Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs

Irina V. Novikova, Scott P. Hennelly, Chang-Shung Tung and Karissa Y. Sanbonmatsu

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

Correspondence to Karissa Y. Sanbonmatsu: *kys@lanl.gov*
http://dx.doi.org/10.1016/j.jmb.2013.02.030
Edited by A. Pyle

Abstract

Novel, profound and unexpected roles of long non-coding RNAs (lncRNAs) are emerging in critical aspects of gene regulation. Thousands of lncRNAs have been recently discovered in a wide range of mammalian systems, related to development, epigenetics, cancer, brain function and hereditary disease. The structural biology of these lncRNAs presents a brave new RNA world, which may contain a diverse zoo of new architectures and mechanisms. While structural studies of lncRNAs are in their infancy, we describe existing structural data for lncRNAs, as well as crystallographic studies of other RNA machines and their implications for lncRNAs. We also discuss the importance of dynamics in RNA machine mechanism. Determining commonalities between lncRNA systems will help elucidate the evolution and mechanistic role of lncRNAs in disease, creating a structural framework necessary to pursue lncRNA-based therapeutics.

Introduction

RNA is primarily known as an intermediary in gene expression between DNA and proteins. Over the past several decades, other roles for RNA have been identified, which include protein synthesis, gene regulation and nucleic acid processing. However, these RNAs were considered as outliers in a transcriptome that consists mainly of protein-coding RNAs. While this may be the case for bacteria and certain unicellular organisms,[1] transcriptomes in higher eukaryotes are markedly different. Recent deep sequencing studies in humans have shown that more than two-thirds of the genome is actively transcribed.[2] Since protein-coding genes constitute a very small fraction of the genome,[3] non-coding RNAs represent the vast majority (more than 80% in mammals).[4] Recent studies estimate approximately 15,000 long non-coding RNAs (lncRNAs) in humans.[5] In light of their expression profiles specific to cell, cell cycle, tissue, developmental stage and disease, it is difficult to precisely quantify the number of human lncRNAs. We note that tens of thousands of lncRNAs have been profiled in 2012.[6–10] This recent explosion of newly discovered lncRNAs suggests that non-coding RNAs may be the norm rather than the exception in the case of eukaryotic organisms.

LncRNAs are defined by the following: (i) lack of coding potential and (ii) transcript length (>200 nt).[11] These transcripts are generally nuclear retained and transcribed by RNA polymerase II, with many that are spliced and polyadenylated.[12–15] LncRNAs may be intronic, intergenic (large intervening non-coding RNAs or lincRNAs) or antisense to the protein-coding genes (overlapping one or more exons). While transcript lengths are normally in the range 1000–10,000 residues, the Air and kcnq1ot1 lncRNAs have lengths that exceed 90 kb.[16–18] Interestingly, in 2013, a novel "monster" non-coding transcript, XACT (252 kb in length), was found to originate from the active X chromosome in human pluripotent cells.[19] Many lncRNAs play key roles in signaling,[20–22] development, cancer,[12,24–29] embryonic stem cell pluripotency,[30] brain function,[31–34] subcellular compartmentalization,[26,35–38] chromatin remodeling,[12,39,40] plant biology[41,42] and stress response.[43] Cell biology studies and functional aspects of lncRNAs have been discussed by Rinn and Chang in a recent review article, as well as in many references therein.[6] Since lncRNAs are often associated with histone modification and chromatin

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

remodeling, epigenetic effects represent a unifying theme. While the mechanism by which epigenetic factors find their targets is largely unknown, many recent studies indicate that lncRNAs may be a necessary component of the network that guides these factors to their chromatin targets. In addition to epigenetic effects, the origin of lncRNAs is not well understood. While the genomes of simple organisms are dominated by protein-coding genes, the advent of non-coding RNAs in higher organisms gave rise to an incredible sequence space for exploration of new function at the RNA level, creating sophisticated regulatory networks. Possible evolution mechanisms may include pseudogenization of protein-coding genes, insertion of transposons and duplication of genes. Investigations into the open area of the origin and evolution of lncRNAs have the potential to yield exciting results.

RNA machines

RNA molecules are able to form complex molecular machines. The ribosome is one of the most well known RNA machines since it has many moving parts and may be powered by GTP hydrolysis.[44,45] In certain circumstances, the ribosome can function without GTP hydrolysis ("factor-free translation").[46] Factor-free translation is quite inefficient in comparison to GTP-based translation and in comparison to the far more efficient protein-based molecular motors (e.g., flagellum).[47] The ribosome also processes information by performing a look-up table operation, converting the 4-letter nucleic acid alphabet into the 20-letter protein alphabet. The group I and group II self-splicing introns are also molecular machines as they use chemical energy to bring together distant regions of RNA.[48–50] RNase P and telomerase RNA also catalyze chemical reactions to accomplish their function.[51–53] The riboswitch is another type of RNA molecular machine.[54] Riboswitches are molecular switches that sense their environment and allow gene expression decisions to be made on the basis of environmental inputs such as ligand concentration.[55–60] More complicated tandem riboswitches act as logical AND gates. In addition, complex combinations of aptamers can be arranged into logic circuitry (e.g., OR, NAND, NOR and NOT gates).[61] Finally, RNA scaffolding that controls information flow could be analogous to an integrated circuit (a device that has no moving parts but controls highly complex information flow in computers). Recent studies suggest that several different lncRNA systems may act as scaffolding, controlling information flow in epigenetic systems.[62]

Structural studies are often critical in deciphering RNA machines and the mechanism of RNA action. Few structural studies of individual lncRNAs have been performed due to their enormous size and very recent discovery.[20,63] Considering that the high-resolution structure of the bacterial ribosome (>4500 nt) required more than two decades for its solution, structural studies of lncRNAs are formidable.[64,65] Fundamental questions regarding the structure of lncRNAs remain unanswered, including:

(1) Is lncRNA mechanism dominated by primary sequence, higher-order structure or both?
(2) Do lncRNAs contain sub-domains?
(3) Are lncRNAs complexed with many proteins or do they exist as isolated RNAs that transiently interact with proteins?
(4) Are lncRNAs compact or extended?

Here, we attempt to provide a starting point for structural studies of lncRNAs. Because tertiary structures have not been solved to date, we review tertiary structures of other RNA machines. ~~We next~~ summarize currently known structural features of eukaryotic lncRNAs (Fig. 1). Finally, we discuss lncRNA structure in the context of previously studied RNA molecular machines and corresponding structure/function relationships.

Crystallographic studies of RNA machines in bacteria and other unicellular organisms

Three-dimensional structure is often integral to the function and mechanism of a biomolecular system. Here, we review crystallographic structures of other RNA systems. This review is by no means comprehensive but highlights the diversity of mechanism and composition of RNA structure. The group I intron, the group II intron, the ribosome and RNase P are the only RNAs with lengths greater than 200 nt that have been studied at high resolution.[48,66]

The group I and group II introns are self-splicing RNAs that catalyze their own cleavage. The structure of the group I intron consists largely of an isolated RNA (~200 nt), organized into a well-defined architecture by co-axially stacked helices connected by single-stranded regions. The pre-2S state of the group I intron includes an exon and also contains mainly co-axially stacked helices connected by single-stranded regions.[67] Four of the helices are capped by tetraloops. This structure consists largely of an isolated RNA with one small spliceosomal protein (U1A) bound to its extremity. The active site contains two base triples, critical for precise positioning of magnesium ions and catalysis.

The structure of the post-catalytic state of the group II intron, solved by Toor *et al.*, contains approximately 400 nt, organized into more than 10 helices, with most capped by RNA stem–loop structures.[68] The RNA contains many co-axially stacked helices, forming a relatively compact overall

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

Q2

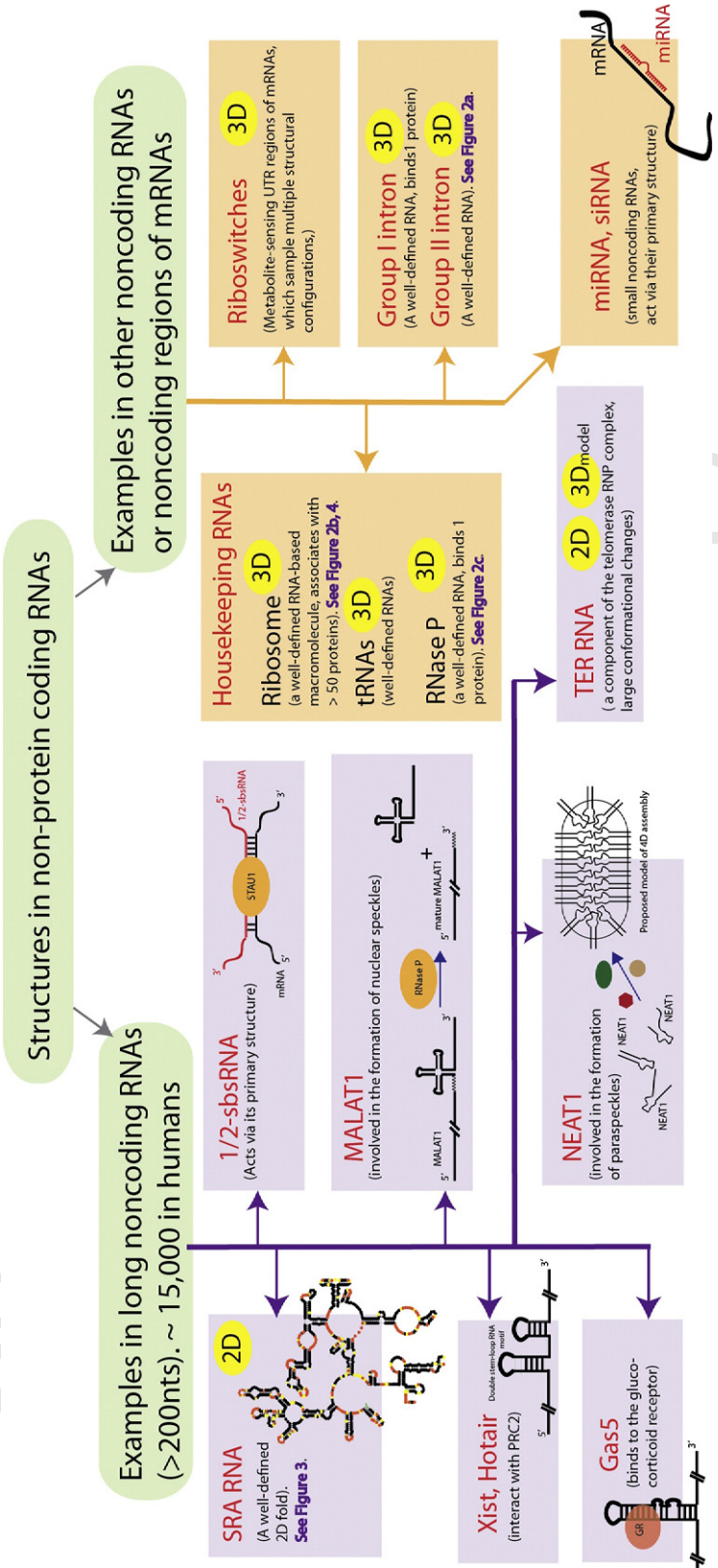

Fig. 1. Examples of non-coding RNAs studied with structural methods. Purple, lncRNAs with existing structural information. Orange, examples of non-coding RNAs with solved crystallographic structures. "3D" denotes systems with crystallographic structures; "2D" denotes systems with experimentally derived secondary structures of the entire RNA.

structure. A large number of long-range tertiary interactions connect the helices, including a kissing loop interaction, a ribose zipper interaction and an intricate five-way junction with many stacked non-helical nucleotides. Several base triples exist in the Z-anchor region, nearby the magnesium binding

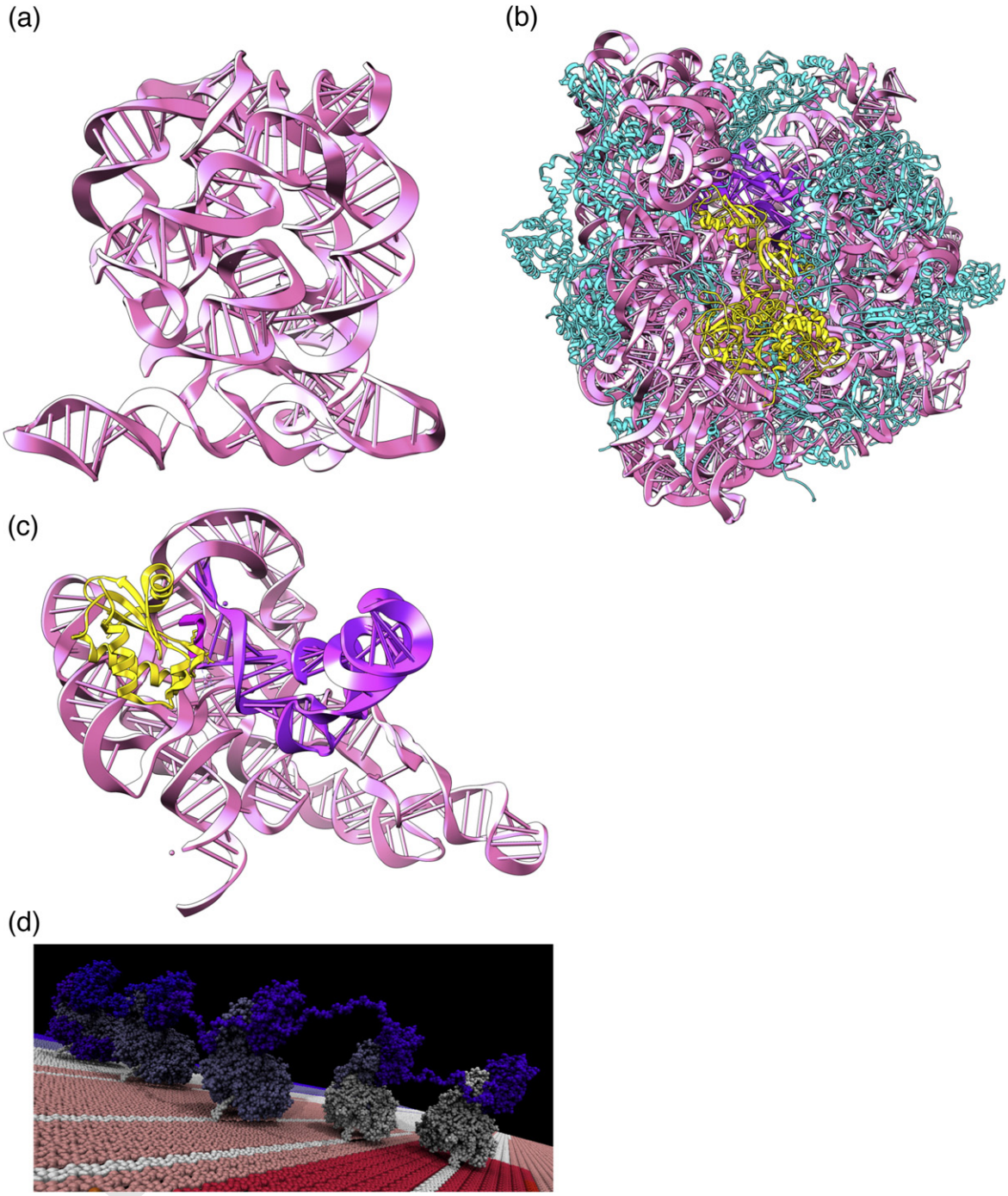

Fig. 2. Tertiary structures of previously solved RNA complexes. (a) The group II intron, solved by Chan *et al.*, is a highly structured isolated RNA with a compact core.[48] (b) The ribosome, solved by Gao *et al.*, is highly structured and compact.[70] Pink, rRNA; purple, tRNA; cyan, ribosomal proteins; yellow, elongation factor EF-G. Different protein factors bind to the same binding sites, regulating protein synthesis. (c) RNase P.[51] RNase P is structured and contains a single protein binding domain. Pink, RNase P RNA; purple, tRNA; yellow, RNase P protein. (d) Cellulosome protein-based molecular machine. The cellulosome represents a relatively disordered but highly efficient molecular machine. Red, white and pink, cellulose substrate; gray, cellulase subunit; purple, dockerin connecting protein; blue, cohesin (disordered linker protein).

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

Q2

sites associated with catalysis. Marcia and Pyle also uncovered a reversible conformational change in a study of the group II intron, which included 14 structures, capturing several different intermediates of splicing consistent with biochemical studies.[69] A structure of the group II intron prior to the first step of splicing[48] shows a high degree of similarity in overall architecture to the post-splicing configuration (Fig. 2a). Interestingly, a significant conformational change occurs in the core of the intron, where a single-stranded region forms a kink motif, producing a reversal in the direction of the RNA strand at this location. It has also been suggested that the spliceosome is mechanistically similar to the group II intron.[68]

The largest RNA tertiary structure solved to date is the ribosome (Fig. 2b).[70] The architecture of the ribosome comprises two major subunits, whose RNA structure is formed by a large number of shorter RNA helices and a small number of longer helices organized into a compact geometry. Approximately 50 proteins (in the case of bacteria) are interspersed throughout the ribosome complex, producing a highly compact, globular architecture. The globular domains of ribosomal proteins themselves typically reside on the exterior of the complex and possess long tails extending to the interior of the ribosome, near the active sites in some cases. The 70S complex is interconnected by an enormous number of tertiary contacts, including both RNA/RNA and RNA/protein interactions. The structure of the ribosome is uniquely designed for its function: the ribosome must (i) accept only correct tRNAs carrying amino acids corresponding to the mRNA codon, (ii) perform chemistry (adding the amino acid to the nascent protein)[71] and (iii) move the tRNAs through the ribosome.[72] A large cavity exists between the small and large subunits. This cavity is used to transport the tRNA ligands through the ribosome. The six major active sites of the ribosome are RNA based. In addition to ribosomal proteins, other protein factors bind to the ribosome, facilitating the initiation and termination of translation, translocation of the ribosome along the mRNA, improved fidelity of tRNA selection and regulation of protein synthesis. While ribosomal proteins bind all over the ribosome, protein factors bind at a few select sites, associating and dissociating with the ribosome at various stages. These include the GTP-associated center and the three tRNA binding sites (the aminoacyl site, the peptidyl site and the exit site). Elongation factors EF-Tu and EF-G each bind to GTP-associated center during different stages of the elongation.

Several different RNA action mechanisms allow the ribosome to accomplish its function. Locally, a conformational selection mechanism may act at the decoding center (the aminoacyl site on the small subunit) during certain stages of elongation. During translocation of the ribosome along the mRNA, many large-scale conformational fluctuations occur simultaneously and at different timescales. Protein binding or GTP hydrolysis events act to synchronize the fluctuations, shifting the equilibrium to the next basin in the energy landscape and allowing the ribosome to progress through the elongation cycle.[45,73]

RNase P (~330–400 nt) can be considered a canonical RNA that is compact, highly structured and binds a single protein (Fig. 2c).[51] RNase P is an RNA-based multiple-turnover enzyme that catalyzes 5′ end maturation of tRNA and other RNAs. While the complex is dominated by RNA, a small protein component (approximately 10% of total mass) increases the affinity of tRNA to RNase P. The overall architecture of the RNA is similar to the self-splicing introns, consisting mainly of co-axially stacked helices, connected by various tertiary interactions.[74]

The structure of the intact telomerase RNA, TER (~450 nt in humans), has yet to be solved; however, this system represents an interesting RNA in the context of lncRNAs.[52] The template for telomeric repeat extension resides within the TER. The telomerase RNA (TER) binds to telomerase reverse transcriptase (TERT), facilitating nucleotide addition to telomeric regions. Crystal structures of portions of the telomerase RNA have been solved. These structures show a high degree of organization with several tertiary interaction motifs, including a well-studied and functionally important pseudoknot near the RNA template sequence. The structural integrity of the RNA is dependent to some degree on the interactions of TERT and accessory proteins.[75] In yeast, the precise location of the telomerase subunit (est1p's binding domain on this RNA) is not the key to its function and can be positioned within the RNA at different locations while maintaining function.[53] These important experiments suggest that the telomerase RNA may act as flexible scaffolding for protein binding. Another telomerase-associated RNA is the telomere-repeat-containing RNA (TERRA). This RNA is transcribed from positions near the telomeric ends and includes sequence from the telomeric repeats. This RNA has been found to associate with chromatin. TERRA appears to localize near the telomeres. Its association is negatively regulated by effectors of the nonsense-mediated RNA decay pathway.[76] The presence of increased concentrations of TERRA may inhibit telomerase.[77]

LncRNA mechanisms based on primary structure

In several cases, mechanistic information has been gained by studying the primary structure of the lncRNA. Here, the role of lncRNA is mainly to provide sequence specificity to a process. Such mechanisms are well known in other RNA systems such as RNAi (*RNA interference*), where base pairing between short regulatory RNAs [siRNA

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

(*small interfering RNA*) or miRNA (*micro RNA*)] and mRNA plays a critical role in gene regulation.[78] RISC (*RNA-induced silencing complex*) complexes use these short RNAs as templates to locate and silence their mRNA targets.[78–84] The protein-based RISC complex provides the structural architecture and is responsible for either directly cleaving the target or down-regulating translation.[85]

A first example is the 1/2-sbsRNA, which facilitates mRNA decay by base pairing with the 3′-untranslated region (STAU1-mediated mRNA decay).[13] Once bound, the lncRNA/mRNA complex is recognized by the double-stranded RNA binding protein, STAU1, triggering mRNA decay. Another example is the DHFR lncRNA, which is transcribed from a minor promoter of the dihydrofolate reductase gene and plays a key role in the epigenetic mechanism of promoter-specific repression of transcription.[86] Here, the lncRNA forms a complex with the major promoter and also interacts with the transcription factor IIB, disrupting the preinitiation complex. Additional evidence exists in other systems for the cotranscriptional recruitment of chromatin modifying complexes utilizing base pairing between the nascent RNA and its antisense transcript.[87] Antisense lncRNAs have been also shown to regulate splicing by interacting with mRNAs. The long non-coding Zeb2 natural antisense transcript is antisense to a 5′ splice site of the Zeb2 mRNA. Binding of Zeb2 inhibits splicing of mRNA and, as a result, preserves the internal ribosome entry site, necessary for efficient translation.[88] Linc-MD1, associated with muscle differentiation, is an example of an lncRNA that serves as a decoy for miRNAs. This lncRNA regulates miRNA action by providing alternative binding sites, effectively titrating the miR-133 and miR-135 away from their targets.[89] These binding sites tend to be relatively short (i.e., much shorter than the total length of the lncRNA). The much larger remainder of the lncRNA may play an additional structural role in the lncRNA's mechanism of action yet to be determined.

Structural probing studies of lncRNAs

Secondary structure often plays a critical role in RNA mechanism, underpinning overall tertiary architecture by defining helices, bulges, stem–loops, internal loops, junctions and sub-domains. Secondary structure alone can provide the basis of function when (i) the secondary structure defines a platform for unique protein recognition or (ii) a single sequence produces two competing secondary structures, as in the case of the riboswitch. The riboswitch represents a quintessential secondary-structure-based RNA mechanism.[55–60] Here, a single sequence of RNA has two different secondary structures, which compete with each other to determine the outcome of gene expression. Often, ligand concentration shifts the equilibrium between secondary structures, resulting in the formation or destruction of a transcriptional terminator helix. In riboswitch studies, the action mechanism (i.e., that the RNA is a ligand-based molecular switch) is determined by studying the secondary structure, without the need for tertiary structure studies. Breaker *et al.*, who have discovered most of the known riboswitches, used chemical probing experiments (in-line probing) to determine the secondary structure of the riboswitch in the presence or absence of the ligand.[56,57,90,91] Their technique has been validated in many crystallographic studies.[54,59,92,93] While many computational techniques have been developed to predict secondary structure, predicting the secondary structure of long RNA sequences remains a significant challenge.[94–98] One promising strategy uses machine learning approaches.[95–97] Our experimentally determined secondary structures are convenient benchmarks for new predictive algorithms.

Our group has produced the first experimentally derived secondary structure of an intact lncRNA, the steroid receptor RNA activator (SRA).[20] This lncRNA co-activates a variety of sex hormone receptors (ER, AR, TR, GR, RAR) and has been shown to directly interact with several proteins (ER, SHARP, SLIRP, DAX-1, SF-1, TR, Pus1p).[99–101] SRA has been found to associate with CTCF and is thought to play a scaffolding role in the transcription complex.[102] SRA is also strongly associated with breast cancer and may be useful as an early onset tumorigenesis marker. As many lncRNAs are similar in size to ribosomal subunits, we followed the footsteps of Noller, Woese and Gutell by (i) performing extensive chemical probing to produce a secondary structure of the full lncRNA (~870 nt) and (ii) validating with covariance analysis across multiple sequences (Fig. 3).[103,104]

Our experiments uncovered an intricate and highly structured two-dimensional architecture of the lncRNA, organized into four major sub-domains. To determine the secondary structure, we employed four methods of structural probing: selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE), in-line probing, DMS probing and RNase V1 digestion. The four methods yielded complementary information about the secondary structure of the intact lncRNA and were highly consistent with each other. Overall, the lncRNA comprises 25 helical segments (H1–H25), 16 terminal loops, 15 internal loops and 5 junction regions (Fig. 3), consistent with the 16S rRNA (~1540 nt), which contains 45 helices, 31 terminal loops, 26 internal loops and 18 junction regions. In the SRA system, helices H1–H7 comprise domain I. Helices H10–H14 comprise domain II. Helices H8, H9 and H15–H21 comprise domain III, while helices H22–H25 comprise domain IV. While domains I–III are highly conserved, representing the core region of SRA, domain IV is evolutionarily

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

Q2

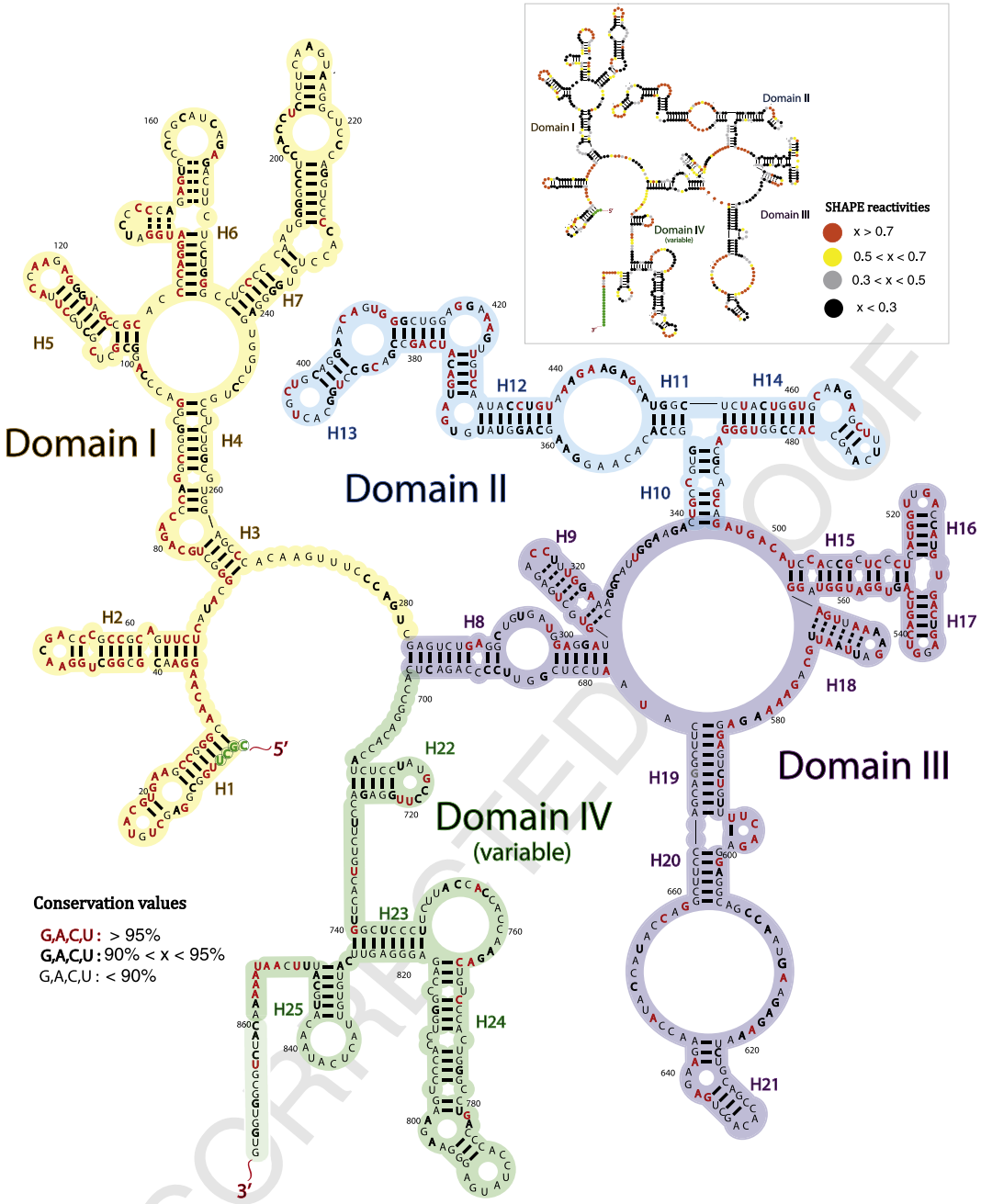

Fig. 3. First experimentally derived secondary structure of an intact higher eukaryote lncRNA, to our knowledge. The human SRA has a length of 870 nt, organized into 4 sub-domains and 25 helices. Four biochemical techniques were used to obtain the secondary structure (SHAPE, in-line, DMS and RNase V1). The vast majority of helices were validated with covariance analysis based on multiple sequence alignment across vertebrates. Yellow, domain I; blue, domain II; purple, domain III; green, domain IV. Inset: secondary structure annotated with SHAPE probing results (SHAPE). Orange/red, high reactivity nucleotides; yellow, medium reactivity; gray/black, low/no mobility.

divergent and highly variable in sequence across mammals. Therefore, our initial expectation was that this region would have little structure because there was no obvious sequence conservation across species. However, we found that this domain does comprise a number of helical components, even though the helical density is slightly lower compared to the rest of the sequence. This is consistent with the recently solved crystallographic structure of the 80S yeast ribosome, whose expansion segments are highly variable in sequence yet highly structured in the crystallographic study.[105]

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

Q2

We also found that the SRA RNA secondary structure contains many highly conserved, purine-rich elements, primarily located in single-stranded regions of RNA. The same preference toward single-stranded regions is also found in riboswitches and ribosomes, attributed to the base-stacking propensity of purines. For example, in the eukaryotic ribosome, expansion segment 6 (ES6) contains purine-rich single-stranded regions.[105] While, at first glance, this secondary structure might suggest a significant amount of disorder, the crystallographic structure reveals a high degree of organization, with many of the single-stranded nucleotides participating in non-Watson-Crick base pairs (Fig. 4). Using a multiple sequence alignment across 45 species for the SRA gene, we were able to calculate the degree of covariation of base pairs between species. Helices H2, H3, H4, H6, H7, H8, H9, H12, H13, H14, H15, H19, H20 and H21 possess at least one covariant base pair.

Generally, the longer the RNA sequence, the more alternative folding choices are present. Probing experiments on SRA were sufficient in most cases to select the correct regional fold, but for some RNA sections, more information was required. In these regions, we opted to use an additional experimental technique, Shotgun Secondary Structure determination, where sub-fragments of the full RNA are probed to determine modularly folded sub-domains of SRA.[20] This technique enabled us to determine much of the secondary structures of domain II and domain III. We note that several interesting studies of viral secondary structures have been performed using SHAPE probing.[106–109] To obtain the secondary structure fold, we incorporated SHAPE reactivity data with structure predictions based on thermodynamic parameters. Covariance analysis and fragment-based probing strategies such as the shotgun methodology can be used to help validate these structures.

In addition to studies of single intact lncRNAs, Chang *et al.* are pioneering genome-wide studies of RNA secondary structure. Performing RNase digestion as a function of temperature, they have produced a technique capable of measuring RNA folding energies on a genome-wide scale.[63] The

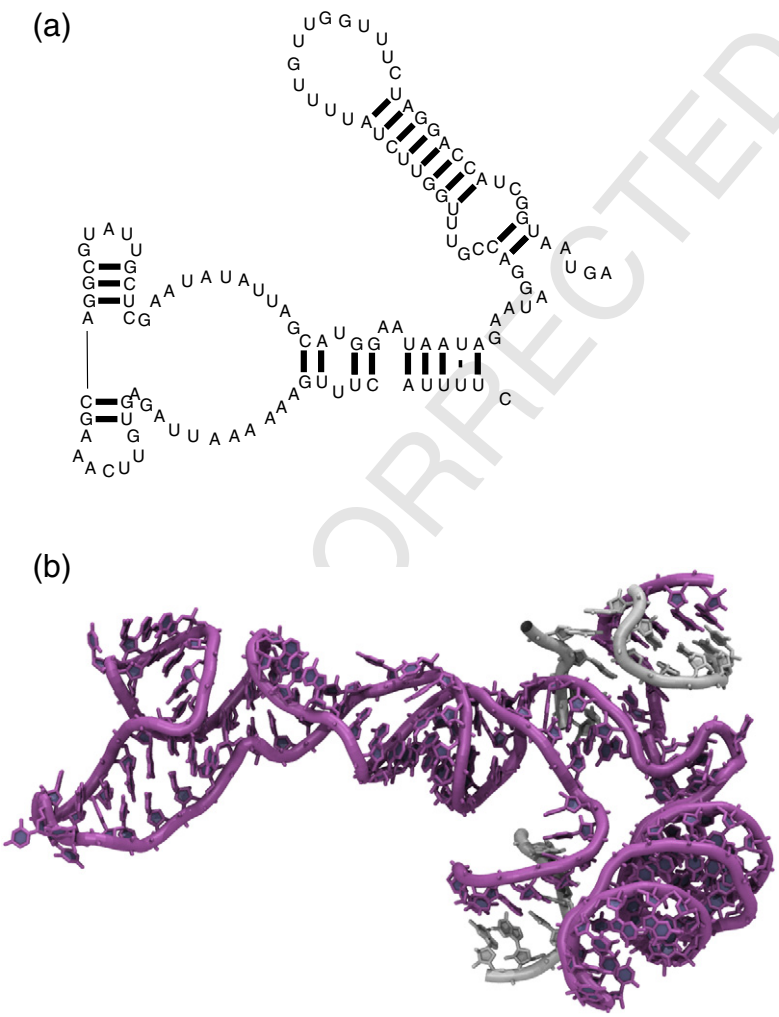

Fig. 4. Expansion segment ES6 of the yeast ribosome small subunit. (a) Secondary structure of ES6 displays a large internal loop (22 nt) and a large stem–loop (13-nt loop). (b) Despite the lack of canonical base pairs, both the internal loop and the stem–loop are highly structured in the recent X-ray structure of the yeast ribosome solved by Ben-Shem *et al.*[105] Many of the purines in the internal loop form noncanonical base pairs.

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

Q2

experiments reveal key features near the start codon in mRNAs. The study also suggests that lncRNAs, on average, appear to be structured, with a level of structural content in between the structural content of the ribosome and the structural content of a typical mRNA. Lucks *et al.* have developed a method to perform SHAPE probing with the Illumina platform, which may also be scalable to genome-wide studies.[110] New SHAPE reagents for *in vivo* RNA probing, with better solubility and increased half-lives, have recently been introduced by Spitale *et al.*[111]

Interesting structural elements involved in non-canonical 3′ end processing mechanisms have been determined for other lncRNA.[26,27,112] Maturation of the NEAT1_v2 (MEN beta) and MALAT1 transcripts relies on the formation of a cloverleaf (four-way junction) secondary structure. It is a highly conserved element, which mimics pre-tRNA substrate and recruits the RNase P complex. Subsequently, RNase P cleaves upfront of this element, generating the mature ends of NEAT1_v2 and MALAT1 transcripts. The cleaved 3′ end fragments undergo an additional processing by RNase Z to yield shorter tRNA-like transcripts, named mascRNA (generated from MALAT1) and menRNA (generated from NEAT1_v2).

The human accelerated region RNA, HAR1, associated with neocortex development, also contains a cloverleaf element, as evidenced by *in vitro* structural probing studies.[113] Interestingly, significant sequence and structure divergence is observed between human and chimpanzee (18 mutations), attributed to brain evolution.[114,115] In chimpanzee, structure probing of this region revealed an extended hairpin structure.[113] It has been suggested that the chimpanzee RNA may form a cloverleaf element in the presence of an additional protein.

Another element observed in lncRNAs is the double stem–loop, often associated with chromatin remodeling. This is a particularly interesting example because many lncRNAs have been shown to play important roles in chromatin remodeling, a focus of many transcriptome-wide lncRNA studies.[15,116,117] In embryonic stem cells, Zhao *et al.* and Surface *et al.* discovered >9000 lncRNAs that interact with the polycomb repressive complex (PRC2).[118,119] Biochemical analysis of PRC2-interacting lncRNAs showed that binding occurs through EZH2, a component of PRC2. In another study, a CLIP-seq investigation of RNAs associated with the SFRS1 splicing factor uncovered >6000 spliced non-coding RNAs. Overall, chromatin remodeling lncRNAs appear to act *in trans* via association with chromatin modifying enzymes.[116] It has been suggested, for the XIST system, that some of these interactions may occur via double stem–loop RNA elements.[120,121]

Double stem–loop RNA motifs have also been implicated in a PRC2-binding region of the HOTAIR lncRNA.[12] HOTAIR (2.2 kb) regulates HoxD genes by recruiting the PRC2 and lysine-specific demethylase (LSD), which each modify histones at targeted loci. Deletions of portions of the HOTAIR RNA show that PRC2 binds to a 300-nt region in the 5′ end of HOTAIR and LSD binds to a 646-nt region at the 3′ end. The intervening sequence may spatially organize the two interaction sites. This region may also contain motifs necessary for targeting. The exact structures of the LSD1 binding motifs have not been determined. Another system, the growth-arrest specific non-coding transcript (gas5), contains a hairpin element. This element is responsible for regulation of the glucocorticoid receptor (GR) via a decoy mechanism, mimicking the DNA hormone-responsive element.[21]

Possibilities for lncRNA three-dimensional structure

In the case of lncRNAs, many interesting questions are yet to be answered. For example, should we expect to encounter structural motifs already met in known RNA systems? What are the structural differences at the RNA level across various organisms? In light of the rapid turnover of lncRNAs, do we expect the evolution of unique lineage-specific structural elements?

We currently lack information on the tertiary structure of lncRNAs. In addition, it is not known if lncRNAs exist in ribonucleoprotein complexes (RNPs) or predominantly as isolated RNAs. With regard to tertiary structure, the ribosome is an interesting system for comparison since the ribosome is the only RNA system >1 kb whose crystallographic structure has been solved. In terms of the composition of a typical lncRNA complex, the total number of unique lncRNAs and unique proteins is a useful constraint. As mentioned above, the number of human lncRNAs has been estimated to be ~15,000,[5] while the number of protein-coding genes is ~21,000.[122] While an lncRNA-based ribosome-like RNP complex could exist, it is unlikely that most lncRNAs are ribosome-like RNPs, given the limited number of unique protein-coding genes relative to the number of unique lncRNA transcripts. Ruling out ribosome-like RNA complexes, the following possibilities remain for the composition of lncRNA complexes: (1) RNP complexes with many repeats of a few proteins, (2) RNP complexes with only a few proteins or (3) isolated RNAs that transiently bind proteins as needed for function. For scenario (1), ~10 protein copies per 1 kb of lncRNA would be required to produce complexes with a similar protein-to-RNA ratio as the ribosome (e.g., 20 protein copies for HOTAIR).

In scenarios (2) and (3), lncRNA complexes would be similar in composition to RNase P, telomerase RNA or the group I/II introns. An "RNase-P-like" lncRNA complex would contain a highly structured and compact RNA core with a single protein binding

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

Q2

site (Fig. 5a). Alternatively, the RNA could be structurally decentralized without a compact core, containing several distinct protein binding sites (Fig. 5c). The lncRNA may also act as a flexible structural scaffold, as suggested for the telomerase RNA (Fig. 5b). In loose terms, such a flexible RNA could be analogous to the cellulosome, where many cellulase subunits are connected by long disordered linker proteins (Fig. 2d). The cellulase subunits do not act coherently but allow the cellulosome to efficiently process cellulose. Finally, the lncRNA could exist predominantly as an isolated, highly structured RNA, similar to the self-splicing introns. Here, the RNA may transiently bind proteins that each perform a specific function.

Supramolecular lncRNA complexes

The large size of certain lncRNA systems has produced new kinds of RNA complexes on a scale not previously studied in the context of structural biology. In X chromosome inactivation, the human XIST is responsible for gene silencing on the inactive X chromosome (Xi). At 17 kb, XIST is one of the largest lncRNAs. A large number of XIST copies are transcribed to physically coat the X chromosome.

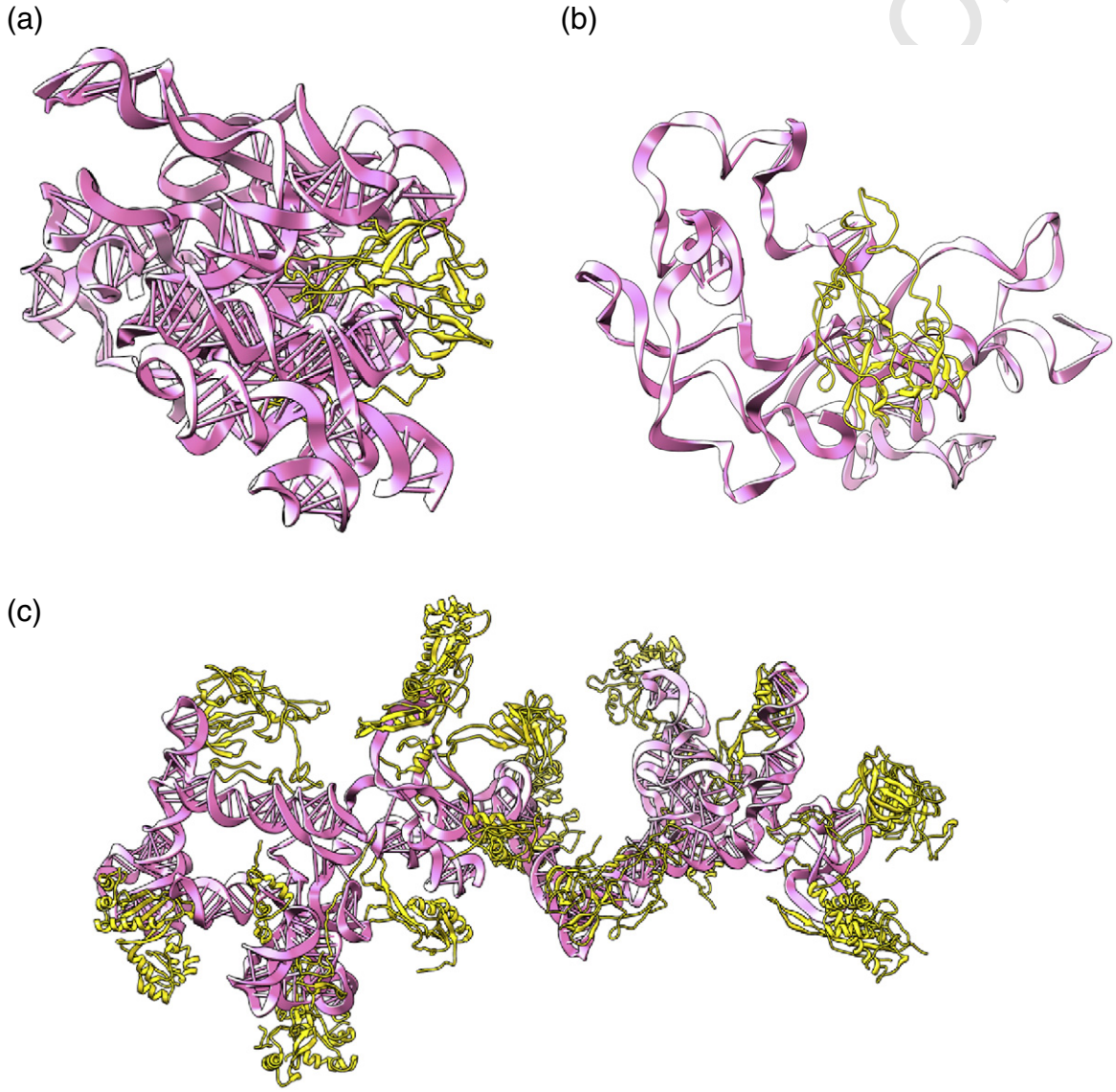

Fig. 5. Homology models representing potential paradigms for lncRNA tertiary structure. (a) Compact core. The lncRNA (pink) is highly structured and has a protein (yellow) binding site that may bind different proteins. (b) Loosely organized protein binding domain with relatively unstructured RNA. (c) Decentralized. The lncRNA possesses several protein (yellow) binding sites but no compact core.

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

FISH experiments suggest that XIST RNAs not only interact with chromatin but also crosstalk to produce a very large structural architecture on the scale of the entire X chromosome.[123–125] The kcnq1ot1 lncRNA (108 kb) is also thought to coat chromosomes.[16,18]

Other lncRNAs participate in subcellular structure formation, including NEAT1 (paraspeckle formation)[27,37,38,126] and MALAT1/NEAT2 (nuclear speckle formation).[28,36,33,127,128] NEAT1 non-coding RNA has two isoforms: a short NEAT1_V1 (3.7 kb) and a long NEAT1_V2 (22.7 kb). While the shorter transcript aids in the paraspeckle assembly, a longer transcript NEAT1_V2 acts as an essential scaffold, creating ~~forming~~ a structural network with sausage-like morphology.[129,130] To date, ~~we identify~~ up to 35 paraspeckle-associated proteins, which range from the splicing factors, the 3′ end processing enzymes and the disease-related proteins.[129]

From structure to mechanism: lncRNA dynamics

The dynamics of lncRNAs must be studied to fully understand their mechanisms. Timescales, in particular, have been shown to play a critical role in understanding the mechanism of other molecular machines. In the ribosome system, rapid kinetics studies placed many useful constraints on the order of events in tRNA selection and translocation prior to the solution of X-ray structures of the ribosome.[131,132] With structural data in hand, single-molecule studies have identified more sub-steps of these processes, improved our understanding of transitions between states and produced a new framework for mechanism.[73]

As lncRNAs have only recently emerged as a class of RNAs, very little is known about the relevant timescales of lncRNA function. One recently discovered lncRNA, DBE-T, is an illustrative example with respect to processes that can occur in lncRNA systems.[40] Here, the DBE-T lncRNA is an important part of the epigenetic switch associated with facioscapulohumeral muscular dystrophy. DBE-T is *cis*-acting RNA and tethers epigenetic factors to the D4Z4 binding element (DBE). As a result, histone methylation occurs. Here, timescales related to transcription, folding, protein binding to the RNA, protein binding to the chromatin and histone methylation should be studied to determine the rate-limiting step.

Summary

LncRNAs have emerged as a new class of RNAs, playing important roles in development, stem cells, cancer, brain disease and epigenetic mechanism. LncRNA mechanism may be based on sequence, secondary structure, tertiary structure or a new combination of these mechanisms. In light of the high degree of diversity among lncRNAs, it is possible that the full mechanistic arsenal of previously studied RNAs may be employed in various lncRNA systems. Because we are in the early stages of lncRNA research, it is not clear if majority of these systems provide scaffolding for gene regulation complexes or engage in more active roles, such as catalysis, molecular switching or information processing. With the identification of many thousands of lncRNAs in recent years, it is clear that new structural studies will play a key role in demystifying these strange, new RNA machines.

Acknowledgements

The work was performed under the auspices of the US Department of Energy via Los Alamos National Laboratory, Laboratory Directed Research and Development.

Received 20 December 2012;
Received in revised form 21 February 2013;
Accepted 25 February 2013

Keywords:
long non-coding RNA;
lncRNA;
lincRNA;
lncRNA structure;
epigenetics

Abbreviations used:
lncRNA, long non-coding RNA; SRA, steroid receptor RNA activator; SHAPE, selective 2′-hydroxyl acylation analyzed by primer extension; LSD, lysine-specific demethylase; RNPribonucleoprotein complex.

References

1. Taft, R. J., Pheasant, M. & Mattick, J. S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays*, **29**, 288–299.
2. Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A. *et al.* (2012). Landscape of transcription in human cells. *Nature*, **489**, 101–108.
3. Mehler, M. F. & Mattick, J. S. (2007). Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiol. Rev.* **87**, 799–823.
4. Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T. *et al.* (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, **316**, 1484–1488.
5. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H. *et al.* (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789.

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

Q2

6. Rinn, J. L. & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166.
7. Ziats, M. N. & Rennert, O. M. (2012). Aberrant expression of long noncoding RNAs in autistic brain. *J. Mol. Neurosci.* **49**, 589–593.
8. Brunner, A. L., Beck, A. H., Edris, B., Sweeney, R. T., Zhu, S. X., Li, R. *et al.* (2012). Transcriptional profiling of lncRNAs and novel transcribed regions across a diverse panel of archived human cancers. *Genome Biol.* **13**, R75.
9. Sati, S., Ghosh, S., Jain, V., Scaria, V. & Sengupta, S. (2012). Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci. *Nucleic Acids Res.* **40**, 10018–10031.
10. Zhang, X., Sun, S., Pu, J. K., Tsang, A. C., Lee, D., Man, V. O. *et al.* (2012). Long non-coding RNA expression profiles predict clinical phenotypes in glioma. *Neurobiol. Dis.* **48**, 1–8.
11. Banfai, B., Jia, H., Khatun, J., Wood, E., Risk, B., Gundling, W. E., Jr *et al.* (2012). Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* **22**, 1646–1657.
12. Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J. *et al.* (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, **464**, 1071–1076.
13. Gong, C. & Maquat, L. E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature*, **470**, 284–288.
14. Ponting, C. P., Oliver, P. L. & Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell*, **136**, 629–641.
15. Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D. *et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223–227.
16. Korostowski, L., Sedlak, N. & Engel, N. (2012). The Kcnq1ot1 long non-coding RNA affects chromatin conformation and expression of Kcnq1, but does not regulate its imprinting in the developing heart. *PLoS Genet.* **8**, e1002956.
17. Sleutels, F., Zwart, R. & Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, **415**, 810–813.
18. Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J. *et al.* (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell*, **32**, 232–246.
19. Vallot, C., Huret, C., Lesecque, Y., Resch, A., Oudrhiri, N., Bennaceur-Griscelli, A. *et al.* (2013). XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat. Genet.* **45**, 239–241.
20. Novikova, I. V., Hennelly, S. P. & Sanbonmatsu, K. Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.* **40**, 5034–5051.
21. Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signaling*, **3**, ra8.
22. Lin, D., Pestova, T. V., Hellen, C. U. & Tiedge, H. (2008). Translational control by a small RNA: dendritic BC1 RNA targets the eukaryotic initiation factor 4A helicase mechanism. *Mol. Cell. Biol.* **28**, 3008–3019.
23. Pauli, A., Rinn, J. L. & Schier, A. F. (2011). Non-coding RNAs as regulators of embryogenesis. *Nat. Rev., Genet.* **12**, 136–149.

Huarte, M. & Rinn, J. L. (2010). Large non-coding RNAs: missing links in cancer? *Hum. Mol. Genet.* **19**, R152–R161.

25. Tsai, M. C., Spitale, R. C. & Chang, H. Y. (2011). Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res.* **71**, 3–7.
26. Wilusz, J. E., Freier, S. M. & Spector, D. L. (2008). 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell*, **135**, 919–932.
27. Sunwoo, H., Dinger, M. E., Wilusz, J. E., Amaral, P. P., Mattick, J. S. & Spector, D. L. (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* **19**, 347–359.
28. Tano, K. & Akimitsu, N. (2012). Long non-coding RNAs in cancer progression. *Front. Genet.* **3**, 219.
29. Tani, H., Nakamura, Y., Ijiri, K. & Akimitsu, N. (2010). Stability of MALAT-1, a nuclear long non-coding RNA in mammalian cells, varies in various cancer cells. *Drug Discoveries Ther.* **4**, 235–239.
30. Dinger, M. E., Amaral, P. P., Mercer, T. R., Pang, K. C., Bruce, S. J., Gardiner, B. B. *et al.* (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18**, 1433–1445.
31. Sone, M., Hayashi, T., Tarui, H., Agata, K., Takeichi, M. & Nakagawa, S. (2007). The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *J. Cell Sci.* **120**, 2498–2506.
32. Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. (2008). Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. USA*, **105**, 716–721.
33. Bernard, D., Prasanth, K. V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z. *et al.* (2010). A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.* **29**, 3082–3093.
34. Dharap, A., Nakka, V. P. & Vemuganti, R. (2012). Effect of focal ischemia on long noncoding RNAs. *Stroke*, **43**, 2800–2802.
35. Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A. & Lawrence, J. B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell*, **33**, 717–726.
36. Tano, K., Mizuno, R., Okada, T., Rakwal, R., Shibato, J., Masuo, Y. *et al.* (2010). MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Lett.* **584**, 4575–4580.
37. Mao, Y. S., Sunwoo, H., Zhang, B. & Spector, D. L. (2011). Direct visualization of the co-transcriptional

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95–101.
38. Ip, J. Y. & Nakagawa, S. (2011). Long non-coding RNAs in nuclear bodies. *Dev., Growth Differ.* [Epub ahead of print].
39. Ietswaart, R., Wu, Z. & Dean, C. (2012). Flowering time control: another window to the connection between antisense RNA and chromatin. *Trends Genet.* **28**, 445–453.
40. Cabianca, D. S., Casa, V., Bodega, B., Xynos, A., Ginelli, E., Tanaka, Y. & Gabellini, D. (2012). A long ncRNA links copy number variation to a polycomb/-trithorax epigenetic switch in FSHD muscular dystrophy. *Cell*, **149**, 819–831.
41. 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.
42. Song, J., Angel, A., Howard, M. & Dean, C. (2012). Vernalization—a cold-induced epigenetic switch. *J. Cell Sci.* **125**, 3723–3731.
43. Mizutani, R., Wakamatsu, A., Tanaka, N., Yoshida, H., Tochigi, N., Suzuki, Y. *et al.* (2012). Identification and characterization of novel genotoxic stress-inducible nuclear long noncoding RNAs in mammalian cells. *PLoS One*, **7**, e34949.
44. Frank, J., Jr & Gonzalez, R. L. (2010). Structure and dynamics of a processive Brownian motor: the translating ribosome. *Annu. Rev. Biochem.* **79**, 381–412.
45. Sanbonmatsu, K. Y. (2012). Computational studies of molecular machines: the ribosome. *Curr. Opin. Struct. Biol.* **22**, 168–174.
46. Gavrilova, L. P., Kostiashkina, O. E., Koteliansky, V. E., Rutkevitch, N. M. & Spirin, A. S. (1976). Factor-free ("non-enzymic") and factor-dependent systems of translation of polyuridylic acid by *Escherichia coli* ribosomes. *J. Mol. Biol.* **101**, 537–552.
47. Kitao, A., Yonekura, K., Maki-Yonekura, S., Samatey, F. A., Imada, K., Namba, K. & Go, N. (2006). Switch interactions control energy frustration and multiple flagellar filament structures. *Proc. Natl. Acad. Sci. USA*, **103**, 4894–4899.
48. Chan, R. T., Robart, A. R., Rajashankar, K. R., Pyle, A. M. & Toor, N. (2012). Crystal structure of a group II intron in the pre-catalytic state. *Nat. Struct. Mol. Biol.* **19**, 555–557.
49. Pyle, A. M. (2010). The tertiary structure of group II introns: implications for biological function and evolution. *Crit. Rev. Biochem. Mol. Biol.* **45**, 215–232.
50. Toor, N., Keating, K. S. & Pyle, A. M. (2009). Structural insights into RNA splicing. *Curr. Opin. Struct. Biol.* **19**, 260–266.
51. Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T. & Mondragon, A. (2010). Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*, **468**, 784–789.
52. Mitchell, M., Gillis, A., Futahashi, M., Fujiwara, H. & Skordalakes, E. (2010). Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat. Struct. Mol. Biol.* **17**, 513–518.
53. Zappulla, D. C. & Cech, T. R. (2004). Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc. Natl. Acad. Sci. USA*, **101**, 10024–10029.
54. Batey, R. T., Gilbert, S. D. & Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, **432**, 411–415.
55. Montange, R. K. & Batey, R. T. (2008). Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.* **37**, 117–133.
56. Winkler, W., Nahvi, A. & Breaker, R. R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952–956.
57. Tucker, B. J. & Breaker, R. R. (2005). Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.* **15**, 342–348.
58. Hennelly, S. P. & Sanbonmatsu, K. Y. (2011). Tertiary contacts control switching of the SAM-I riboswitch. *Nucleic Acids Res.* **39**, 2416–2431.
59. Stoddard, C. D., Montange, R. K., Hennelly, S. P., Rambo, R. P., Sanbonmatsu, K. Y. & Batey, R. T. (2010). Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure*, **18**, 787–797.
60. Hayes, R. L., Noel, J. K., Mohanty, U., Whitford, P. C., Hennelly, S. P., Onuchic, J. N. & Sanbonmatsu, K. Y. (2012). Magnesium fluctuations modulate RNA dynamics in the SAM-I riboswitch. *J. Am. Chem. Soc.* **134**, 12043–12053.
61. Penchovsky, R. & Breaker, R. R. (2005). Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes. *Nat. Biotechnol.* **23**, 1424–1433.
62. Guttman, M. & Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature*, **482**, 339–346.
63. Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R. *et al.* (2012). Genome-wide measurement of RNA folding energies. *Mol. Cell*, **48**, 169–181.
64. Appelt, K., Dijk, J., Reinhardt, R., Sanhuesa, S., White, S. W., Wilson, K. S. & Yonath, A. (1981). The crystallization of ribosomal proteins from the 50S subunit of the *Escherichia coli* and *Bacillus stearothermophilus* ribosome. *J. Biol. Chem.* **256**, 11787–11790.
65. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M. & Cate, J. H. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science*, **310**, 827–834.
66. Pyle, A. M., Fedorova, O. & Waldsich, C. (2007). Folding of group II introns: a model system for large, multidomain RNAs? *Trends Biochem. Sci.* **32**, 138–145.
67. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J. & Strobel, S. A. (2004). Crystal structure of a self-splicing group I intron with both exons. *Nature*, **430**, 45–50.
68. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. (2008). Crystal structure of a self-spliced group II intron. *Science*, **320**, 77–82.
69. Marcia, M. & Pyle, A. M. (2012). Visualizing group II intron catalysis through the stages of splicing. *Cell*, **151**, 497–507.
70. Gao, Y. G., Selmer, M., Dunham, C. M., Weixlbaumer, A., Kelley, A. C. & Ramakrishnan, V. (2009). The

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

Q2

structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science*, **326**, 694–699.

71. Rodnina, M. V., Beringer, M. & Wintermeyer, W. (2006). Mechanism of peptide bond formation on the ribosome. *Q. Rev. Biophys.* **39**, 203–225.
72. Whitford, P. C., Ahmed, A., Yu, Y., Hennelly, S. P., Tama, F., Spahn, C. M. *et al.* (2011). Excited states of ribosome translocation revealed through integrative molecular modeling. *Proc. Natl Acad. Sci. USA*, **108**, 18943–18948.
73. Blanchard, S. C. (2009). Single-molecule observations of ribosome function. *Curr. Opin. Struct. Biol.* **19**, 103–109.
74. Reiter, N. J., Chan, C. W. & Mondragon, A. (2011). Emerging structural themes in large RNA molecules. *Curr. Opin. Struct. Biol.* **21**, 319–326.
75. Cole, D. I., Legassie, J. D., Bonifacio, L. N., Sekaran, V. G., Ding, F., Dokholyan, N. V. & Jarstfer, M. B. (2012). New models of tetrahymena telomerase RNA from experimentally derived constraints and modeling. *J. Am. Chem. Soc.* **134**, 20070–20080.
76. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, **318**, 798–801.
77. Schoeftner, S. & Blasco, M. A. (2008). Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell Biol.* **10**, 228–236.
78. Lau, N., Lim, L., Weinstein, E. & Bartel, D. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
79. Reinhart, B., Weinstein, E., Rhoades, M., Bartel, B. & Bartel, D. (2002). MicroRNAs in plants. *Genes Dev.* **16**, 1616–1626.
80. Lewis, B., Shih, I., Jones-Rhoades, M., Bartel, D. & Burge, C. (2003). Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.
81. Tang, G., Reinhart, B., Bartel, D. & Zamore, P. (2003). A biochemical framework for RNA silencing in plants. *Genes Dev.* **17**, 49–63.
82. Zamore, P. (2002). Ancient pathways programmed by small RNAs. *Science*, **296**, 1265–1269.
83. Zamore, P. D. & Haley, B. (2005). Ribo-gnome: the big world of small RNAs. *Science*, **309**, 1519–1524.
84. Ghildiyal, M. & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nat. Rev., Genet.* **10**, 94–108.
85. Tomari, Y. & Zamore, P. D. (2005). Perspective: machines for RNAi. *Genes Dev.* **19**, 517–529.
86. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*, **445**, 666–670.
87. Bertani, S., Sauer, S., Bolotin, E. & Sauer, F. (2011). The noncoding RNA mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol. Cell*, **43**, 1040–1046.
88. Beltran, M., Puig, I., Pena, C., Garcia, J. M., Alvarez, A. B., Pena, R. *et al.* (2008). A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial–mesenchymal transition. *Genes Dev.* **22**, 756–769.
89. Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M. *et al.* (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*, **147**, 358–369.
90. Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. (2002). An mRNA structure that controls gene expression by binding FMN. *Proc. Natl Acad. Sci. USA*, **99**, 15908–15913.
91. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. (2003). An mRNA structure that controls gene expression by binding *S*-adenosylmethionine. *Nat. Struct. Biol.* **10**, 701–707.
92. Montange, R. K. & Batey, R. T. (2006). Structure of the *S*-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, **441**, 1172–1175.
93. Gilbert, S. D., Rambo, R. P., Van Tyne, D. & Batey, R. T. (2008). Structure of the SAM-II riboswitch bound to *S*-adenosylmethionine. *Nat. Struct. Mol. Biol.* **15**, 177–182.
94. Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. (2009). Accurate SHAPE-directed RNA structure determination. *Proc. Natl Acad. Sci. USA*, **106**, 97–102.
95. Hamada, M., Sato, K. & Asai, K. (2010). Prediction of RNA secondary structure by maximizing pseudo-expected accuracy. *BMC Bioinformatics*, **11**, 586.
96. Hamada, M., Sato, K. & Asai, K. (2011). Improving the accuracy of predicting secondary structure for aligned RNA sequences. *Nucleic Acids Res.* **39**, 393–402.
97. Rivas, E., Lang, R. & Eddy, S. R. (2012). A range of complex probabilistic models for RNA secondary structure prediction that includes the nearest-neighbor model and more. *RNA*, **18**, 193–212.
98. Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R. & Stadler, P. F. (2008). RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics*, **9**, 474.
99. Ghosh, S. K., Patton, J. R. & Spanjaard, R. A. (2012). A small RNA derived from RNA coactivator SRA blocks steroid receptor signaling via inhibition of Pus1p-mediated pseudouridylation of SRA: evidence of a novel RNA binding domain in the N-terminus of steroid receptors. *Biochemistry*, **51**, 8163–8172.
100. Xu, B., Yang, W. H., Gerin, I., Hu, C. D., Hammer, G. D. & Koenig, R. J. (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.
101. Colley, S. M. & Leedman, P. J. (2009). SRA and its binding partners: an expanding role for RNA-binding coregulators in nuclear receptor-mediated gene regulation. *Crit. Rev. Biochem. Mol. Biol.* **44**, 25–33.
102. Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R. D. & 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.
103. Woese, C. R., Magrum, L. J., Gupta, R., Siegel, R. B., Stahl, D. A., Kop, J. *et al.* (1980). Secondary structure

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030

model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res.* **8**, 2275–2293.
104. Noller, H. F. & Woese, C. R. (1981). Secondary structure of 16S ribosomal RNA. *Science*, **212**, 403–411.
105. Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G. & Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, **334**, 1524–1529.
106. Legiewicz, M., Badorrek, C. S., Turner, K. B., Fabris, D., Hamm, T. E., Rekosh, D. *et al.* (2008). Resistance to RevM10 inhibition reflects a conformational switch in the HIV-1 Rev response element. *Proc. Natl Acad. Sci. USA*, **105**, 14365–14370.
107. Legiewicz, M., Zolotukhin, A. S., Pilkington, G. R., Purzycka, K. J., Mitchell, M., Uranishi, H. *et al.* (2010). The RNA transport element of the murine musD retrotransposon requires long-range intramolecular interactions for function. *J. Biol. Chem.* **285**, 42097–42104.
108. Purzycka, K. J., Legiewicz, M., Matsuda, E., Eizentstat, L. D., Lusvarghi, S., Saha, A. *et al.* (2013). Exploring Ty1 retrotransposon RNA structure within virus-like particles. *Nucleic Acids Res.* **41**, 463–473.
109. Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr, Swanstrom, R. *et al.* (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.
110. Lucks, J. B., Mortimer, S. A., Trapnell, C., Luo, S., Aviran, S., Schroth, G. P. *et al.* (2011). Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl Acad. Sci. USA*, **108**, 11063–11068.
111. Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T. & Chang, H. Y. (2013). RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **9**, 18–20.
112. Wilusz, J. E. & Spector, D. L. (2010). An unexpected ending: noncanonical 3′ end processing mechanisms. *RNA*, **16**, 259–266.
113. Beniaminov, A., Westhof, E. & Krol, A. (2008). Distinctive structures between chimpanzee and human in a brain noncoding RNA. *RNA*, **14**, 1270–1275.
114. Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M. A., Coppens, S., Pedersen, J. S. *et al.* (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, **443**, 167–172.
115. Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S. *et al.* (2006). Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* **2**, e168.
116. Guttman, M., Donaghey, J., Carey, B. W., Garber, M., Grenier, J. K., Munson, G. *et al.* (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, **477**, 295–300.
117. Sanford, J. R., Wang, X., Mort, M., Vanduyn, N., Cooper, D. N., Mooney, S. D. *et al.* (2009). Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* **19**, 381–394.
118. Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K. *et al.* (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell*, **40**, 939–953.
119. Surface, L. E., Thornton, S. R. & Boyer, L. A. (2010). Polycomb group proteins set the stage for early lineage commitment. *Cell Stem Cell*, **7**, 288–298.
120. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, **322**, 750–756.
121. Wutz, A., Rasmussen, T. P. & Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* **30**, 167–174.
122. Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C. & Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
123. Jeon, Y., Sarma, K. & Lee, J. T. (2012). New and Xisting regulatory mechanisms of X chromosome inactivation. *Curr. Opin. Genet. Dev.* **22**, 62–71.
124. Lee, J. T. (2011). Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nat. Rev., Mol. Cell Biol.* **12**, 815–826.
125. Kim, D. H., Jeon, Y., Anguera, M. C. & Lee, J. T. (2011). X-chromosome epigenetic reprogramming in pluripotent stem cells via noncoding genes. *Semin. Cell Dev. Biol.* **22**, 336–342.
126. Sasaki, Y. T., 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.
127. Gutschner, T., Hammerle, M., Eissmann, M., Hsu, J., Kim, Y., Hung, G. *et al.* (2012). The non-coding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* **73**, 1180–1189.
128. Miyagawa, R., Tano, K., Mizuno, R., Nakamura, Y., Ijiri, K., Rakwal, R. *et al.* (2012). Identification of *cis*- and *trans*-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA*, **18**, 738–751.
129. Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y. F., Goshima, N. & Hirose, T. (2012). Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J.* **31**, 4020–4034.
130. Souquere, S., Beauclair, G., Harper, F., Fox, A. & Pierron, G. (2010). Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies. *Mol. Biol. Cell*, **21**, 4020–4027.
131. Rodnina, M. V. & Wintermeyer, W. (2001). Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. *Annu. Rev. Biochem.* **70**, 415–435.
132. Rodnina, M. V. & Wintermeyer, W. (2001). Ribosome fidelity: tRNA discrimination, proofreading and induced fit. *Trends Biochem. Sci.* **26**, 124–130.

Please cite this article as: Novikova, I. *et al.*, Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs, *J. Mol. Biol.* (2013), http://dx.doi.org/10.1016/j.jmb.2013.02.030