Mou TC, Schmid MF, Das R, Chiu W (2019) Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat Commun* 10:5511
- Zhang K, Zheludev I.N, Hagey R.J, Wu M.T, Haslecker R, Hou Y.J, Kretsch R, Pintilie G.D, Rangan R, Kladwang W, Li S, Pham E.A, Bernardin-Souibgui C, Baric R.S, Sheahan T.P, D'Souza V, Glenn J.S, Chiu W, Das R 2020 Cryo-electron Microscopy and Exploratory Antisense Targeting of the 28-kDa Frameshift Stimulation Element from the SARS-CoV-2 RNA Genome. *bioRxiv*
- Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* 40:939–953