Abstract

A Pipeline for Discovery and Determination of Functional Tertiary Structures in Large RNAs

Shivali Patel

2022

RNAs can adopt complex folded motifs and higher-order three-dimensional (3-D) structures that play a key role in various cellular processes. It has recently become clear that large RNAs (>200 nucleotides in length) contain regions of higher-order structures; however, identifying them remains challenging. To unambiguously elucidate their mechanism and effectively target these RNAs, we need to understand the interplay between their higher-order structure and biological function. In this dissertation, I describe my efforts towards developing techniques to discover and visualize RNA structure in diverse systems ranging from ribozymes to long non-coding RNAs (lncRNAs) and viral RNA genomes.

First, I developed a high-throughput sequencing-based approach to systemically identify compact tertiary structures using the lanthanide metal terbium (Tb^{3+}). With this approach (Tb-seq), I scanned for stable structural modules and potential riboregulatory motifs in self-splicing group II introns, human RNase P, HCV genomic RNA, and SARS-CoV-2 genomic RNA. Tb-seq reveals sites in which RNA forms sharp, stable turn motifs that result in complex arrangements of nucleotides in 3-D space, thereby providing a

powerful new approach for pinpointing regions of complex RNA structure that are potentially associated with RNA functional elements.

Next, to directly visualize the molecular level details of these discovered tertiary structures, I used electron cryo-microscopy (Cryo-EM). I studied the molecular architecture of the self-splicing group IIB intron aI5γ from *Saccharomyces cerevisiae*. This intron has undergone extensive biochemical characterization over the last several decades revealing, it as a model system for understanding RNA folding and catalysis. However, its sub-nanometer resolution structure remains unknown. Here, I report a 3.6 Å structure of this ribozyme, observing for the first time the intricate structural interplay and extensive network of tertiary interactions that regulate the tight folding and catalytic activity. Given its broad applicability, the cryo-EM pipeline can be readily adapted to understand the molecular basis in a diverse set of large RNA systems.

Finally, I integrated these two approaches as a systematic workflow to thoroughly characterize the structural architecture of the lncRNA Pnky, known to play a key role in neurogenesis. I discovered that Pnky exhibits a high degree of structural organization, including an RNA pseudoknot and several long-range tertiary interaction sites. Preliminary cryo-EM demonstrates that Pnky adopts a compact and globular architecture to regulate neuronal stem cell differentiation. These findings provide a roadmap of candidate riboregulatory motifs for targeted functional investigation. Taken together, the approaches described in this thesis will serve as a systematic pipeline for discovering and visualizing higher-order structures in complex RNA systems.

A Pipeline for Discovery and Determination of Functional Tertiary Structures in Large RNAs

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
In Candidacy for the Degree of
Doctor of Philosophy

by
Shivali Patel

Dissertation Director: Dr. Anna Marie Pyle

December 2022

Acknowledgments

I want to thank all those who have supported me in my scientific journey. I want to offer my deepest gratitude to my advisor, Anna Pyle, for creating an environment where I was given the independence to pursue my scientific interests and collaborate with a fantastic group of people inside and outside the lab. I would also like to thank my committee members, Wendy Gilbert and Matt Simon, for their valuable feedback and support throughout the years. I want to thank all my collaborators for their support and acting as a soundboard for all my ideas.

I express my sincerest appreciation for past and present members of the Pyle lab. Since day one, everyone has made the lab a nourishing and positive work environment. We are constantly supporting and challenging each other to develop into great scientists. I want to especially thank Rafael Tavares, Ananth Kumar, Olga Fedorova, Becky Adams, and Wenshuai Wang for all their mentorship throughout the years. I would also like to thank my undergraduate mentor Sasha Serganov for all his guidance and training. I want to thank Alla Peselis, Jake Weaver, and Nikita Vasilyev for taking me in as an undergraduate and laying the foundation for me to develop into a scientist.

To all the amazing friendships, old and new, who have supported me throughout the ups and downs of graduate school. My scientific journey would not have been possible without all of you there. Finally, I want to thank my family for their love and encouragement to pursue my passion throughout the years. I especially want to thank my brother, Shivam Patel, for always being there every step of the way.

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

1. Introduction

1.1 Versatility of RNA

The central dogma of molecular biology states that the role of nucleic acids is to store biological information that will be translated into proteins which will carry out cellular functions.[1] Therefore, research efforts have primarily focused on characterizing the importance of protein structure and function. It is becoming increasingly clear that RNA plays critical biological roles through its primary sequence, secondary or higher order structure. RNA can act as a catalyst,[2] regulate gene expression via *cis* or *trans* mechanisms,[3,4] and encode signals for localization or decay.[5] RNA can even interact with a host of cellular factors, such as proteins, to form dynamic complexes that undergo multiple conformational rearrangements to translate mRNA,[6] catalyze intronic removal through splicing pathways,[7] or initiate the excision of 5' end of tRNAs.[8] The versatility of these RNA molecules underscores the need to closely examine their structure to comprehensively understand their function.

1.2 What is RNA structure?

Classically, RNA was thought to be a single-stranded entity whose function was defined by its primary sequence, such as an mRNA encoding for translating proteins or binding by micro RNAs to regulate RNA expression.[9] In reality, the ability of RNAs to fold back on themselves and form stable structures in concert with cellular factors underlies their complex functional roles.[10] On a two-dimensional level, RNAs can form canonical Watson-Crick base paring or non-conical Hoogsteen/wobble and sugar-edge pairing

resulting in stem-loop and hairpin-like structures. These structures can affect mRNA stability by influencing the speed of translating ribosomes.[11] In some systems, the ability of these structures to rapidly change conformations in response to a ligand,[12] temperature,[13] or ion osmolarity[14] can finely tune gene expression.

The ability of these single-stranded and double-stranded regions to come together in three-dimensional space and stabilize the RNA's electrostatic negative regions through intra and inter-molecular interactions forms the basis of the complex RNA folding process. RNA helices can form long-range contacts to orientate and stabilize an active site.[7] They can position themselves to form platforms for the recruitment of proteins[15] or form pockets for binding to small molecules.[16] Even single-stranded regions are not always flexible strings of nucleotides.[10] Oftentimes, in bulge or terminal stem loops, nucleotides can form stabilizing base stacking interactions or long-range base-pairing interactions. RNAs can even remodel their 3-D structure when bound to cellular proteins,[8] ligands,[17] or as it progresses through step-wise processes.[18]

1.3 Techniques to assess RNA structure

Throughout the years, substantial efforts have been made to improve our understanding of RNA secondary and tertiary structure. Here, I briefly outline some of the techniques.

1.3.1 Probing RNA secondary structures

To identify the likeliness of a nucleotide to be single- versus double-stranded, classical efforts employed the use of enzymatic nucleases that would specifically cleave single-stranded or double-stranded regions in an RNA.[19,20] In recent years, small molecule chemical probes that modify nucleotides have emerged as powerful tools to assess the secondary structure.[21-23] Coupling the chemical probing readout with reverse transcription and high-throughput sequencing technologies has expanded the characterization of the structural landscape in whole genomes. Furthermore, the ability of these probes to permeate cell membranes and modify RNAs within the cellular context has transformed our understanding of functional RNA structures.[22,24,25] By comparing the probing reactivities between *in-vitro* and *in-cellulo* systems, one can assess which regions of RNA are likely recognized by cellular factors and can influence function.[26]

1.3.2 Tertiary structure probes

There are many hallmarks to assess the tertiary structure of an RNA. One of the most fundamental manipulations is the role of metal ions, specifically magnesium, in the formation of RNA tertiary structure.[27] Determining where these metals bind can be done using lanthanide metals such as terbium (Tb^{3+})[28] and europium (Eu^{3+})[29] or "in-line" probing

techniques.[30] Both utilize the inherent property of these metals to deprotonate the 2'OH, cleave the adjacent phosphodiester bond[31] and elucidate structural motifs in the RNA.

A second approach to study higher-order RNA structure is by investigating solvent accessibility of nucleotides during a folding pathway or binding to a small ligand.[32-35] Furthermore, the application of this technique to *in-cell* systems has enhanced our understanding of structural changes during RNA-protein complex formation and processing.[36] Lastly, the development of cross-linking and proximity ligation approaches to study long-range RNA-RNA base pairing have revealed higher-order organization in multi-kilobase RNAs.[37,38]

1.3.3 *In-silico* predictions

Throughout the years, a host of *in-silico* algorithms have emerged to predict macromolecular structure. Many of these algorithms use minimum free energy estimations[39,40] to determine the likelihood of an individual structural confirmation. More recently, there has been a tremendous surge in machine learning[41,42] to predict RNA structures de-novo. While these programs provide an initial start for small RNAs (<500 nucleotides), their prediction accuracy is reduced in large RNA systems. Moreover, such programs are unable to account for how proteins or other nucleic acid interactions influence the folding of RNAs.

Simultaneously, the development of computational pipelines to calculate reactivities from small molecule chemical probes and use them as constraints has increased the accuracy of secondary structure prediction models. With recent expansion into mapping tertiary contacts through statistical correlations in mutational profiling[43,44] or pseudoknot

prediction programs, we are beginning to identify tertiary structure elements in large RNAs. Ultimately integrating these approaches with probing data, non-conical base pairing, and protein occupancy will improve and expand our arsenal of tertiary structure prediction methods. It is worth mentioning that, despite the recent revolution in 3-D protein structure prediction,[45] there remains a long way to go for similar accurate RNA and RNP structure prediction.

1.3.4 Biophysical techniques

Biophysical techniques such as X-ray crystallography and NMR have revolutionized our atomic-level understanding of how RNA can adopt complex folds to function.[46,47] While crystallography works well with tightly compact and structured RNAs that can be purified to milligram amounts, there are serious limitations if such high amounts of starting materials cannot be isolated or synthesized. Additionally, the presence of flexible or dynamic regions in an RNA can prevent the formation of ordered crystalline packing. NMR is a versatile technique to study the dynamics of RNA folding pathways,[48] but current technical limitations make it applicable to only small RNAs.

Complementary techniques such as sedimentation velocity analytical ultracentrifugation (SV-AUC), small-angle x-ray scattering (SAXS), size-exclusion chromatography with multi-angle light scattering (SEC-MALS), and atomic force microscopy (AFM) have been successfully applied to access large RNA's globally shape and conformational states,[49,50] however, these techniques lack the achievable resolution needed for a molecular level assessment.

In recent years, advances in microscopy and computational tools have revolutionized electron cryo-microscopy (cryo-EM), resulting in a mechanistic understanding of complex and diverse systems.[51] With the reduced concentration of starting material and downstream data processing software that can deconvolute dynamic functional states of large complexes,[52] it has increasingly become a valuable tool for elucidating details at the molecular level. Nevertheless, its application to study RNA-only systems is limited by serious bottlenecks in the sample preparation process, underscoring the need to develop an effective pipeline for high-resolution structure determination and transform RNA structural biology.

1.4 Dissertation Summary

Sequencing of the human genome has revealed a vast expanse of multi-kilobase RNAs present in cells whose functional role is largely still not understood.[53] These RNAs likely harbor structural elements that control and fine-tune their diverse function. This underscores the need to expand the RNA toolbox to adapt and develop approaches to identify functional riboregulatory elements. This dissertation will focus on several frontiers of RNA structural biology and the development of new approaches to study RNA tertiary structure.

High-resolution structures reveal that RNAs can adopt specific 3-D shapes by compressing their RNA backbones together, bringing negatively charged phosphates in close proximity. This is often accomplished by recruiting multivalent cations (usually Mg^{2+}), which stabilize these sites by neutralizing regions of local negative charge. One way to assess these electrostatic negative reservoirs is by the lanthanide metal terbium (III) (Tb^{3+}). Coordinated Tb^{3+} can substitute for Mg^{2+} at these sites and rapidly cleave the RNA backbone, revealing the presence of these compact tertiary modules. While Tb^{3+} has been used in the past, these cleavage sites were monitored via low-throughput biochemical methods only applicable to small RNAs. In **Chapter 1,** I present a high-throughput sequencing method for detecting compact tertiary structures in large RNAs, termed Tb-seq. Tb-seq reveals sites in which RNA forms sharp, stable turn motifs that result in complex arrangements of nucleotides in 3-D space and thus provide a powerful new approach for pinpointing regions of complex RNA structure that are potentially associated with RNA functional elements.

With recent technological advances, we have now entered an era where we can perform high resolution structure determination by cryo-EM. Yet, its adaptability to study RNA-only systems is limited. Advancing the application of cryo-EM to RNA systems will allow for a mechanistic insight into regulatory structural features governing biological processes. In **Chapter 2,** I discuss the development of a robust pipeline to perform "RNA only" cryo-EM using the classically studied Group II intron aI5γ. Found in *Saccharomyces cerevisiae*, this intron has undergone extensive biochemical characterization, revealing it to be a model system to understand RNA folding and splicing. However, atomic-level structural information has been lacking for over 30 years due to difficulties in determining optimal crystallization conditions. My efforts resulted in a 3.6Å resolution structure of aI5γ, revealing the presence of novel functional motifs that regulate the tight folding and catalytic activity of this ribozyme. These results provide a long-waited structural insight into the 3-D architecture of this ribozyme and lay the foundation for a new era of RNA cryo-EM.

Finally, in **Chapter 3,** I apply these techniques to study the structure function relationship of long non-coding RNA, Pnky. This trans-acting lncRNA plays a critical role in repressing neuronal stem cell differentiation,[54] though the molecular mechanism of how Pnky functions is limited. I believe the higher order structure of Pnky can modulate protein recruitment and function in neuronal stem cells. Towards this, I performed a combination of biophysical, chemical probing, and cryo-EM to understand how Pnky's secondary and tertiary structural features regulate neurogenesis. My findings reveal that Pnky exhibits a high degree of structural organization with many sharp and stable turn motifs which result

in a complex 3-D architecture and thus provide a road map of candidate riboregulatory motifs for targeted functional investigation. This systematic workflow lays the foundation for exploring the three-dimensional architecture in the diverse spectrum of lncRNAs.

2. Systematic detection of tertiary structural modules in large RNAs and RNP interfaces by Tb-seq

2.1 Introduction

RNAs can adopt complex folded motifs and higher-order 3-D structures that are essential across a variety of specific cellular processes.[6,12,55-57] It has recently become clear that many types of multi-kilobase RNA transcripts contain regions of tertiary structure that, either alone or in concert with protein partners, carry out biological functions.[58,59] However, identifying these regions of complex RNA structure remains challenging. Current structure prediction methods on long RNAs are unable to pinpoint regions containing stable RNA tertiary structure modules or complex protein binding sites from sequence alone.[60] While biophysical techniques such as NMR,[61] x-ray crystallography,[62] and cryo-EM[63] are invaluable tools for the observation of RNA structure, they are time consuming and difficult to perform on a multi-kilobase long RNA that contains a mixture of both structured and flexible regions. As our understanding of their biological functions becomes increasingly important, and interest in small molecule targeting of RNAs grows, it is vital to develop new tools for identifying regions of tertiary structure in long RNA molecules.

In recent years, chemical probing has become a powerful tool for studying RNA structure. Many important advances have improved our ability to identify single- versus double-stranded nucleotides in RNA,[19,22,64-66] these data have primarily been used to infer secondary but not tertiary structures of RNA. Fewer methods have been developed to detect higher-order structures and these protocols are limited to assessment of solvent-accessible

regions [32-34] or identification of long-range RNA-RNA base-pairs by cross-linking methods[37,38] or statistical correlations in mutational profiling.[43,44] The field would benefit from a readily adaptable, high throughput approach for identifying regions of local tertiary structure, which are often hallmarks of functional RNA motifs and riboregulatory elements.

High-resolution RNA structures show that regions of tightly packed tertiary structure often contain phosphate backbones that are packed in close proximity, within the same strand or on adjacent strands. These local regions of intense negative electrostatic potential act as sinks for multivalent ion coordination.[27,67] One way to probe these electrostatically negative reservoirs is to monitor the cleavage patterns catalyzed by coordinated metal ions.[31] When nucleotides in such regions adopt an "in-line geometry", which aligns an upstream 2'-OH with the downstream 3'-OR group of a phosphodiester linkage,[68] adjacent metal hydroxide ions can behave as a general base, deprotonating the 2'-OH group and producing a 2' oxyanion nucleophile that attacks the adjacent phosphate and causes strand scission.[68] While this type of Mg^{2+}-catalyzed cleavage (known as in-line probing[30]) normally occurs on a slow timescale that ranges from hours to days,[30] the same phenomenon is greatly accelerated by lanthanide ions such as terbium (Tb^{3+}) and europium (Eu^{3+}).[29,69] Tb^{3+} and Mg^{2+} share similar ionic radii (0.92 Å and 0.72 Å,[70] respectively) and coordination geometry preferences, but lanthanide ions have an additional positive charge and the pKa of coordinated water molecules is much lower for ions such as Tb^{3+}, (pKa ~ 7.9 for Tb^{3+}-H_2O versus ~ 11.4 for Mg^{2+}-H_2O).[28,70] Therefore, at low ion concentrations and neutral pH, Tb^{3+} coordinates with structured RNA binding sites in a manner that is almost identical to that of Mg^{2+}, but the more potent Tb^{3+} general base rapidly facilitates RNA backbone cleavage at sites of metal ion binding.[28] Tb^{3+} probing of RNA has been

used extensively in the past, but until now, it was a low-throughput method that relied on electrophoretic quantification.[70-72]

Here I present Tb-seq, a sequencing-based approach that employs Tb^{3+} to detect regions of tertiary structure in long RNAs. To demonstrate the efficacy of this technique, I first applied the Tb-seq pipeline to identify tertiary structural motifs within structurally well-characterized RNA molecules. I then applied it to probe known and unknown RNA structures in a cellular context to investigate RNA motifs and protein binding sites. These studies show that Tb-seq detects regions of RNA involved in RNA tertiary structure motifs and within RNP complexes, thereby providing a powerful new approach for pinpointing regions of complex RNA structure that are potentially associated with RNA functional elements.

2.2 Results

2.2.1 Developing a high throughput sequencing-based approach to detect Tb^{3+} cleavage sites

To precisely identify tertiary RNA structural elements in a high throughput manner, I adapted a previously established Tb^{3+} RNA cleavage assay for accurate single nucleotide detection in an RNA of interest. In the classical version of this experiment, the RNA of interest is end-labeled with ^{32}P, probed with Tb^{3+}, and the sites of hydrolysis are visualized after electrophoresis of the RNA.[70-73] To adapt this assay to a sequencing readout, I first determined if the expected Tb^{3+}cleavage sites could be detected as termination events upon reverse transcription (RT). I used the D135 ribozyme derived from yeast group II intron aI5γ, which has been extensively characterized using the classical version of Tb^{3+} probing.[71] I found that reverse transcription stops (Figure 2.1a, right) recapitulate the previously-published Tb^{3+} cleavage pattern (Figure 2.1a, left), thereby validating RT as a tool to detect Tb^{3+}-induced cleavage. I then adapted this approach to a sequencing assay using a previously-developed pipeline for assessing RT termination events[74] to quantify termination (Figure 2.1b). This sequencing and analysis approach, Tb-seq, recapitulated the previously published D135 Tb^{3+} cut sites (Figure 2.2). To better understand whether Tb-seq could be used as a discovery tool for assessing higher-order RNA structure in a variety of RNA types, I next applied this method to evaluate the patterns of cleavage in RNAs with well-determined tertiary structures.

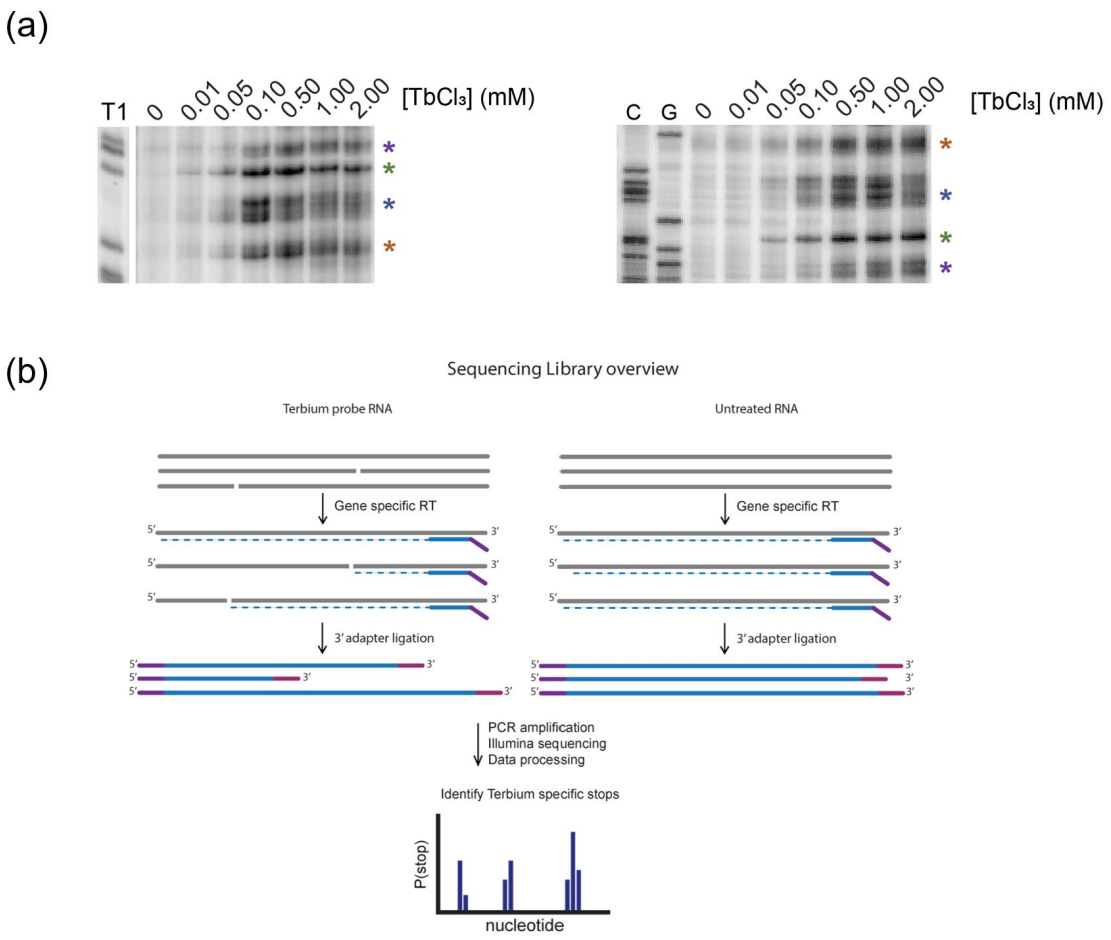

Figure 2.1. Developing a sequencing-based approach to detect Tb^{3+} cleavage sites.

(a) Denaturing electrophoresis of ^{32}P-labeled aI5γ RNA probed at the indicated $TbCl_3$ concentrations (left). Primer extension electrophoresis of corresponding cDNA products from reverse transcription (right). Asterisks indicate banding pattern.

(b) Tb-seq library construction workflow. Tb^{3+} cleaved RNA is reverse transcribed with a gene-specific RT primer containing a 5' adapter handle. The resulting cDNA is 3' adapter ligated and PCR amplified with Illumina multiplex handles. Stop sites are processed using the RTEvents counter script.

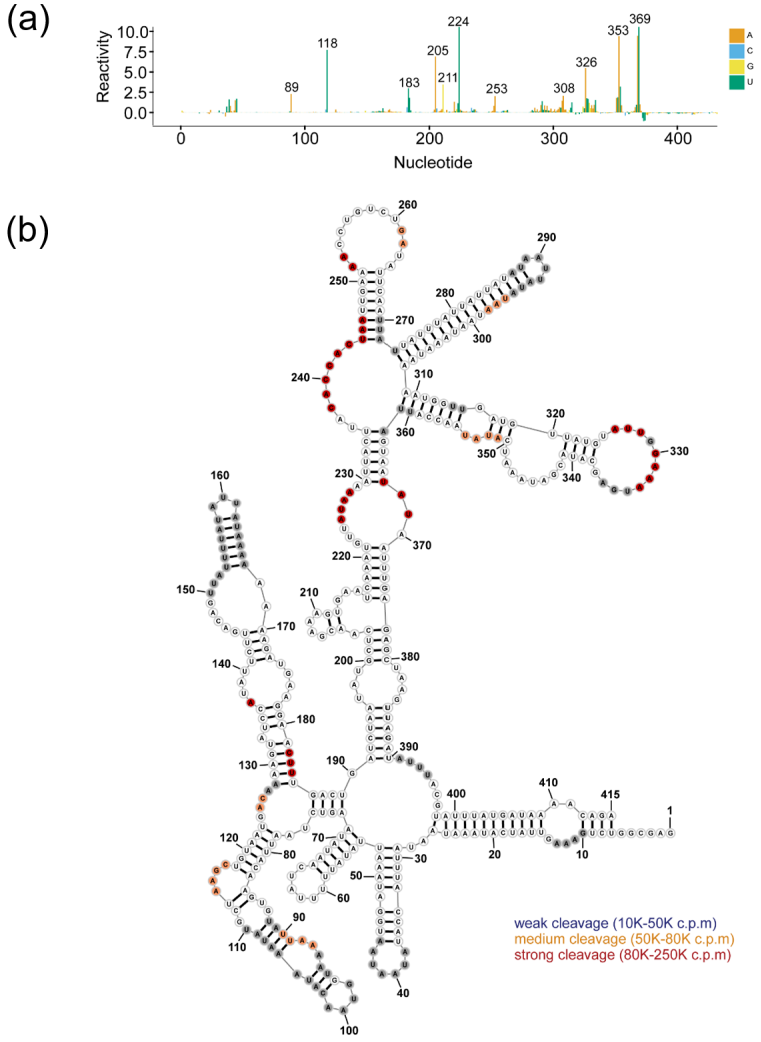

Figure 2.2. Tb-seq on D135 recapitulates previously reported cleavage sites.

(a) Bar plot showing reactivity profile obtained when probing D135.
(b) Secondary structure of D135 domain I with Tb^{3+} cleavage sites determined from electrophoresis.[71]

2.2.2 Tb-seq reveals well-folded RNA tertiary elements

To benchmark Tb-seq on RNAs that have never been analyzed with Tb^{3+} cleavage before, I probed *in-vitro* transcribed RNAs that contain both well-folded RNA tertiary elements and known metal sites. I chose a group II intron from *Oceanobacillus iheyensis* (*O.i.*) that has been well characterized biochemically and crystallographically.[18] First, I utilized primer extension gels to test reagent concentrations and reaction times to obtain an ideal reactivity signal and ensure the RNA is not over-cleaved (Figure 2.3). I find low concentrations of $TbCl_3$ and 10 minutes of probing is sufficient to produce a cleavage pattern.

I then performed Tb-seq using a range of Tb^{3+} concentrations from 0.01 mM - 2 mM for 10 min to evaluate the intensity and location of the cleavage patterns (Figure 2.4). Notably, I observe that the cleavage signal is abolished if the *O.i.* intron is denatured prior to probing, supporting my interpretation that cleavage signals are indicators of RNA structure. To determine whether secondary structure alone is sufficient to produce the cleavage pattern, the intron was folded only in the presence of monovalent ions, under conditions lacking the magnesium ions known to promote its characteristic tertiary structure[75] (Figure 2.5a). I found that secondary structure was insufficient to establish the signals, once more supporting the interpretation that Tb-seq signals correspond to sites of tertiary structure. Instead, I found that at certain Tb^{3+} concentrations (0.5 mM), non-specific cleavage is observed (Figure 2.5b). These results demonstrate that a correctly folded intron containing well-defined tertiary elements is required for Tb^{3+} coordination and site-specific RNA cleavage.

Next, I established a three-point criteria set for selecting nucleotide stop sites that are likely to result from specific, site-bound Tb^{3+}-dependent cleavage, which I will call "strong Tb signal". First, a reactivity value of >0.5 was established for detecting strong sites of cleavage and maximizing probe specificity (see Methods). Second, these sites must be observed in two independent replicates to demonstrate reproducibility. Third, the selected sites must show a dependence of signal on Tb^{3+} concentration to ensure that stop signals are not due to spontaneous RT termination events. Nucleotide sites that satisfy these criteria are highlighted in red in the secondary structure diagram of the *O.i.* intron (Figure 2.6a). Upon initial inspection, I observed that the strongest Tb^{3+} sites are in short loop regions within the RNA secondary structure. Upon close inspection, it became clear that these cleavage sites fall within or are adjacent to the most evolutionarily conserved long-range RNA tertiary interactions that are essential for correctly folding the ribozyme (annotated by Greek letters, Figure 2.6a).[76] To further understand the conformation of these sites in 3-D space, the Tb^{3+} signal was visualized on the crystal structure of *O.i.* intron (Figure 2.6b). I found that Tb^{3+} causes backbone cleavage at regions where the phosphate backbone compresses together to form sharp, stable turns. These turns are all components of RNA tertiary motifs required for the correct folding of the active ribozyme.

To examine the sites of cleavage in greater detail, I focused on two regions that are specifically recognized by Tb^{3+}. The first is the ζ-ζ' tetraloop-receptor interaction, which is among the best characterized and most important interactions for positioning catalytic intron domain 5 (D5).[76] Here, a single G236 residue in D1 flips out of a sharp backbone turn and base-stacks with A370 in D5 (Figure 2.6b top insert). Several nucleotides (234-237) in the bulge that mediates this ζ-ζ' tetraloop were identified to contain strong Tb-seq

signal. The second motif, λ-λ' is within the z-anchor, a module that forms multiple higher order structures and serves as a scaffold for properly positioning the 5' splice site. Notably, strong Tb-seq signal was observed in A106 in D1, which forms a minor groove base triple with nucleotides C267 and G374 in D5 (Figure 2.6b bottom insert). These results demonstrate that Tb^{3+} detects functionally important interactions in group II introns where RNA phosphate backbones come into very close proximity, thereby allowing for multi-helix base stacking and long-range interactions.

To further test and expand Tb-seq, I probed another class of RNA that contains a well-defined tertiary structure. For this, I chose the Hepatitis C. Virus (HCV) internal ribosome entry site (IRES), specifically focusing on domain II, which has well-characterized structural features identified by both cryoEM[77] and NMR.[78] Implementing the criteria described above, I observed strong Tb^{3+} signal clustering in two regions. The first is a loop region containing nucleotides 92-95, where the phosphate backbones kink and come into close proximity (Figure 2.7). The second region is near nucleotides 52-54, where the phosphate backbone forms a nearly 90˚ bend in the RNA (Figure 2.7). This bend is implicated in positioning of the downstream terminal loop near the 40S E site of the ribosome, which allows for translation of viral proteins. [77,79] Interestingly, this region has been targeted by functional inhibition studies where multiple small molecules bind and structurally extend the bend into an elongated conformation, inhibiting viral translation.[80,81] Together these results indicate terbium probing can detect functionally important structures in RNAs, allowing it to be used as a screening tool for identifying regions that are likely to contain compact motifs.

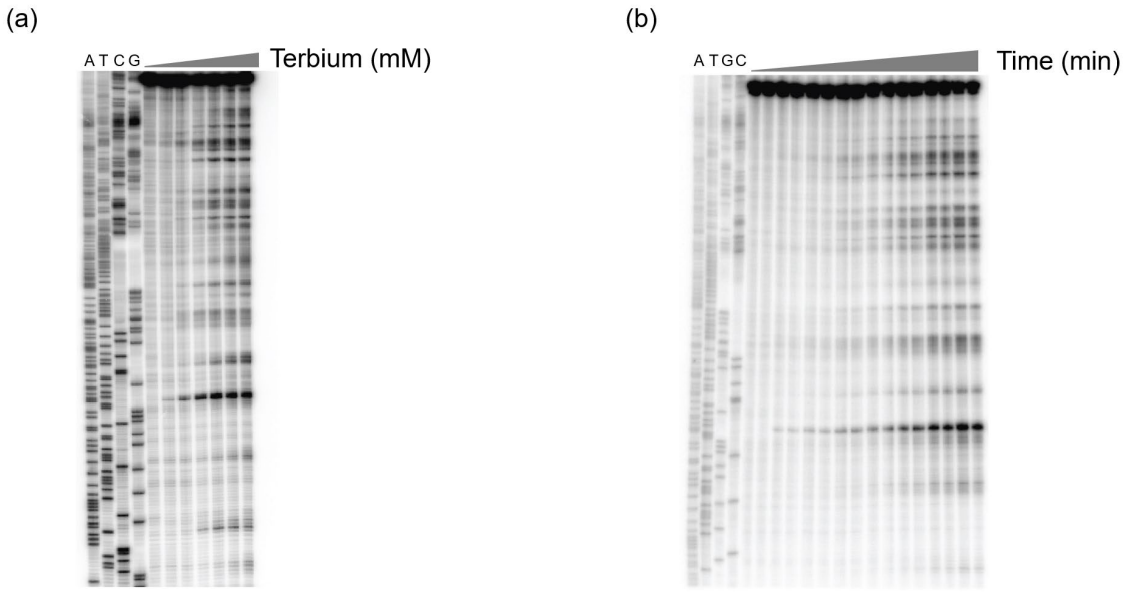

Figure 2.3. Optimizing Tb^{3+} probing conditions.

(a) Primer extension gel of D135 Tb^{3+} probed for 10 min at the indicated concentrations. The $TbCl_3$ concentrations are 0, 0.01, 0.1, 0.25, 0.5, 1, and 2 mM.
(b) Time-resolved primer extension gel of D135 probed at 0.5mM $TbCl_3$. Times of probing are: 0, 0.1, 0.2 0.5, 1, 2, 3, 5, 7, 10, 20, 30, 45, 60, and 90 min.

(a)

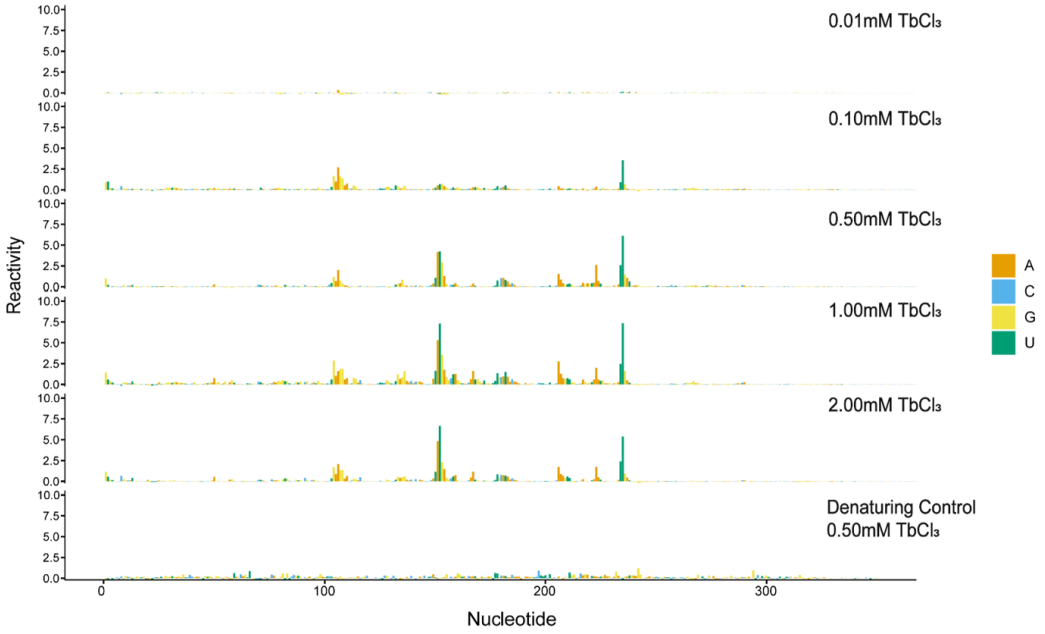

Figure 2.4. Tb-seq reactivity profile of *O.i.* intron.

(a) Bar plot showing reactivity profile obtained when Tb^{3+} probing *O.i.* intron at the indicated concentrations.

Figure 2.5. Tb^{3+} probing the secondary structure of *O.i.* intron does not result in a distinct cleavage pattern.

(a) Electrophoretic mobility shift assay of *O.i.* intron. The $MgCl_2$ concentrations are 0, 0.1, 0.5, 1, 2, 5, 10, 15, 20, 30 and 50 mM.
(b) Bar plot displaying reactivity values when probing *O.i.* intron containing only secondary structure.

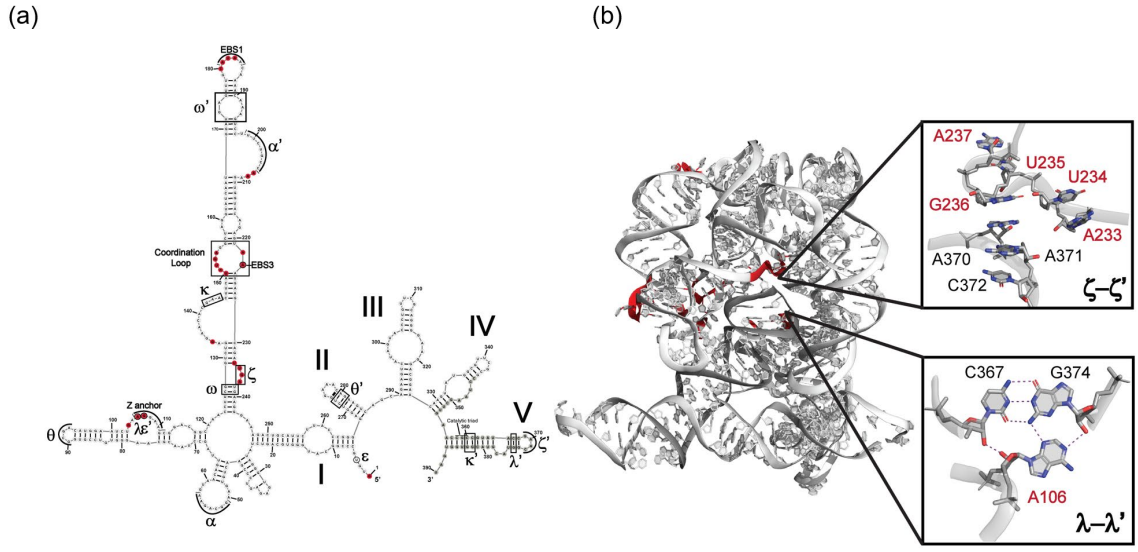

Figure 2.6. Tb-seq of *O.i.* intron detects long range, evolutionarily conserved RNA-RNA interactions.

(a) Secondary structure of *O.i.* intron displaying sites of strong Tb^{3+}cleavage (red). Long range RNA-RNA interactions are indicated by Greek letters. EBS1 and EBS2 correspond to exon binding sites. Gray nucleotides indicate a lack of sequencing data in this region.

(b) 3-D structure of *O.i.* intron showing sites of strong Tb^{3+} cleavage on the RNA backbone (red highlight). Inserts showing close up view of two long range RNA-RNA interactions (ζ-ζ' and λ-λ') with nucleotides that display Tb^{3+} cleavage (red) and hydrogen bonds (purple dashed lines). Adapted from PDB 4E8M.

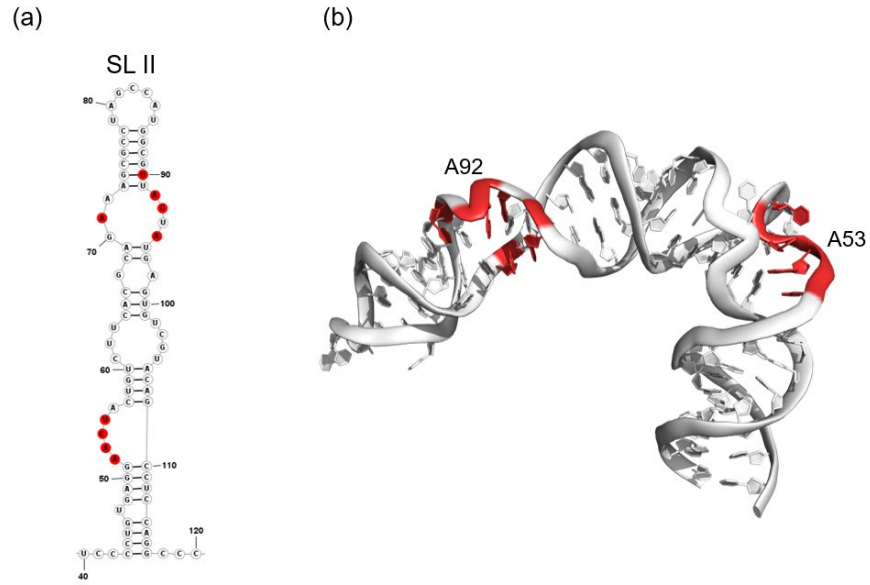

Figure 2.7. Tb-seq of HCV IRES detects conserved L-shaped bend in stem loop II.

(a) Secondary of HCV 5' UTR Stem loop (SL) II displaying sites of strong Tb^{3+}cleavage (red).
(b) 3-D structure of SL2 showing sites of strong Tb^{3+} cleavage on the RNA backbone (red highlight). Adapted from PDB 52AQ.

2.2.3 Tb-seq detects key RNA-protein interactions in a cellular context

Having established the versatility of Tb-seq on RNAs that have been *in-vitro* transcribed, I sought to extend it to cellular contexts, where RNA can fold together with proteins, small molecule ligands, and other nucleic acids.[59] I decided to conduct the first experiments on a structurally well-defined cellular RNA with known protein binding sites. To this end, I probed human RNase P to understand how terbium can be used to reveal higher-order RNA structural motifs in that stable RNP. In order to circumvent the issues of introducing lanthanide ions into cells, I developed an approach for gently lysing mammalian cells in a way that maintains intact RNA-Protein (RNP) complexes (Figure 2.8a). I then treated the resulting extract with Tb^{3+} and implemented the Tb-seq pipeline, using the criteria I established for identifying strong sites of specific Tb^{3+} cleavage (Figure 2.8b). By comparing the Tb-seq signal with the cryo-EM structure of human RNase P H1 RNA,[8] I observed that the strongest cleavage sites were found in regions where the RNA backbone bends sharply, notably at the top and bottom of the H1 RNA (nt 47-50 and 169-173) (Figure 2.9a).

Human RNase P consists of ten protein components that wrap around and bind the H1 RNA at multiple regions, presumably stabilizing its elongated conformation[8] (Figure 2.9a). While a number of sites are observed, here I highlight two examples where Tb-seq reveals regions containing critical RNA-protein interactions. The first is a backbone turn located in the loop of stem P9 (Figure 2.9a, bottom insert). The bases of nucleotides C125 and U126 form hydrogen bonding interactions with the side chains of the essential core protein, Rpp29. This protein makes multiple contacts with stem P9 and P1, bringing them together in close proximity and stabilizing the downstream helical core of the H1 RNA,

which recognizes the 5' end of pre-tRNA for cleavage. The second site of strong Tb-seq signal is observed in the loop region of stem P3. Here, the backbone, bases, and sugars of the nucleotides targeted by Tb^{3+} (C61, C63, A64, U65), form networks of hydrogen bonds with proteins Rpp20 and Rpp30b (Figure 2.9a, top insert). In this context, Tb-seq signals correspond to exposed regions of the RNA which form structural motifs that are stabilized by protein interactions within RNase P.

To further explore the ability of Tb-seq to reveal RNP interactions and to understand the role of the protein in Tb^{3+} detection at these sites, I used Tb^{3+} to probe human RNase P in the absence of proteins. To this end, Tb^{3+} cleavage was conducted on cell lysates that were treated with a proteolytic enzyme (Proteinase-K), which strips proteins from RNA (Figure 2.8a). As in studies with other chemical probes,[24] I then performed a differential reactivity comparison, termed ΔTb, to compare changes in H1 RNA structure in the presence and absence of proteins (Figure 2.9b and Figure 2.10). Consistent with a disruption of a stabilizing protein interaction, the two regions described above became less reactive in the absence of proteins (show a loss in terbium reactivity) (Figure 2.9b inserts). By contrast, other nucleotides became more reactive after Proteinase-K treatment (see stem P3, Figure 2.9b), which may result from conformational rearrangement that occurs in the absence of proteins. These data suggest that ΔTb detects modules of protein-stabilized RNA structures within RNase P, thereby broadening the applicability of the method to probing of RNP interfaces.

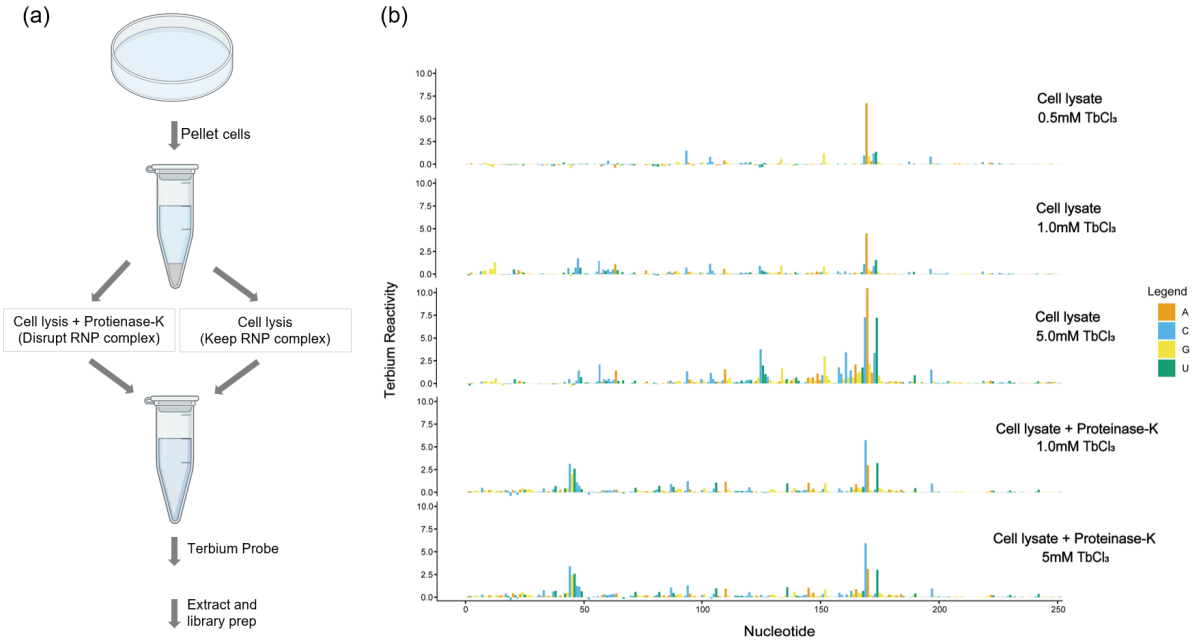

Figure 2.8. Establishing cell lysis probing on human RNase P in the presence and absence of proteins.

(a) Schematic displaying an overview of cell lysis probing to either retain or disrupt RNA-protein (RNP) complexes.
(b) Bar plot displaying Tb^{3+} reactivity values when probed at the indicated concentrations and probing conditions.

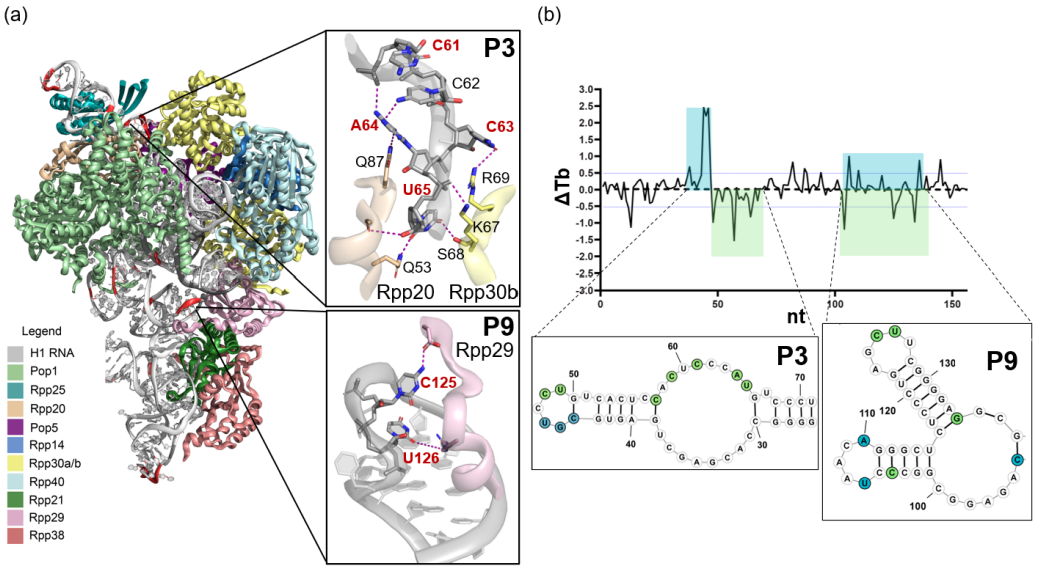

Figure 2.9. Probing RNA-Protein interactions in human RNase P.

(a) 3-D structure of RNase P complexed with its protein components showing sites of strong Tb^{3+} cleavage on the RNA backbone (red highlight). Inserts show close up views of two regions containing RNA-Protein interactions with nucleotides that display Tb^{3+} cleavage labeled (red) and hydrogen bonds (purple dashed lines). Adapted from PDB 6AHR.

(b) ΔTb reactivity profile for the first 150nt of RNase P. Nucleotides indicated in green become less reactive when probed in the absence of protein components whereas nucleotides indicated in blue become more reactive.

(a)

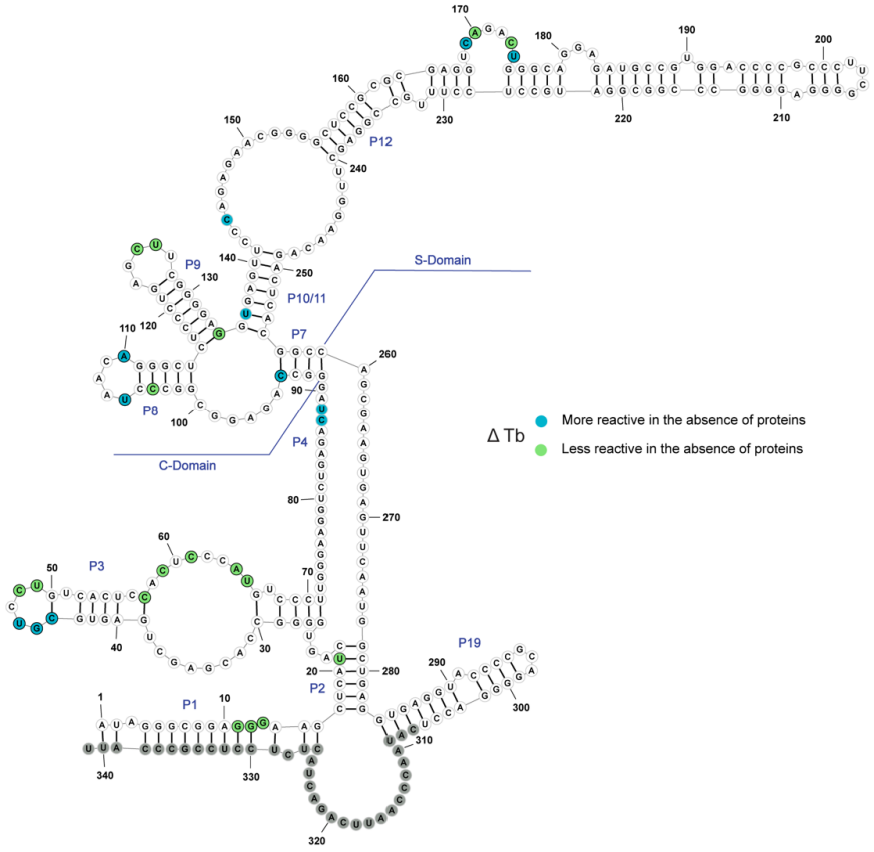

Figure 2.10. Δ Tb analysis of human RNase P.

(a) Secondary structure of RNase P displaying Δ Tb reactivities. For nucleotides in gray, no sequencing data is available.

2.2.4 Tb-seq reveals modules of higher-order structure in viral RNAs

Having validated Tb-seq as an RNA tertiary structure probe, I sought to apply it to discover novel RNA structures in multi-kilobase RNAs, such as long viral RNA genomes. Numerous studies have demonstrated that viral RNA genomes contain secondary and tertiary structures both in the UTRs and coding regions that are important for function.[82,83] Indeed, I utilized Tb-seq to detect functional RNA structures within the HCV IRES (Figure 2.7). Given the urgency of detecting functional RNA elements within SARS-CoV-2 RNA[82,84] and the limited tools available to detect them, I performed cell lysate Tb-seq in SARS-CoV-2 infected cells. I specifically examined the 5'-terminal 1400nt of the RNA genome, which contains the 5'UTR, the coding region of Nsp1, and part of the Nsp2 ORF.

Upon inspection of the Tb-seq signal profile, (Figure 2.11), I observed the emergence of a distinct set of cleavage patterns consisting of clusters of consecutive nucleotides that display Tb^{3+} cleave. Interestingly, this signal profile is similar to our previous work probing ribozymes indicating a high degree of structuredness in the viral genome. Overlaying these sites onto the predicted secondary structure,[82] I observed signal not only in the UTR but in the coding region of the genome as well (Figure 2.12). Upon closer inspection, I observed the majority of Tb-seq signal present in small stem-loop or bulge regions implicating these regions to be small modules of tight RNA structure (Figure 2.12).

To furthermore understand the role of protein occupancy on this structured genome and narrow down on potential functional modules, I probed in the absence of proteins and implemented our Δ Tb pipeline. Overall, numerous changes across the RNA were seen indicating a global conformational change in the genome structuredness (Figure 2.12). I

observed some sites where the reactivity signal increased, implying a conformational change in RNA tertiary structure or new backbone accessibility in the absence of proteins. Additionally, there are sites that became less reactive upon the loss of proteins (Figure 2.11 inserts). Given my findings probing RNase P, these sites are implicated to be a structural module containing a sharp backbone bend that is stabilized by protein components. The limited information available on the proteome of this genome makes it difficult further assess the interactors. Nevertheless, together these data underscore the utility of combinatorial Tb-seq for narrowing down structural modules and provide a roadmap of candidate functional structures regulating viral lifecycle.

(a)

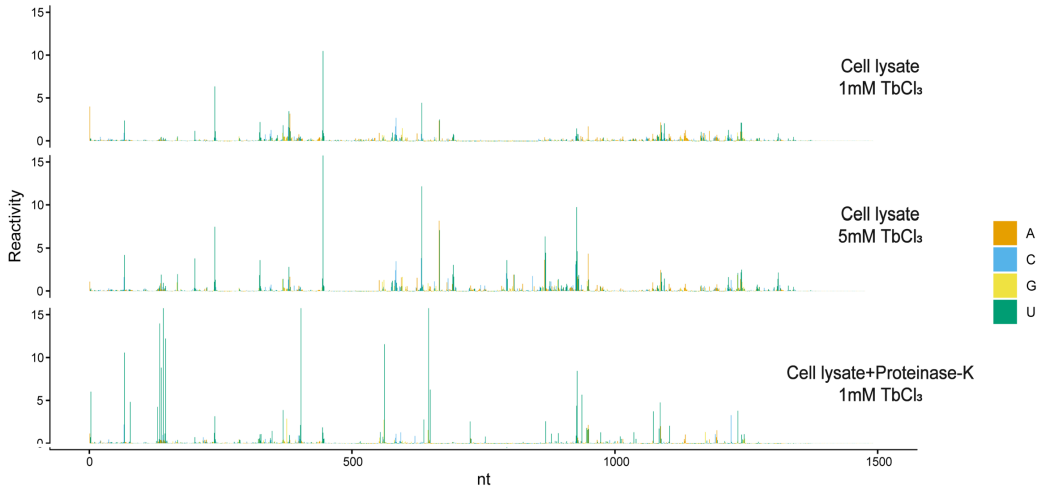

Figure 2.11. Cell lysis probing on of SARS-CoV-2 in the presence and absence of proteins.

(a) Bar plot displaying reactivity values when probed at the indicated concentrations and probing conditions.

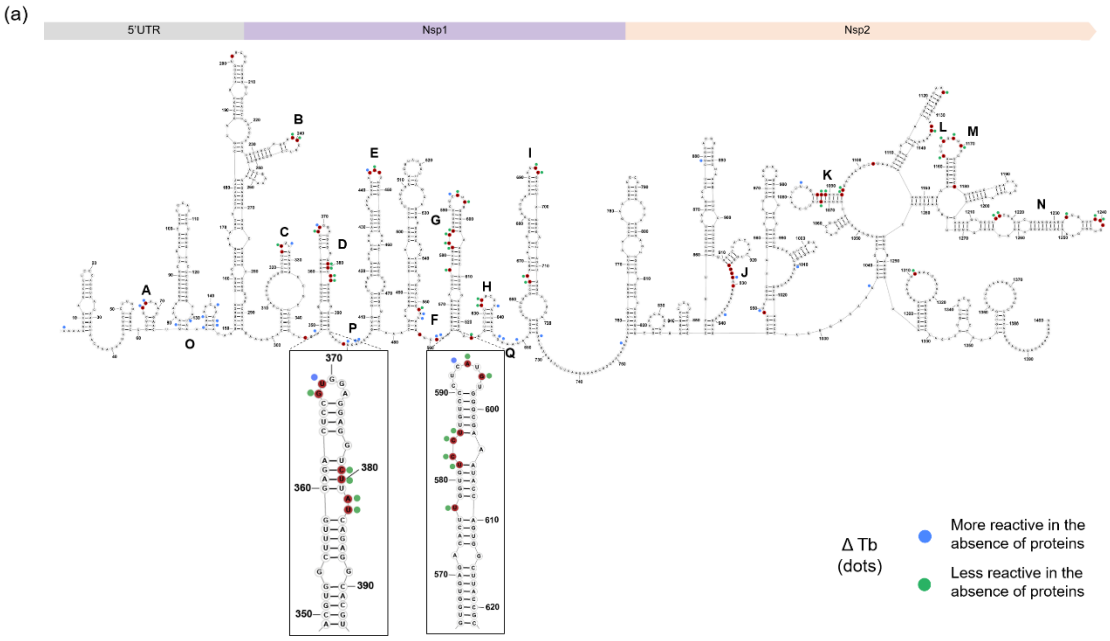

Figure 2.12. Tb-seq identifies novel structural modules in SARS-CoV-2.

(a) Cell lysate probing of the 5' terminal of SARS-COV2. Nucleotides mapped in red displaying strong Tb^{3+}cleavage. Inserts display close up views of two regions displaying Δ Tb reactivities.

2.3 Discussion

As biologists explore the growing landscape of biologically important multi-kilobase RNAs, such as viral genomes, unprocessed mRNAs, primary miRNAs and long noncoding RNAs, new tools are needed that will enable researchers to focus their attention on specific regions of RNA for detailed functional analysis. The Tb-seq pipeline presented here provides one such filter, yielding valuable information about structurally compact local RNA motifs that differs from the information reflected in other probes of secondary and tertiary structure. In addition, by using the Δ Tb probing strategy, and probing in the presence and absence of protein components, one can narrow down tertiary structures that undergo protein-dependent conformational differences. Ultimately, integrating Tb-seq with orthogonal chemical probes, pull-down methods,[85,86] cross-linking agents,[37,38] and functional assays will allow for a comprehensive mechanistic understanding of individual RNA molecules.

With recent technological advances, we have now entered an era where it is possible to rapidly determine high resolution structures of large RNAs.[87] However, multi-kilobase RNAs cannot be visualized in their entirety as they often contain conformationally flexible regions.[88-90] For this reason, most RNAs are amenable to high-resolution structure determination only after a careful study of their overall structural landscape. This requires a methodical approach for identifying RNA regions and RNP substructures that can be visualized with powerful tools such as cryo-EM and SAXS. In addition, there are many cases where one must rationally design or isolate stable motifs of RNA and/or RNP complexes. Here, I provide a way to identify the most structurally compact regions of a

large RNA and, in tandem with other long-range probing methods, choose the best regions for high-resolution investigation.

Performing Tb-seq on RNAs with known structures provided a useful starting point for assessing the types of RNA motifs that are recognized and cleaved by Tb^{3+}. I initially attempted to correlate Tb-seq signals with specific torsion angles and with atomic distance vectors between different base and backbone atoms, but no clear correlation emerged. In order to develop a less fine-grained correlation metric, I visualized the structures and noted that most Tb^{3+} cleavage sites occur in regions where multiple phosphate backbone residues pinch together in close proximity. To reflect this, I computed a metric for assessing the "sharpness" of turns in the RNA backbone at Tb^{3+} cleavage sites, deriving our values from high-resolution structures of the *O.i.* intron.[18] Specifically, I measured the backbone phosphate distances between nucleotide n to nucleotide n+2 at sites displaying strong Tb-seq signals (P_n -> P_{n+2}, or every other phosphate) and compared these data to the corresponding distances in a simple helical structure within domain 4 of the intron (Figure 2.13). I found that RNA regions with strong Tb-seq signals tend to have very small P_n -> P_{n+2} values (5.5 - 9.1Å) relative to the same distances calculated from a simple helix (9.6-12.5 Å), indicating local compression of the RNA backbone. I speculate that additional data will allow for refinement of these parameters. For example, while some Tb^{3+} cleavage sites, such as those in region 7, are not characterized by small P_n -> P_{n+2} values, visual inspection of the structure shows that these same nucleotides are part of a larger motif in 3-D space that contains adjacent pinched backbones that are characterized by strong Tb-seq signatures and small P_n -> P_{n+2} values (in region 5). Therefore, the same bound metal ion may be catalyzing both cleavage events. Additional data and analysis will enable a

more quantitative description of specific structural features that are recognized and cleaved by Tb^{3+}.

The human transcriptome contains a vast set of large, complex RNA molecules, and until recently, we have lacked the tools to assess their 3-D structural content. However, the biochemical methods that were initially developed to study tRNAs, riboswitches and ribozymes are being gradually being adapted to explore the growing repertoire of multi-kilobase RNAs that are central to gene expression and pathogenicity. Here I present a much-needed expansion of the RNA probing toolbox that allows investigators to rapidly pinpoint candidate RNA tertiary structures efficiently and precisely, paving the way for downstream mechanistic study and therapeutic targeting.

(a)

	O.i. Domian 1	
	Nucleotide	Backbone distance (Å) from Phosphate of nucleotide n to Phosphate of nucleotide n+2
Region 1	1	11.837
	2	10.171
	3	10.279
Region 2	102	11.107
	103	11.607
	104	8.311
	105	6.929
	106	11.209
	107	8.661
	108	10.947
	109	5.604
	110	5.797
	111	10.975
Region 3	133	6.652
	134	11.817
	135	9.408
	136	8.653
	137	10.009
Region 4	148	10.970
	149	10.735
	150	11.544
	151	6.524
	152	11.606
	153	10.518
	154	8.700
	155	11.660
	156	11.808

	O.i. Domian 1	
	Nucleotide	Backbone distance (Å) from Phosphate of nucleotide n to Phosphate of nucleotide n+2
Region 5	178	8.358
	179	8.238
	180	7.588
	181	8.930
	182	11.244
	183	11.862
	184	10.275
	185	11.133
Region 6	204	12.313
	205	11.963
	206	9.530
	207	9.283
	208	10.029
	209	10.451
Region 7	221	12.307
	222	10.117
	223	11.392
	224	10.577
	225	13.071
	226	11.594
Region 8	233	10.520
	234	10.581
	235	11.009
	236	6.836
	237	9.108
	238	9.256
	239	11.196

(b)

O.i. Domain 4	
nucleotide	Backbone distance (Å) from Phosphate of nucleotide n to Phosphate of nucleotide n+2
328	10.88
329	10.898
330	11.178
331	11.591
332	12.855
333	10.274
334	9.631
335	11.004
336	12.559
337	11.802
338	11.401
339	11.958
340	11.584
341	10.985
342	12.388
343	12.043
344	11.523
345	11.715
346	10.965
347	12.552
348	11.614
349	9.972
350	11.221
351	10.892
352	10.642
353	11.657
354	9.948

Figure 2.13. Correlating Tb-seq signal with backbone phosphate distances for Group II *O.i.* intron.

(a) Backbone distances (Å) for regions displaying Tb^{3+} cleavage. Distances were calculated from phosphate of nucleotide n to phosphate of nucleotide n+2 (P_n -> P_{n+2}). In red text are sites which display strong Tb^{3+} cleavage. Highlighted boxes indicate distances deviating from values in a simple helix in *O.i.* domain 4[18].

(b) Backbone distances (Å) of helix in *O.i.* domain 4. Distances were calculated from phosphate of nucleotide n to phosphate of nucleotide n+2 (P_n -> P_{n+2}). All values calculated from PDB 4E8M.

2.4 Materials and Methods

In-vitro transcription and purification

The *in-vitro* transcription of aI5γ D135[91] and *Oceanobacillus iheyensis (O.i.)* group II intron D1-5,[18] was performed as previously described.[92] In brief, RNAs were transcribed by runoff transcription using T7 RNA polymerase[93] in a buffer containing 12mM $MgCl_2$, 40mM Tris-Cl pH8, 2mM Spermidine, 10mM NaCl, 0.01% Triton X-100, 10mM DTT, 5µl SUPERase-In and 3.6mM of each NTP. The reactions were incubated at 37˚C for 2 hours followed by purification on a denaturing 5% polyacrylamide gel. The *in-vitro* transcription of full length HCV genome (JC1)[94] were performed as previously described.[95] Transcribed RNA was buffer exchanged into a filtration buffer (50mM HEPES pH7.2 and 150mM KCl) using 50-kDa Amicon Ultra filtration columns. The RNA was purified by size exclusion chromatography at room temperature using a self-packed 24ml Sephacryl S-1000 column equilibrated with filtration buffer. RNA from the peak fraction was used for subsequent folding and probing.

RNA folding and Tb^{3+} probing

For D135, Tb^{3+} cleavage was performed using two approaches. The first was direct visualization of Tb^{3+} mediated RNA cleavage by electrophoresis.[71] In brief, D135 was 5′ end labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase. Thereafter, 3nM of ^{32}P - labeled RNA and 1µg of unlabeled RNA were mixed in a monovalent buffer containing 50mM MOPS pH7 and 500mM KCl to a final volume of 18µl. For visualization of Tb^{3+} mediated cleavage sites by reverse transcription and sequencing 1µg of unlabeled RNA

was used. For all reactions, the mixture was heated up to 90°C for 1 min and cooled at room temperature for 2 min. Thereafter, 2µl of 1M $MgCl_2$ (final concentration 100mM) was added and folded at 37˚C for 30 min. Subsequently, probing was performed by incubating 18µl of the folded RNA with 10x $TbCl_3$ stocks prepared in the monovalent buffer (final 1x concentration from 0.01mM-2mM $TbCl_3$) or 2µl of monovalent buffer (negative control) for 40 min at 25 °C. For the time-course experiments, probing was performed at the indicated times. All reactions were quenched with the addition of 3µl of 50mM EDTA pH 8 and precipitated by adding 1/10 volume of Na-Acetate (3M, pH 5.2), 0.5µl of glycogen (Invitrogen), and three volumes of ethanol. RNAs were resuspended in 4µl of loading buffer (82 % (v/v) deionized formamide, 0.16 % (w/v) xylene cyanol (XC), 0.16 % (w/v) bromophenol blue (BB), 10 mM EDTA, pH 8.0) and resolved on a denaturing 5% polyacrylamide gel. The gel was dried, exposed to phosphor screens overnight and scanned using a Typhoon FLA9500 phosphorimager (GE Healthcare) or Typhoon RGB Biomolecular imager (Cytiva).

For Tb^{3+} probing of *O.i.* and HCV, 1µg of RNA was diluted in the respective monovalent ion buffers (50mM HEPES pH7 and 150mM KCl for *O.i.* or 50mM HEPES pH7.2 and 150mM KCl for HCV) to a final volume of 18µl. Thereafter, 2µl of 100mM $MgCl_2$ (final concentration 10mM) was added and incubated at 37˚C for 30 min. Subsequently, probing was performed by incubating 18µl of the folded RNA with 2µl of 10x $TbCl_3$ stocks prepared in the respective monovalent ion buffers (final 1x concentration from 0.01mM-2mM) or 2µl of respective monovalent buffer (negative control) for 10 min at 25 °C. Reactivities were compared under conditions where 0.5mM $TbCl_3$ was employed and used in all figures, unless indicated otherwise. All reactions were quenched with the

addition of 3µl of 50mM EDTA pH 8. For the denaturing control, RNA was folded as described above but afterwards, deionized formamide was added to a final concentration of 50%. The denatured RNA was probed with a final concentration of 0.5mM $TbCl_3$. For the secondary structure control, native gel electrophoresis was carried out as previously described.[75] RNA was initially incubated in a monovalent buffer in the absence of $MgCl_2$. Probing was carried out at the indicated $TbCl_3$ concentrations. All RNA samples were cleaned up using a Zymo RNA clean and concentrator column according to the manufacturer's instructions.

Cell culture of human RNase P and SARS-COV-2 infection

For in-cell studies of RNase P RNA structure, Huh7.5 cells (ATCC) were cultured in Dulbecco's Modified Eagle Medium (DMEM w/o sodium pyruvate) that was supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 1mM non-essential amino acids. Cells were cultured to ~80% confluency (~$5x10^6$ cells) in a 150 cm tissue culture-treated dish.

For studies of SARS-COV-2 RNA, Huh7.5 cells (ATCC) were cultured in DMEM supplemented with 10% FBS and 1% Penn/Strep. Approximately $5x10^6$ cells were plated in T-150 tissue culture-treated flasks and incubated overnight at 37°C/5% CO_2. The next day, media was removed and $5x10^5$ PFU (MOI ~ 0.1) of SARS-Related Coronavirus 2 Isolate USA/WA2020 (BEI Resources #NR-52281) was added to each flask in fresh media. Cells were incubated with virus inocula until three days post-infection (dpi).

Cell lysis probing

For all flasks the media was aspirated, cells were washed once with cold wash buffer (50mM HEPES-KOH pH7.2, 150mM NaCl, 3mM KCl), and then dislodged in 2ml of cold wash buffer with a cell scraper. The cells were collected and centrifuged at 200g x 5 min at 4˚C. The supernatant was removed and the cells were resuspended in 2ml lysis buffer (1% TritonX-100, 50mM HEPES-KOH, pH7.2, 150mM KCl, 18mM NaCl, 1mM $MgCl_2$, 1mM $CaCl_2$, 30ul SUPERase-In (20U/µl) and 1x cOmplete Protease Inhibitor Cocktail EDTA-free). To 250µl of resuspended cells, 50/µl of Turbo DNase (2U/µl) was added and the mixture was incubated at 37˚C for 20 min. For cell lysis + Proteinase-K probing experiments, cells were prepared, lysed and DNase digested as described above, but the lysis buffer did not contain protease inhibitor. Subsequently, 25µl of 20mg/ml Proteinase-K was added to each 250µl of lysed cells and the mixture was incubated at 37˚C for an additional 20 min.

All reactions were centrifuged at 200g x 15sec. Probing was performed by incubating 225µl of supernatant with 25µl of freshly made 10x $TbCl_3$ (final concentrations from 0mM-5mM, prepared in wash buffer). The reactions were immediately placed on a rocker and allowed to incubate at 25°C for 10 minutes before quenching with 20µl of 0.1M EDTA. RNA was extracted using Trizol according to manufacturer's instructions. For experiments involving RNase P, total RNA was ribosome depleted using a Ribominus kit that was used according to manufacturer's protocol with the following exception: the ribodepleted supernatant was purified using a Zymo RNA clean and a concentrator to retain RNAs that are greater than 17 nucleotides in size. For experiments involving SARS-CoV-2, total RNA was cleaned using a Zymo RNA clean and concentrator column.

Reverse transcription (RT)

For each probing condition, 1-4µg of *in-vitro* transcribed or cellular RNA was mixed with 1-2pmol of gene specific primers (Appendix, Table 1), and brought to a volume of 7µl. To anneal primers, the mixture was heated at 90˚C for 1 min followed by 30˚C for 2 min. To initiate reverse transcription, 2ul of Marathon RT,[96] 10µl of 2x MarathonRT buffer (100 mM Tris-HCl pH 8.3, 400 mM KCl, 4 mM MgCl2, 10 mM DTT and 40% glycerol), 1µl of 10mM dNTP mix (NEB) were added and incubated at 42˚C for 30 min. RNA was degraded with the addition of 1µl of 3M KOH, heated to 95˚C for 5 min, and snap cooled to 4˚C for 5 min. Thereafter, 1µl of 3M HCl was added to neutralize the reaction. For primer extension reactions that would be visualized using electrophoresis, reverse transcription was carried out as described, but using a ^{32}P labeled primer. After the reaction, cDNA was ethanol precipitated at -20˚C overnight. The cDNA pellets were dissolved in 5µl of loading buffer and resolved on a denaturing 5% polyacrylamide gel. The gel was dried, exposed to phosphor screens overnight and scanned using a Typhoon RGB Biomolecular imager (Cytiva). For ladder generation, RT was carried out using Thermo Sequenase cycling kit according to manufacturer's instructions with an input of 500ng of template.

Sequencing library preparation

The cDNA products from reverse transcription were purified using AMPure XP beads by adding a 1.2x bead to sample ratio and incubating at room temperature for 10 min. The beads were captured using a magnetic rack for 5 min and washed 3 times with 180µl of fresh 80% ETOH. The beads were air dried for 5 min and resuspended in 12µl of water to elute the cDNA. Thereafter, 3' adaptor ligation was performed by mixing 8µl of

purified cDNA with 0.2µl of 50µM 3' adaptor (Appendix, Table 1), 1µl of 10mM ATP, 2µl of T4 RNA Ligase buffer, 8µl of 50% PEG 8000. The mixture was incubated at 25˚C for 16 hours, followed by enzyme deactivation at 65˚C for 15 min. Ligated products were purified with AMPure XP beads using a 1.2x bead to sample ratio. The products were PCR amplified 4-12 cycles with Q5 HF DNA polymerase using Illumina TruSeq forward primer and indexed reverse primers (NEB Next multiplex oligos) (Appendix, Table 1), with cycle times of 98˚C for 10 sec, 62˚C for 45 sec, and 72˚C for 60 sec. PCR products were purified with 1.2x volume of AMPure XP beads. Library concentrations were determined using a Qubit dsDNA HS Assay Kit and a BioAnalyzer High Sensitivity DNA Analysis. Libraries were diluted, pooled, and sequenced using a NextSeq 500/550 or NextSeq 2000 platform.

Tb-seq data analysis

All FASTQ files were processed using Cutadapt to remove Illumina adaptor sequences and then aligned to the respective RNA sequence using HISAT2 (v2.10). Stop information was extracted using RTEventsCounter.py script.[74] The probability of stop per nucleotide was calculated as the number of stops divided by the sum of the total number of read-through events plus the number of stops (equation (1)). Probabilities were background subtracted against a no-probe control (equation (2)). Only nucleotides that contained more than 10,000 read-throughs were considered. To better compare probing experiments conducted in different contexts, including *in-vitro* and in cell conditions where efficiencies of cleavage might differ, values were normalized to the top 10th percentile of stop rates, then scaled from 1-8 (termed "reactivity," below, based on[97]).

$$containsP(stop) = \frac{n_{stop}}{n_{stop} + n_{read-through}} \quad (1)$$

$$Reactivity = P(stop)_{treated} - P(stop)_{untreated} \quad (2)$$

Reactivities were compared under conditions where 1mM $TbCl_3$ was employed and used in all figures unless indicated otherwise. For the Δ Tb analysis, the reactivity obtained from cell lysate + Proteinase-K probing experiments was subtracted from the reactivity obtained from cell lysate probing experiments (equation (3)). In order to take a conservative approach, a stringent cutoff of +/-0.5 was implemented to detect strong differences in reactivities.

$$\Delta Tb = Reacitivty_{Cell\ lysate+Proteinase-K} - Reactivity_{Cell\ lysate} \quad (3)$$

Structure and graphical display

All secondary structures were visualized and drawn using StructureEditor.[98] All three-dimensional structure renderings were done using PyMOL Molecular Graphics System, Version 2.0, Schrödinger, LLC. Graphical displays were made using GraphPad Prism 8 or RStudio Version 1.2.5001. Solvent Accessible Surface Area was calculated using POPScomp.[99]

3. Visualizing the ai5γ group IIB intron by cryo-EM

3.1 Introduction

Group II introns are large, catalytic RNAs that facilitate a multitude of reactions, including self-splicing,[100] reverse splicing,[101] and retrohoming.[102] Found in nearly all domains of life, they can be subdivided into three main phylogenetic families, the IIA, IIB, and IIC intron.[7] While the primary sequence in these families is not well conserved, their secondary structure organization is highly conserved, consisting of six major domains (D1-D6) radiating from a central wheel.[103] Each of these six domains form complex intra-and inter-domain interactions that result in a catalytically active ribozyme. Throughout the years, this catalytic activity of self-splicing group II introns ribozyme has made them an excellent reporter system for understanding RNA folding mechanisms and the key role of RNA structure in function.[76,104]

While previous structural work has characterized the three-dimensional architecture of class IIA[105] and IIC[18,106] introns, limited information is available on class IIB introns.[107] Group IIB introns are the oldest, most structurally complex of the classes and are closely related to human spliceosome machinery. Located in the yeast mitochondrial cytochrome oxidase gene *cox1,* the group IIB aI5γ intron plays an essential role in respiration[108]. While significant progress has been made to understand the role of aI5γ tertiary structure in catalysis,[92,100,109-111] a high-resolution molecular model remains elusive. Given its biological significance and current therapeutic targeting,[16] understanding the structural

details of its catalytic active site would aid in designing more potent and specific small molecule inhibitors.

This lack of structural details of the group IIB class of introns can be attributed to the limited structural methods to assess such tertiary structures, especially of large RNAs. Traditionally, labs use X-ray crystallography as the primary means to determine high-resolution 3-D structures of biomolecules.[112] However, this requires the purification of large quantities of sample and involves screening thousands of chemical conditions to obtain crystals of the sample. In recent years, the revolution in single particle electron cryo-microscopy (cryo-EM) has resulted in a rapid explosion of newly determined high-resolution structures of diverse macromolecular complexes.[51] Cryo-EM has increasingly become a useful tool to understand molecular level details of biomolecules primarily because it has two advantages over traditional methods like X-ray crystallography. First, it requires a reduced concentration of starting material and does not require the formation of a compact crystalize structure.[113] Second, the development of downstream data processing software can deconvolute dynamic functional states of large complexes.[52] Nevertheless, cryo-EM's application to study RNA-only systems is limited as serious bottlenecks exist in the sample preparation and subsequent model building process.

Here, I describe my efforts toward developing a pipeline to perform single particle cryo-EM of large RNAs. I first established conditions to form a homogeneous, well-folded species of RNA in solution. I then optimized cryo-freezing conditions to obtain a reproducible, even distribution of well-separated particles in a thin layer of vitreous ice. Finally, I implemented a strategy to collect and process a single particle cryo-EM data set. The resulting effort, lead to a 3.6 Å reconstruction of group II aI5γ intron revealing novel

elements contributing to the folded three-dimensional architecture of this ribozyme. Given its broad applicability, this cryo-EM pipeline can be readily adapted to understand the molecular basis in a diverse set of large RNA systems. This work was done in collaboration with Ananth Kumar.

.

3.2 Results

3.2.1 Folded D135 adopts a homogenous, compact state

To study a defined conformation of the intron, a minimal functional ribozyme construct, termed D135, containing full-length domains 1, 3, and 5 and shortened domains 2 and 4, was used.[109] This construct had all the necessary components to catalyze the cleavage of its exonic substrates as a multi-turnover enzyme.[75]

Before beginning any structural studies, it was important to identify ionic conditions that promote the formation of a homogenous, compact RNA. Extensive biochemical characterization throughout the years revealed that this intron requires high ionic conditions to promote its folding and subsequent splicing *in-vitro.*[109] However, these studies were often done with low concentrations of RNA (nano molar range). Given that higher RNA concentrations (micro molar range) are needed for cryo-EM sample preparation, I wanted to assess the quality of folded D135 at these higher concentration ranges. Towards this goal, I used two complementary techniques to study the folded state of this intron, Sedimentation velocity analytical ultra-centrifugation (SV-AUC) and negative stain electron microscopy.

SV-AUC is a powerful method to quantitively assess macromolecular size, shape, and interactions in solution.[114] These molecules are characterized in their native state in various ionic solutions that can be simultaneously run and compared. Performing SV-AUC on D135 revealed the unfolded species (no Mg^{2+}) sedimented at 9S (Figure 3.1a) while the folded species (+ Mg^{2+}) sedimented at 13S (Figure 3.1a). This faster sedimentation

indicated an overall increase in compaction of the RNA upon folding in the presence of Mg^{2+}. Furthermore, consistent with the hyperchromicity effect, the compacted RNAs absorbed less UV and displayed a decrease in peak intensity. Importantly, the presence of a single peak indicated a monodisperse state with minimal aggregate or broken species (Figure 3.1a).

Next, to directly visualize the three-dimensional architecture of the RNA, negative stain electron microscopy was employed. Negative stain EM is a simple and rapid way to assess the size and homogeneity of the particle and can be used to evaluate low-resolution features. In negative stain EM, the molecule of interest was mixed with a heavy stain such as uranyl acetate, embedding the particles onto the grid. Inside the vacuum chamber of the microscope, a low dose of electrons interacted with the stain, enhancing the contrast and visualization of the particles. Electron microscopy images (micrographs) of the negatively stained unfolded D135 revealed the presence of elongated, conformationally heterogeneous particles along with some degree of aggregation. Strikingly, micrographs of the folded D135 RNA revealed compact, globular particles of 9-13 nM size. Moreover, these particles were uniformly distributed with minimal to no aggregation (Figure 3.1b). Reference-free 2D classification of these particles resulted in well-aligned classes containing a diverse distribution of orientations. Importantly the overall features, including size and shape, were comparable to that of the computationally predicted model of aI5γ[115] (Figure 3.2b).

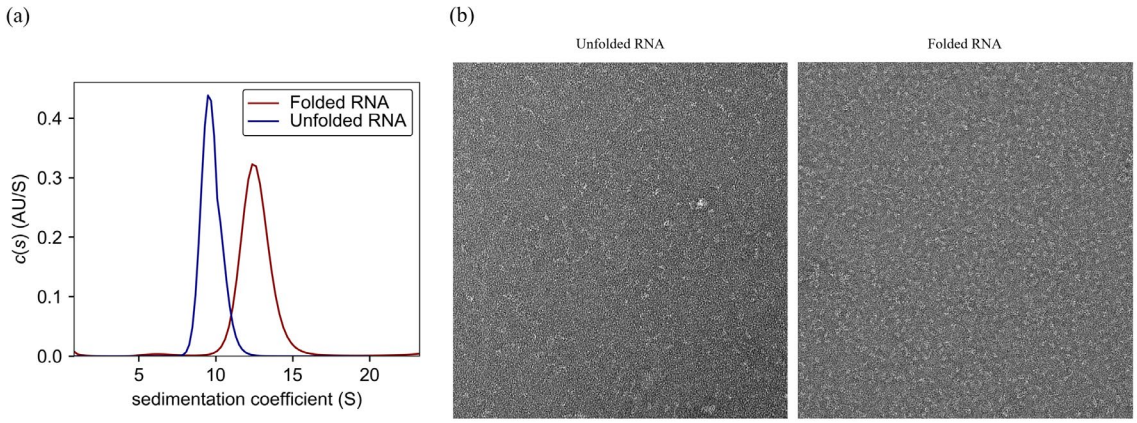

Figure 3.1. Group II intron folds into a compact homogenous structure.

(a) SV-AUC profile of unfolded group II intron and folded group II intron.
(b) Negative stain electron microscopy of the RNA at ~60,000x magnification.

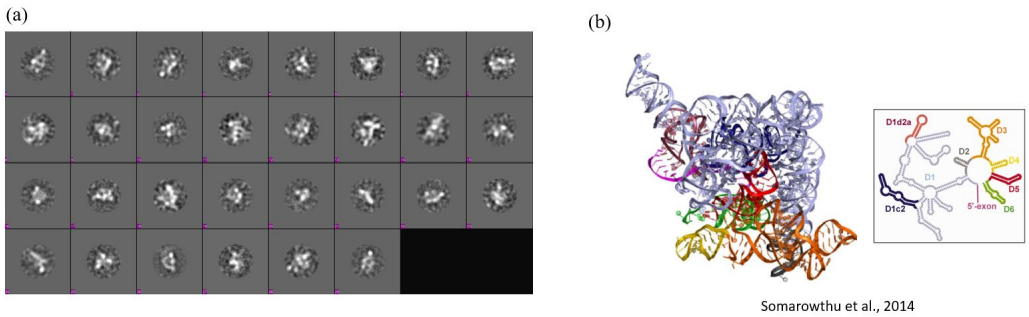

Figure 3.2. Negative stain EM of D135 intron.

(a) Selected 2D class averages of ~50,000 aligned particles.
(b) Computationally predicted 3-D model of aI5γ intron.

3.2.2 Optimizing cryo freezing conditions

Having established biophysical conditions to obtain homogenous particles, the next steps involved freezing the RNA sample into a thin layer of vitreous ice on an EM grid, using a commercially available automated plunge freezing instrument (Vitrobot). However, this process entailed extensive trial and error as the embedded particles needed be evenly distributed with a minimum inter-particle distance of ~50-100 nM, and must be in a broad range of orientations. I first tried freezing D135 using Quantifoil grids consisting of a copper frame and a carbon mesh at a variety of blot times and forces. I then screened these grids on a 200 keV Glacios microscope equipped with a K2 camera, and observed little to no particles present in the holes. A closer inspection of the surrounding carbon film revealed that most particles adhered to the carbon. Therefore, I sought to change the grids to UltraAufoil grids consisting of a gold frame and mesh. These grids revealed that most of the holes were empty and only a few contained a low number of particles (~10-60 per hole). Nevertheless, I collected a small dataset on the Glacios microscope and processed the micrographs.

Preliminary 2D class averages revealed features consistent with those obtained from negative stain (Figure 3.3). An initial 15 Å reconstruction revealed structural features consistent with SV-AUC and modeling.[115] However, at this resolution, it was difficult to model in any domains of the intron into the density. Notably, the final reconstruction only had ~30,000 particles in the class. Recent work revealed that high-resolution RNA reconstructions often had an order of magnitude more particles to achieve sub 4 Å resolution.[17] In the current cryo-EM girds, few numbers of holes had particles in them, and they only contained <60 particles. Due to the physical properties of gold, it was technically

challenging to know whether a hole contained thick or thin ice and particle, until I obtained a cryo-EM image of the hole. It was clear that to have an efficient data collection strategy, I had to improve the particle concentration on the grids and obtain grids with uniformly varying ice thickness.

To overcome these challenges, I decided to systematically vary and test a multitude of variables including varying hole sizes, support layers, and the use of additives to alter surface tension. During the blotting process, thin films of ice containing the RNA particles are produced prior to plunge freezing the sample. Though the precise physical forces during this are not well characterized, the slightest variation in ice thickness can drastically affect the particle distribution, homogeneity, and size. I tested two ways to produce a thicker film and obtain more particles in the holes. The first involved reducing the blot time and blot force and the second involved reducing the grid hole diameter, as the reduced surface area has been shown to produce thicker film/ice regions.[116] Unfortunately, both these attempts did not aid in the distribution of particles.

Along a similar line of reasoning, I sought to alter surface tension at the air-water interface to promote the distribution of particles in the holes. For this, I used a variety of detergents that were added to my sample seconds before blotting and plunge freezing. While detergents have successfully been used for membrane protein cryo-EM,[117] their effects on RNA vitrification are not well tested. The use of non-ionic detergents, Tween-20 and Triton X-100, resulted in the unfolding and aggregation of the RNA (Figure 3.4c). Zwitterionic and ionic detergents, CHAPS and deoxycholate respectively, resulted in some holes having particles, but numerous aggregates were present and furthermore, the particles were not consistently present in all holes (Figure 3.4d).

Since my previous micrographs revealed particles adhering to the support film, I sought to add a solid carbon support layer over the holes.[118] This process has previously resulted in the successful distribution of particles.[105] I first used commercially available grids containing a thin 2 nM layer of carbon on the holes. However, these grids showed the presence of large aggregates consisting of unfolded RNA particles (Figure 3.4e). Next, I attempted to manually add a thin layer of graphene oxide to the holes. This proved to be a technically challenging feat, resulting in an uneven distribution of graphene thickness and little to no particles present in the holes (Figure 3.4f).

Finally, using a reverse side blotting technique where the RNA sample was applied to the non-glow discharged side (see Methods), I was able to reproducibly create grids with an even distribution of particles (Figure 3.4h).

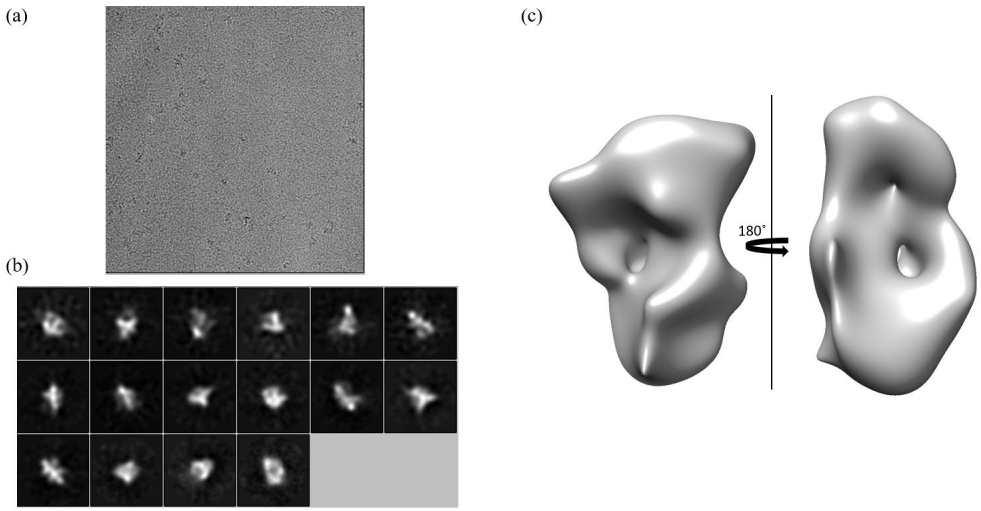

Figure 3.3. Preliminary structural characterization of D135.

(a) Representative micrograph of D135.
(b) Selected 2D class averages of ~50,000 aligned particles.
(c) 3-D reconstruction of the D135 consisting of ~30,000 particles.

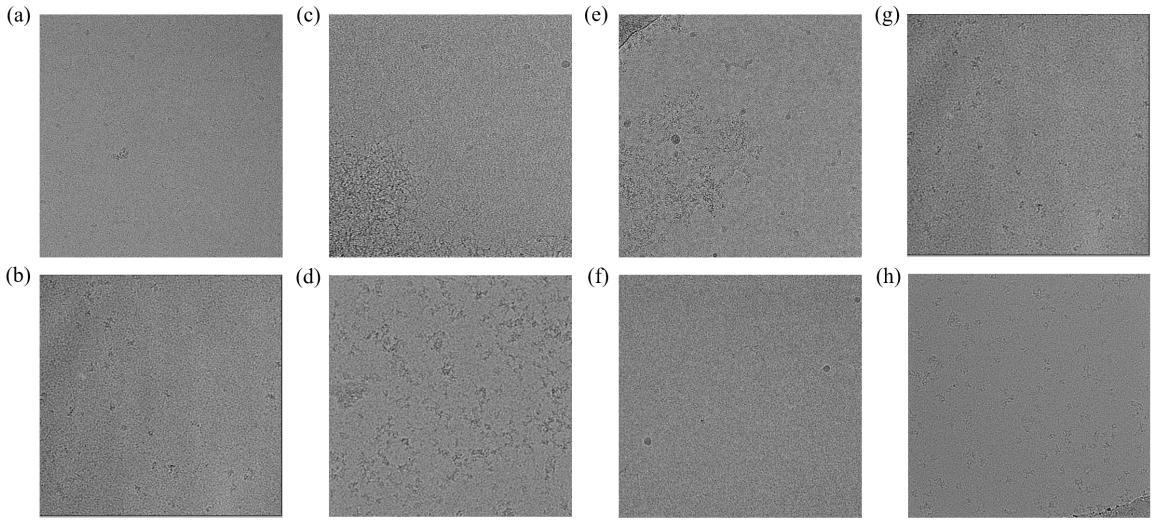

Figure 3.4. Optimizing cryo-freezing conditions.

(a) Representative particle distribution using Quantifoil holey carbon grids. Hole sizes of R 0.6/1 μm, R 1.2/1.3 μm, and supports of Cu, Ni and Au were varied.
(b) Representative particle distribution using C-flat holey carbon grids. Hole sizes of R 0.6/1 μm and R 1.2/1.3 μm, and supports of Cu, Ni and Au were varied.
(c) Representative particle distribution using UltrAUfoil holey carbon grids R 1/2/1.3 μm, 300 mesh, Au support grids with non-ionic detergents Triton X-100 and Tween-20.
(d) Representative particle distribution using UltrAUfoil holey carbon grids R 1/2/1.3 μm, 300 mesh, Au support grids with ionic detergent sodium deoxycholate.
(e) Representative particle distribution using Pelco lacey carbon, R 63 μm, 300 mesh, Cu support.
(f) Representative particle distribution using Quantifoil with 2nm ultrathin carbon R 2/2 μm, 300 mesh, Cu support.
(g) Representative particle distribution using UltrAUfoil grids. Hole sizes of R 0.6/1 μm, R 1.2/1.3 μm and R 2/2 μm, and mesh sizes of 200 and 300 were varied.
(h) Representative particle distribution using reverse side blotting technique on Quantifoil grids R 1.2/1.3 μm, 300 mesh, Au support.

3.2.3 Krios dataset collection

After obtaining well-distributed, uniform particles reliably and reproducibly, I collected a single particle cryo-EM dataset of the D135 RNA using a 300 kV FEI Titan Krios electron microscope equipped with a K3 camera (Gatan). A general overview of the data processing is outlined in Figure 3.5. All micrographs were processed through RELION3.0[119] and corrected for beam induced motion using MOTIONCORR2[120] followed by CTF estimation and correction using CTFFIND.[121] Thereafter, reference-free 2D class averages revealed the presence of high-resolution features and an overall shape consistent with negative stain EM. An initial 3-D model was generated and low pass filtered to 60 Å for subsequent 3-D classification steps. Upon inspection, two classes (class 1 and class 3) showed detailed secondary structural features, though some inherent heterogeneity was present (indicated by the green arrow in Figure 3.5). These two classes were subject to 3-D refinement and post processing each resulting in a 4.62 Å structure. Comparing both classes by superimposition of the final maps (Figure 3.6), revealed the core of the intron to be identical and the regions shown in blue arrows (Figure 3.5) in different conformational states. Perhaps, the presence of such inherent conformational dynamics within D135 explains why it was impossible to crystallize the RNA over the last three decades.

The particles from class 1 and class 3 were merged and further processed to obtain a final 3.62 Å reconstruction based on the gold standard Fourier shell correlation (FSC) at 0.143 (Figure 3.7a). Upon closer inspection, local resolution estimation (Figure 3.7b) revealed the core of the intron to be ~3.4 Å. This allowed for de-novo model building of all core helices.

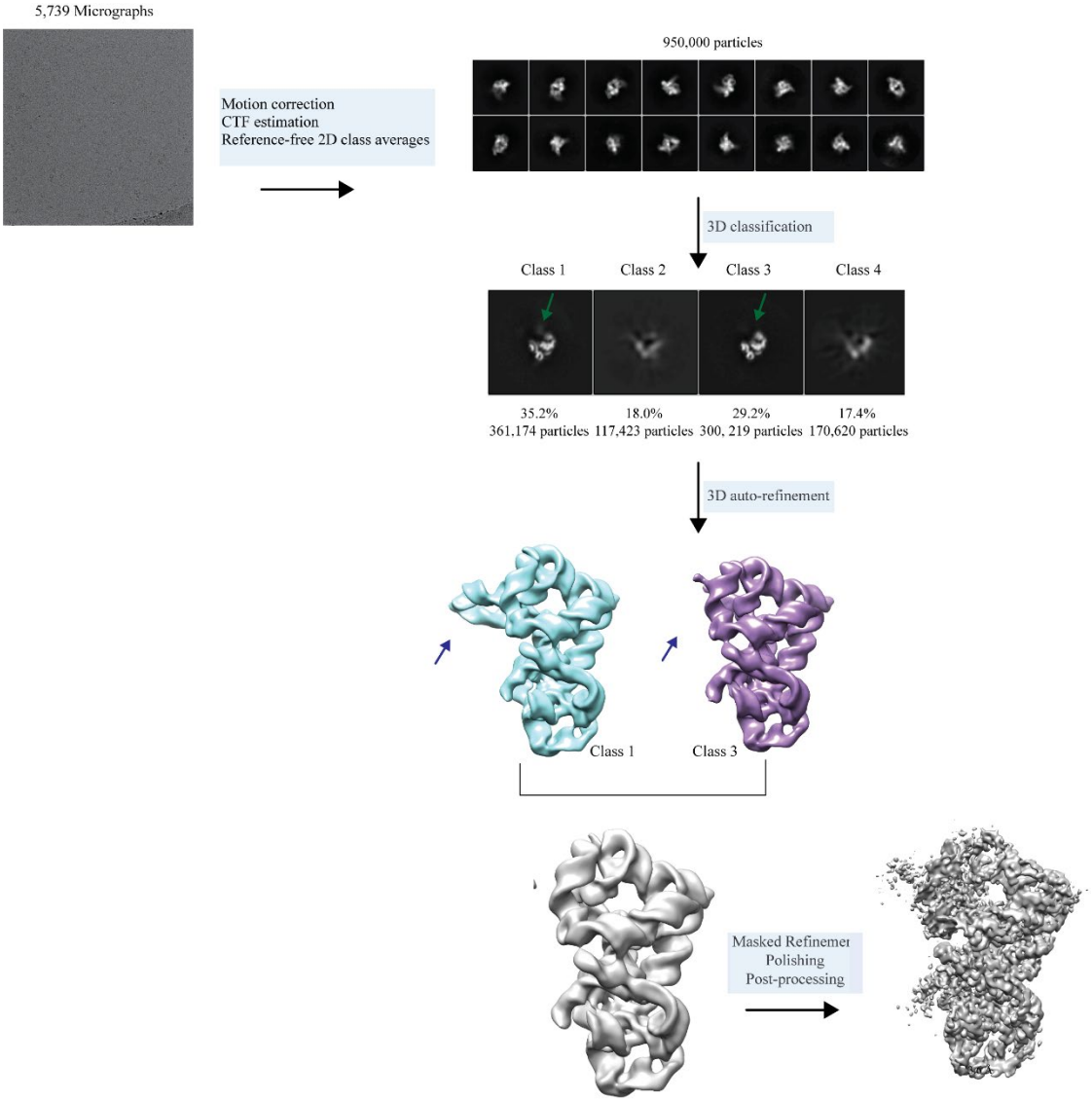

Figure 3.5. Cryo-EM data processing workflow.

(a) One data set was collected on a 300 keV Titan Krios equipped with a Falcon-II detector. Class 1 and class 3 were merged for a final 3.6 Å reconstruction.

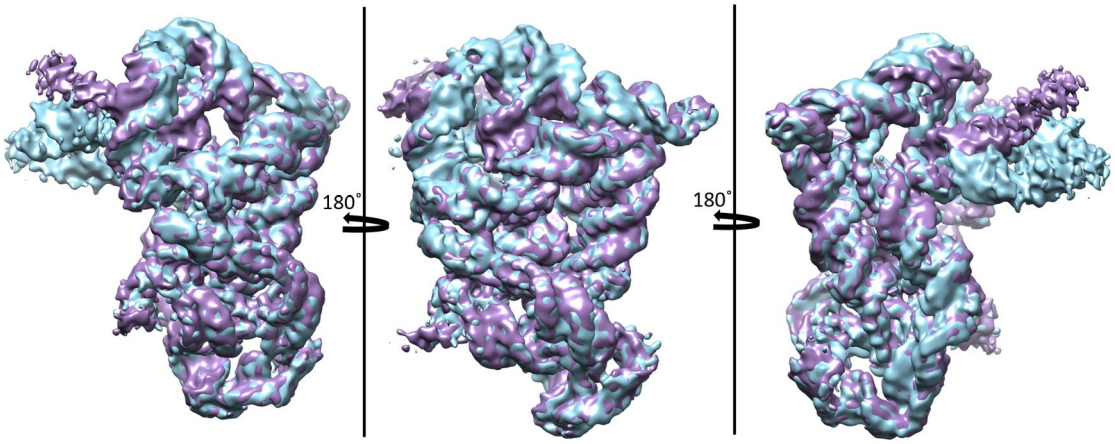

Figure 3.6 Comparison of class 1 and class 3 reconstructions.

(a) Superimposed reconstructions of class 1 (blue) and class 3 (purple)

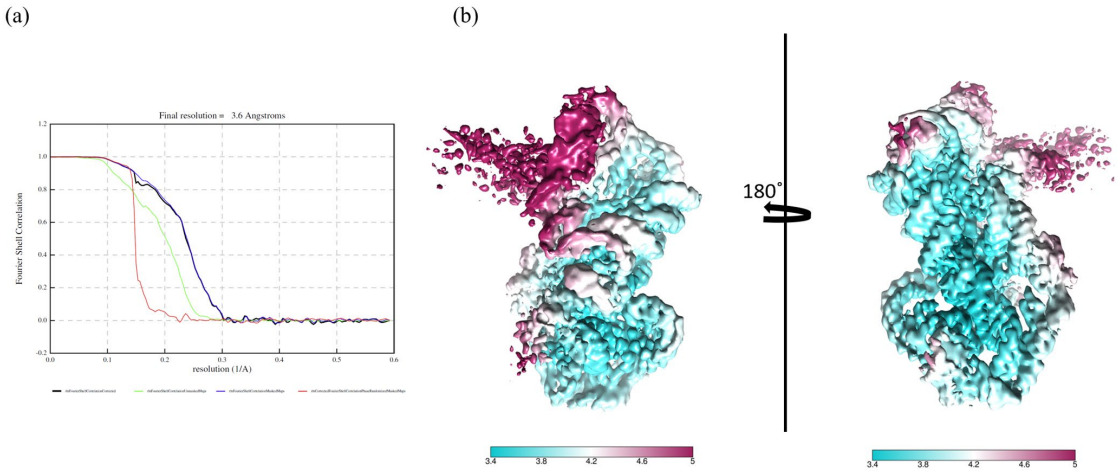

Figure 3.7. Final reconstruction resolution estimation.

(a) Fourier shell correlation curve.
(b) Local resolution map resolution displaying varied resolution from ~3.4 Å in the core of the intron to ~5 Å near the flexible surface.

3.2.4 D135 ribozyme architecture

This model represents the first high-resolution structure available for this intron. The cryo-EM map, colored according to the domain organization (Figure 3.8), revealed a characteristic group II intron tertiary architecture. Catalytic domain 5 (D5) resided in the center of the intron. This 34-nucleotide hairpin contained an asymmetric two nucleotide bulge and made an intricate network of tertiary contacts organizing active site components. D1 maintained its role as a scaffolding element by forming an upside-down V shape bend, positioning itself as a cap above D5. This allowed the large domain to make extensive contacts with the rest of the intron (Figure 3.8b). Compared to group IIC *O.i* intron, D3 formed an intricate, complex fold, acting as a brace to stabilize the base of the helix (Figures 3.9 and 3.10). Finally, D2 positioned itself beneath D5, extending downwards and sandwiching between D2 and D1, helix c1 (Figure 3.10).

The outlined domain folds were maintained by a plethora of intra- and inter-domain contacts throughout the intron. Positioning of D5 was made possible through a series of internal loops in D1, helix d1 that serves as a "folding control element." [122] One of the oldest predicted tertiary interactions was the ζ-ζ', where a terminal GNRA tetra loop in D5 formed a base stacking interaction with the bulge loop in D1, helix d1 anchoring D5 into the D1 scaffold (Figure 3.11a).

A four-way junction in D3 formed two tetra-loop interactions with D2 (τ - τ') and D5 (μ-μ') (Figure 3.10). This domain has been called the "catalytic effector" as its presence accelerated group II catalysis[123] but its molecular basis was not fully understood. Inspection of the current structure sheds light on this. Remarkably μ-μ' and κ- κ' interactions came together to form a five nucleotide A platform and position itself into the

minor groove of helix D5 through a series of base-triples and 2'OH interactions. (Figure 3.1b).

On the opposite side of the D3 junction are the τ and ρ interactions. In ρ-ρ', a bulge U base paired with an A located in the linker region between D1 and D2, and initiated the positioning of the terminal stem of D2 (Figure 3.10). Each side of the D2 helix is then stabilized through two tetraloop interactions with D1 and D3. The τ - τ' interaction involved a GAAA tetraloop interaction in D3 with a GA and GC base pairs in D2 (Figure 3.10). The θ–θ′ interaction involved a GUAA tetraloop in D1, helix c1, and two CG base pairs in D2 (Figure 3.10). Together these interactions positioned the D2 helix to extend down the intron, stabilizing the tertiary fold of the ribozyme (Figure 3.10).

The largest domain, D1 is transcribed and folded first to form the basic scaffold for organizing the rest of the domains. While this complex domain contained some essential inter-domain interactions discussed above, it also contained several intra-domain contacts that aid in positioning its nine helices. Observed in the structure is the ε-ε' interaction present in all classes of group II introns. Here two G-C base pairs (Figure 3.11c) and base stacking interaction between G5 and A570 in the 2nt bulge bring the 5' exon near the catalytic active site (Figure 3.11c).

The inherent flexibilities in the rest of the D1 helices (c2 and d1-d4) resulted in a low-resolution map; therefore, I was unable to build these regions. Current work is underway to try and resolve this region through computational approaches.[52,124]

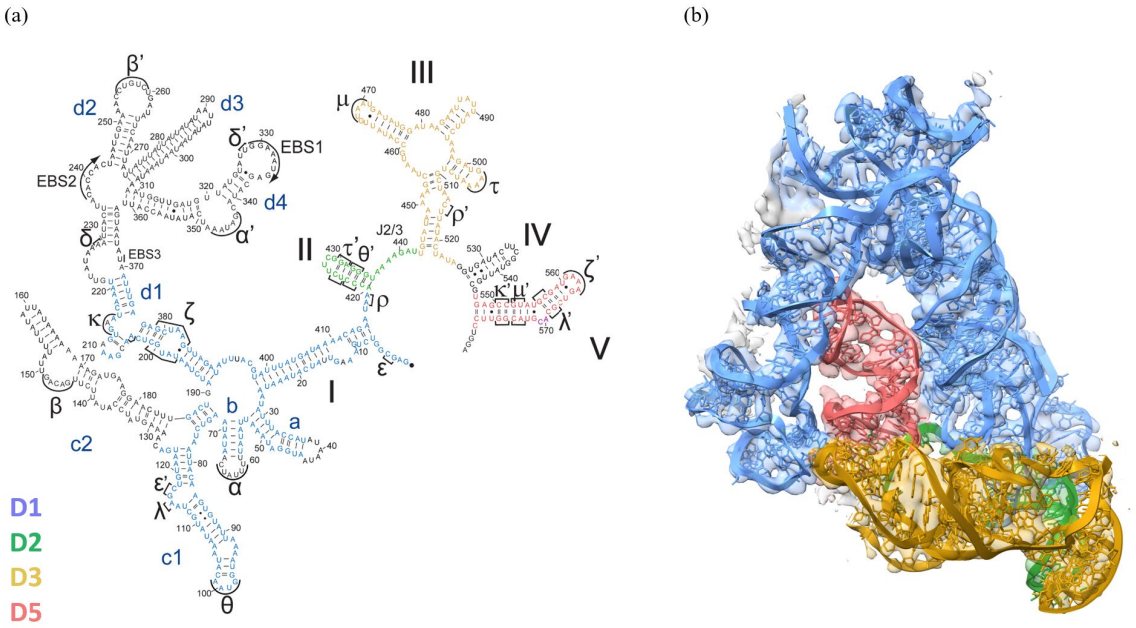

Figure 3.8. Overall structural architecture of D135.

(a) Secondary structure of D135 with domains colored according to legend. Black nucleotides indicate flexible region and were not modeled. Greek letters indicate sites of evolutionarily conserved tertiary interactions. EBS1 and EBS2 correspond to exon binding sites 1 and 2.
(a) 3-D model of D135 colored by domain.

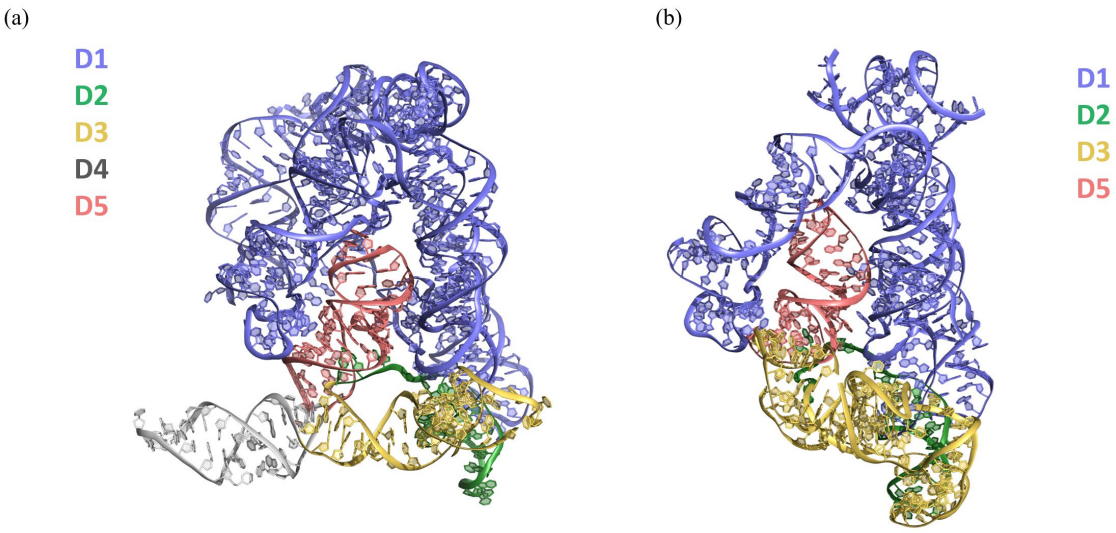

Figure 3.9. Comparison of group II introns.

(a) 3-D structure of group IIC *O.i.* intron. Adapted from PDB 4E8M.
(b) 3-D structure of group IIB D135.

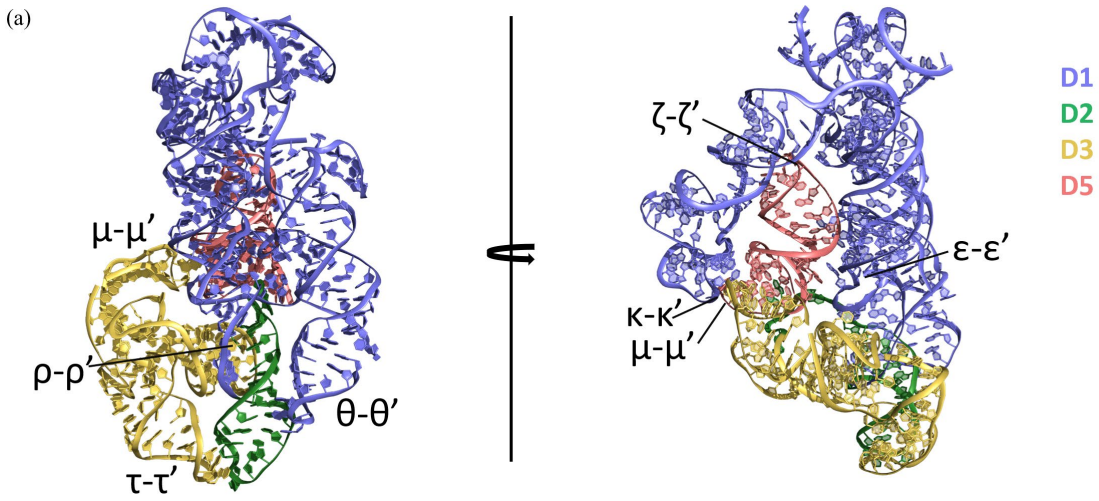

Figure 3.10. Mapping D135 evolutionarily conserved sites of tertiary interactions.

(a) 3-D structure of D135 with Greek letters indicating sites of evolutionarily conserved tertiary interactions.

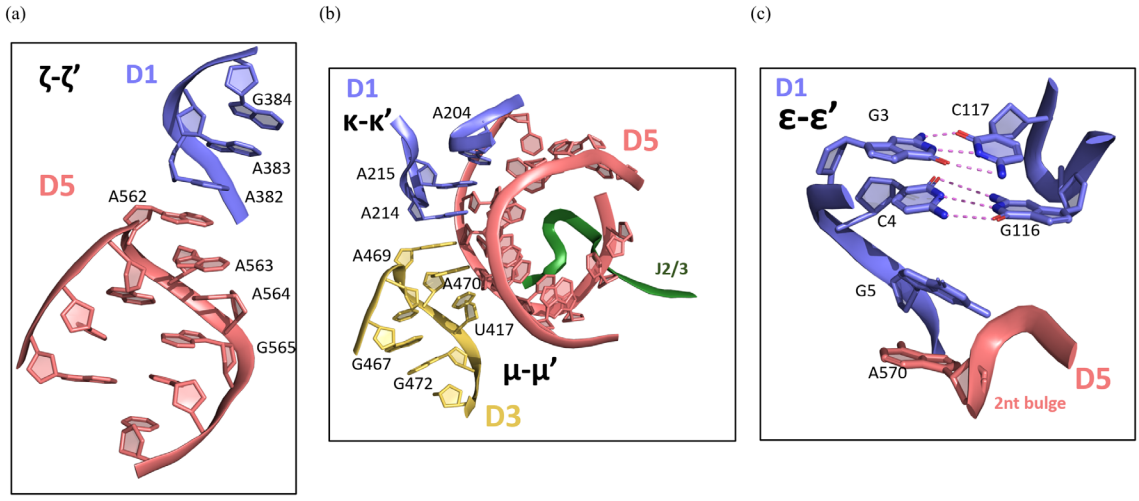

Figure 3.11. Essential tertiary interactions in D135.

(a) 3-D structure of the ζ-ζ' tetraloop base stacking interaction between D1 and D5.
(b) 3-D structure of the μ-μ' and κ-κ' interactions are forming a five base A platform.
(c) 3-D structure of the ε-ε' base pairing.

3.3 Discussion and future outlook

D135 ribozyme has been extensively well characterized biochemically, but its complete structural architecture remained elusive. With this structure, I finally visualized molecular-level details of interactions and understood how the interplay of these domains results in a catalytically active ribozyme.

The dynamic nature of helices in Domain 1 revealed in this structure explains why previous attempts to crystallize this RNA were in vain. These flexible helices play a key role in intron folding, exon recognition, and splicing. Therefore, current work is underway to capture D135 with a short, cis 5' exon bound to its binding site on the intron (EBS1 and EBS2) in D1. Capturing this stalled state will stabilize these flexible helices and reveal the full architecture of this ribozyme. Additionally, the current pipeline opens avenues for future studies to visualize the dynamic interplay and reorganization of key structural elements that catalyze each step in the splicing pathway of aI5γ.

Given the current interest in small molecule targeting of this intron and the development of anti-fungal agents, this model can now be used to perform a more targeted virtual small molecule screening and identify specific group IIB inhibitors. Current work is underway to try and solve this intron bound the previously published splicing inhibitor, Intronistat B.[16]

The workflow presented here demonstrates the underutilized power of cryo-EM in identifying functional structural elements in previously "un-crystallizable" classes of RNA. One of the largest challenges that has prevented this field from rapidly expanding is the bottleneck in grid preparation. By systematically testing a plethora of grid freezing

conditions, I have established an approach to reproducibly obtain an even distribution of particles in the holes. This pipeline can now be readily applied to study structure-function relationships in a diverse set of RNA systems, such as mRNAs or lncRNA, where we are just beginning to understand their structural landscapes. In the upcoming years, cryo-EM will transform our understanding of RNA structural biology.

3.4 Materials and Methods

In-vitro transcription and purification

AI5γ D135 plasmid (20 μg) was linearized and transcribed as previously described.[92] In brief, the construct was *in-vitro* transcribed at 37˚C for 2 hours using T7 RNA polymerase in a buffer containing 12mM $MgCl_2$, 40mM Tris-Cl pH8, 2mM Spermidine, 10mM NaCl, 0.01% Triton X-100, 10mM DTT, 2μl SUPERase-In and 3.6mM of each NTP. Thereafter, 4U of TURBO DNase was added, and the mixture was incubated at 37˚C for 30 min. To chelate excess divalent ions, 5μl of 0.5M EDTA was added. Transcription products were purified on a denaturing 5% polyacrylamide gel and eluted overnight at 4˚C in a gel elution buffer (10 mM MOPS-NaOH pH 6.0, 300 mM NaCl, and 1 mM EDTA). RNA was ethanol precipitated at -20˚C overnight. The next day RNA was spun at 15,000g for 20 min and washed with fresh 70% ethanol. The pellet was air dried for 10 min and resuspended in a monovalent buffer (50mM K-Mops pH7, 300mM KCl). The RNA was folded at the indicated concentrations (100nM-20μM) by heating up to 90°C for 1 min, followed by cooling at room temperature for 2 min. A final concentration of 75mM $MgCl_2$ was added and the mixture was incubated at 37˚C for 30 min. For the unfolded control, an equivalent amount of monovalent buffer was added.

Sedimentation velocity-analytical ultracentrifugation (SV-AUC)

SV-AUC experiments were performed with a Beckman XL-1 centrifuge and an AN-50 Ti Analytical Rotor. For each condition, 500μl of the RNA at an A260 of 0.6 was used. RNA was folded as described above. For each condition, 500μl of buffer was also prepared. AUC cells were cleaned thoroughly with RNase Zap and 100% ETOH, dried, and assembled according to the manufacturer's instructions. Using a round gel loading tip,

2x 210µl of buffer was added to the left side of the cell chamber. Similarly, 2x 195µl of RNA was added to the right side of the cell chamber. Cells were sealed according to the manufacturer's instructions, placed into the rotor, and equilibrated to 20˚C for 1.5-2 hours under vacuum. Thereafter, cells were spun at 25,000 rpm for 16 hours for a total of 120 scans. Data was analyzed in Sedfit using the continuous distribution model.[125,126] Hydrodynamic radii (R_h) were calculated in Sedfit using a partial specific volume of 0.53 cm^3/g and a hydration of 0.59 g/g.[127] Buffer viscosity and density were estimated using Sednterp and figures were made using Gussi c(s). [128]

Negative stain EM

For negative stain, EMS continuous carbon grids (300 mesh, CU) were glow discharged for 30 sec at 26mAmps, 0.36 mBar using a PELCO easiGlow Discharge Cleaning System. Immediately after, 4µl of 100nM RNA was added to the glow discharged surface of the grid and incubated for 40 seconds. The RNA was wicked away using Whatman paper and the grid was immersed into a 30µl drop of 2% uranyl acetate (UA) stain. After 10 seconds, the stain was wicked away and the grid was immersed in a second 30µl drop of 2% UA. After 30 seconds, the excess stain was once more wicked away and the grid was air dried for 2 min. Grids were imaged using a 120 keV Talos L120C (Thermo Fisher Scientific) at a magnification of ~45-60,000x and a defocus of -2µm.

Optimizing cryo-freezing conditions

Cryo-EM grids were prepared using a Vitrobot Mark IV (Thermo Fisher Scientific). Grids were glow discharged for 30-90 seconds at 26 mAmps and 0.36 mBar using a PELCO easiGlow Discharge Cleaning System. Aliquots of 3.5µl of RNA (concentrations varying from 100nM-20µM) were applied to the surface of the grid inside the Vitrobot chamber, equilibrated at 25˚C and 100% humidity. Blotting was done for 1.5-5 sec with a

force of -4 and using filter paper (Ted Pella) equilibrated in the chamber 20 minutes prior to starting the session. Grids were plunge frozen in liquid ethane chilled to -170˚C and stored in liquid nitrogen until further use. Preliminary screening was done using a 200keV Glacios (Thermo Fisher Scientific). Grid freezing was tested with a variety of companies: Quantifoil holey carbon (varied hole size, mesh, and support), C-flat (varied hole size, mesh, and support), Pelco lacey carbon, (hole size of R ~63 μm, 300 mesh, Cu support), Quantifoil with 2nm ultrathin carbon (hole size R 2/2 μm, 300 mesh, Cu support) and size and UltrAuFoil (varied hole size and mesh).

For detergent screening, RNA was folded as described above, and thereafter 0.01-0.5% (w/v) of detergents (indicated below) were added to the sample right before blotting and plunge freezing. The following detergent concentrations were tested: CHAPSO (final concentration 0.2% -0.5% (w/v)), NP-40 (final concentration 0.005-0.02% (w/v)), Sodium deoxycholate (final concentration 0.04- 0.17% (w/v)), Triton X-100 (final concentration 0.0- 0.1% (w/v)), Tween-20 (final concentration 0.0-0.1% w/v)). For reverse side blotting, Quantifoil grids with a hole size of 1.2/1.3 μm, Au support, and 300 mesh were glow discharged carbon side facing up. However, the RNA was added to the Au support size, blotted, and plunge frozen as described above.

Cryo-EM data acquisition and processing

All micrograph data were collected using a 300keV Titan Krios microscope (Thermo Fisher Scientific) equipped with a Bio Quantum filter (Gatan) and a spherical aberration of 0.01mm. The micrographs were collected with a Post-GIF K3 camera (Gatan) in super-resolution mode at a magnification of 81,000 and a pixel size of 0.844 with a total dose of 50 e^- per $Å^2$ per sec fractionated over 50 frames. Data were collected using Serial

EM with a defocus range of -0.8 μM to -2.5 μM. A total of 5,844 micrographs were collected. The data were processed as described above, and ~600,000 particles were subject to auto-refinement, CTF refinement, Bayesian polishing, and post processing resulting in a 3.6 Å map. All model building was performed in Coot[129] and refinement in Phenix.[130] All figure images were created in ChimeraX[131] or PyMOL Molecular Graphics system, Version 2.0 Schrödinger, LLC.

4 Long non-coding RNA, Pnky adopts an intricate tertiary architecture

4.1 Introduction

Long non-coding RNAs (lncRNAs) are RNA transcripts over 200 nucleotides in length that contain no protein-coding potential. These RNAs are implicated in playing functional roles in many cellular processes and human diseases.[132-134] They can regulate gene expression by recruiting protein factors to modulate RNA splicing,[135] nuclear organization[136] and chromosome remodeling[137-139] or even act as a sponge to sequester away proteins and regulate degradation.[140-142] While significant progress has been made to understand their secondary structure,[26,143-150] we are only beginning to understand their tertiary landscape.[49,88,89,151]

One of the most limited areas in understanding the role of lncRNAs is in neuronal development. Given that many lncRNAs are dysregulated in neurological disorders,[152,153] recent work has begun to characterize functional lncRNAs in neurogenesis using a systematic approach.[154-156] In these studies, *Pnky* was identified as an evolutionarily conserved, trans-acting, nuclear lncRNA involved in repressing V-SVZ NSCs (ventricular-subventricular zone neuronal stem cells) differentiation into neuronal cells.[157] RACE cloning demonstrated Pnky to be an 825 nucleotide RNA encoded from three exons.[158] Robust genetic studies revealed knockdown of Pnky by shRNAs or conditional knock-out in cultured NSCs increased differentiation of neurons by 3-5-fold without perturbing expression of nearby by locus *Pou3f2*.[158] Furthermore, expression of Pnky from BAC transgene rescued the deletion phenotype indicating trans expression of the RNA can

regulate development.[157] Given the essential role of this lncRNA and strong genetic evidence, further insights into its mechanism of action require a structural characterization. Towards this, it remains unknown whether this lncRNA has any stable secondary structure elements and whether it can form higher-order structures that can, alone or in concert with protein partners, regulate its function.

Therefore, to better understand the structure-function relationship of Pnky, I used an integrative structural biology approach. I first *in-vitro* transcribed and purified full length Pnky. Then, I determined ionic conditions for folding Pnky into a homogenous, monodisperse, and compact form. Thereafter, I combined SHAPE-MAP and Tb-seq to understand Pnky's secondary and tertiary structural landscape. Finally, I visualized its 3-D architecture by cryo-EM. Remarkably, Pnky contains an intricate three-dimensional architecture consisting of compact structural modules comparable to "well-folded" ribozymes. These novel findings provide a road map of candidate riboregulatory motifs for targeted functional investigation of Pnky's role in neurogenesis. This comprehensive approach can be extended to explore many other lncRNA structure functional relationships. This work was done in collaboration with Rafael Tavares and Parna Saha.

4.2 Results

4.2.1 Pnky adopts a compact, stable monodisperse state after folding

For structural studies on a large RNA, it is important to establish conditions that result in a well-folded, homogenous distribution of RNA molecules. Towards this, I used a semi-native *in-vitro* transcription and purification protocol that preserves the co-transcriptional secondary structure.[159] To monitor the degree of molecular compaction as a function of magnesium concentration, I performed sedimentation-velocity analytical ultracentrifugation (SV-AUC) experiments. SV-AUC is a powerful method to quantitively assess and compare macromolecular size, shape, and interactions in a variety of ionic solutions.[114] In the case of Pnky, as $MgCl_2$ was titrated from 0mM-100mM, the sedimentation coefficient of Pnky increased and UV absorption decreased indicating an overall global compaction of the molecules[160] (Figure 4.1a). Importantly a single, monodisperse peak was present at low $MgCl_2$ concentrations, signifying a conformationally homogenous species (Figure 4.1a). Representing (R_h) of the RNA as a function of $MgCl_2$ concentration and fitting the generated curve to a hill equation[159] resulted in a $K_{½\,Mg}$ of ~4.9 mM (Figure 4.1b). This value is comparable to that of lncRNA RepA[144] and is lower than that of lncRNA HOTAIR[145] and aI5γ group IIB intron.[110] Taken together, I determined that 15mM $MgCl_2$ (3x the experimentally determined $K_{½\,Mg}$) is sufficient to produce a fully folded, compact RNA species which can be used to perform subsequent probing and structural studies.

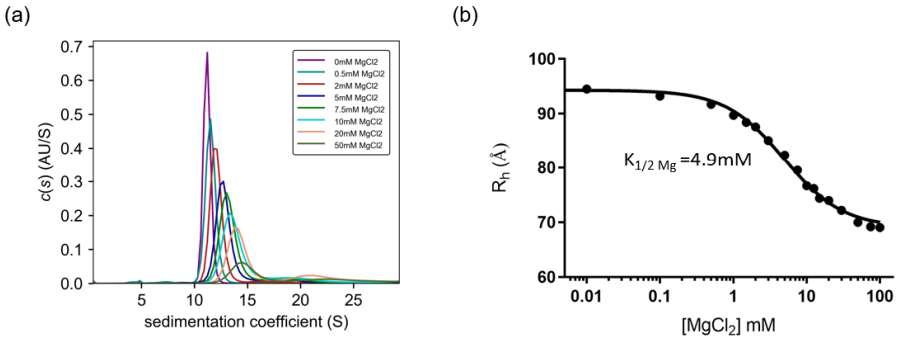

Figure 4.1. Pnky compacts to a homogenous, monodisperse species upon the addition of $MgCl_2$.

(a) SV-AUC profiles of Pnky folded in increasing $MgCl_2$ concentrations.
(b) Hill plot of the hydrodynamic radii (Rh) obtained from the SV-AUC profile. $K_{1/2\ Mg}$ indicates the concentration of $MgCl_2$ at which 50% of the RNA is folded.

4.2.2 Evaluating the *in-vitro* SHAPE-MAP data

Having established folding conditions, I first examined Pnky's structure *in-vitro*, in the absence of protein components. To obtain a secondary structure model, I implemented the SHAPE-MAP strategy (selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling) and probed with the SHAPE reagent 1M7. This electrophile selectively acylated flexible 2'OH (single-stranded regions) depositing a bulky adduct. These adducts were then recorded as mutations during specific reverse transcription conditions and read out by HTS technologies.[21] Downstream computational processing of these mutations resulted in a reactivity profile that was used as constraints to improve structure prediction accuracy.[161]

Overall, two high-quality datasets were obtained with an average read depth >50,000x and effective reactivity data for 97.8% (763/780) of nucleotides in Pnky. To evaluate the SHAPE-MAP data quality, relative mutation rates of 1M7 treated and DMSO treated samples were compared. In both replicates, the modified mutation rate was significantly higher than that of the untreated samples (Figure 4.2a; p-value <0.001). The Pearson correlation coefficient across both replicates was 0.94, indicating a highly reproducible dataset (Figure 4.2b).

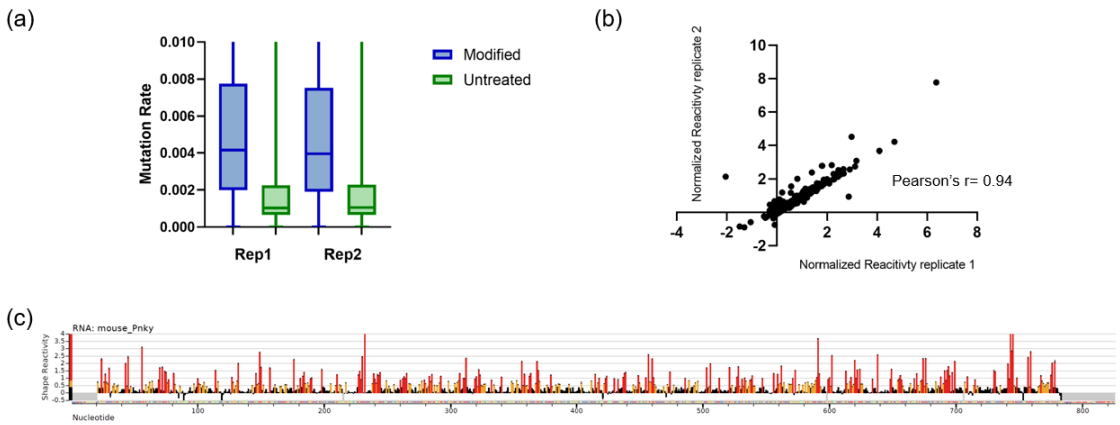

Figure 4.2. SHAPE reactivity profile of *in-vitro* Pnky.

(a) Mutation rate plot of two replicates. Box indicates interquartile range and line indicates average. Values outside this range are not shown.
(b) Correlation plot of normalized SHAPE reactivities from two replicates.
(c) Representative SHAPE reactivity profile. Red bars indicate high reactivity >0.85, orange bars indicate medium reactivity 0.4-0.84, and black bars indicate low reactivity <0.4.

4.2.3 *In-vitro* secondary structure of Pnky reveals the presence of highly structured modules

Before building the secondary structure model of Pnky, a global pseudoknot prediction was performed using ShapeKnots. Pseudoknots are formed through the pairing interactions between single-stranded nucleotides in one loop region to a complementary region in the RNA. These stable tertiary structures have been found in catalytic RNAs[162,163] and regulatory elements of viral RNAs[164] and play diverse functional roles in biological processes.[165] Incorporation of these nucleotides as hard constraints in structure prediction further improved the accuracy of the secondary structure model. One pseudoknot between nucleotides 151-156 and 269-274 was predicted and further inspection of experimental SHAPE reactivities supported its presence (Figure 4.2c). After incorporating the pseudoknot constraints, the full structure of Pnky was predicted using Superfold[21] (Figure 4.3a).

Initial inspection of the secondary structural map indicated Pnky was a highly structured RNA consisting of 7 modules (Figure 4.3a). The numerous long stem structures present in these modules, could come together to form a higher order three-dimensional architecture. In particular, modules 5-7 radiated from a central-wheel like structure (Figure 4.3a) and were reminiscent of the group II intron secondary structure organization. Interestingly this wheel-like structure falls in the evolutionarily conserved regions between mouse Pnky and human PNKY (Figure 4.3b). Additionally, the terminal regions of modules 5 and 6 formed dumbbell-like structures could serve as a three-way junction to recruit cellular factors.

The predicted kissing loop pseudoknot was located in the terminal regions of modules 2 and 4 and could be recognized by a cellular protein or allow for the proper 3-D organization of downstream helices. Interestingly, Module 2, which contains one side of the pseudoknot, shares sequence conservation with human PNKY (Figure 4.3b), implicating that this human lncRNA may also contain functional tertiary structures in its genome.

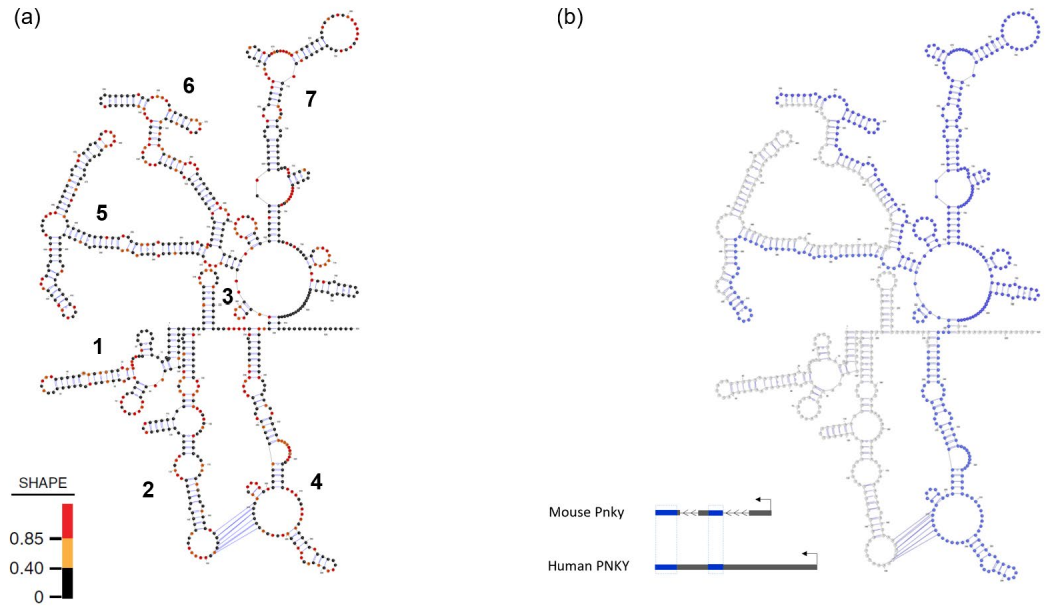

Figure 4.3. Secondary structure of Pnky reveals a highly structured, modular organization.

(a) *In-vitro* secondary structural map of Pnky colored by SHAPE reactivities. The predicted pseudoknot is indicated by blue lines. Modules are numbered 1-7.
(b) *In-vitro* secondary structural map of Pnky colored by sequence conserved nucleotides (blue).

4.2.4 Optimizing the *in-cellulo* probing of Pnky

While the high-quality *in-vitro* map was a start in understanding Pnky's secondary structure, it was necessary to obtain a complementary map that captured Pnky's structure in a cellular context. However, the low abundance of Pnky and limited passaging of the NSCs posed several challenges. During RT, the low copy number of Pnky molecules would result in an insufficient capture of the landscape of modified sites and lead to under-sampling of the data. To combat this, three approaches were used to try and maximize the number of Pnky molecules in the RT reaction.

First, these cells were cultured and expanded to their maximum potential, allowing for ~70-90 million cells per probing condition. Second, after probing, these cells were nuclear enriched, removing the pool of cytoplasmic RNA. Third, the extracted nuclear RNA was then depleted of rRNAs, as removing these abundant species has previously improved the detection of low abundance transcripts.[166] These efforts lead to an input of ~800-1000ng of RNA for each RT condition. Thereafter, library preparation, sequencing, and processing were performed as described above. The resulting profile displayed promising results at the 3' end of Pnky (Figure 4.4a); however, upstream regions often contained mutation rates similar to the background DMSO control. This indicated a need to further increase the copy number of Pnky during RT. Therefore, we switched to a cell line that expressed *Pnky* from a BAC transgene and contained 2-5x more copies of the RNA compared to WT NSCs. Following the enrichment procedure and library preparation outlined above, a preliminary profile containing global mutation rates just greater than the background control was obtained (Figure 4.4b).

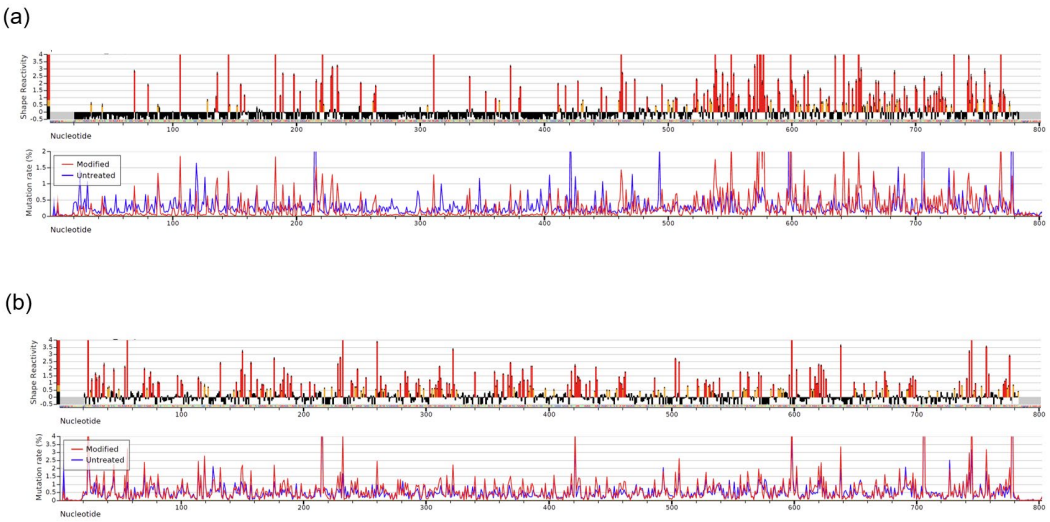

Figure 4.4. Optimizing in-cellulo SHAPE reactivity profiles.

(a) Representative SHAPE reactivity profile and mutation rate for modified (1M7 treated) and untreated (DMSO) in WT NSCs.
(b) Representative SHAPE reactivity profile and mutation rate for modified (1M7 treated) and untreated (DMSO) in NSCs expressing BAC *Pnky* transgene.

4.2.5 *In-cellulo* probing of Pnky secondary structure reveals a similar, highly structural organization

A preliminary map of *in-cellulo* Pnky was generated (Figure 4.5). Overall, I observed a high degree of structural organization consisting of 7 structural modules (Figure 4.5). Consistent with the *in-vitro* structure, many of these modules contained long stems which may compact to form a higher-order 3-D fold.

Upon inspection of the structure, I observed many modules, such as 1, 2, and 3, to be nearly identical in both maps. The high degree of structure conservation suggested the presence of a robust platform that may recruit and bind splicing factor PTBP1 (current work by Parna). I next inspected the shape reactivities near the predicted pseudoknot between modules 2 and 4. Interestingly, these reactivities supported the presence of this pseudoknot in NSCs. A similar kissing loop type pseudoknot was predicted in lncRNA MEG3 and mutagenesis studies revealed this interaction to play a critical role in the p53 regulation.[167] Current mutagenesis work is underway to test the function of this tertiary structure in neurogenesis.

De-novo prediction of an *in-cellulo* RNA can often result in a convoluted model as interactions with cellular factors such as proteins could affect the reactivity profile as to what is “single-stranded” or “double-stranded.” Therefore, I performed a second analysis approach to compare reactivity differences between cell-free and in-cell RNA. This ΔSHAPE analysis has been successfully used to understand protein occupancy by identifying nucleotides on an RNA that recognize a protein.[24]

This analysis (Figure 4.6a) revealed several conformational changes across the RNA. My previous work implied that consecutive nucleotides displaying differential reactivity profiles often confer functional specificity. Therefore, I scanned the RNA and highlighted one region between nucleotides 685 and 700 that became less reactive in cells (Figure 4.6, top insert). Inspection of the *in-vitro* map, showed these nucleotides resided on a terminal stem-loop (Figure 4.6 bottom insert), implicating this region to be recognized by a protein or involved in a protein-mediated RNA-RNA interaction (both of which would result in the loss of reactivity seen in the *in-cellulo* profile). Current work is underway to produce a second biological replicate and draw more definitive conclusions.

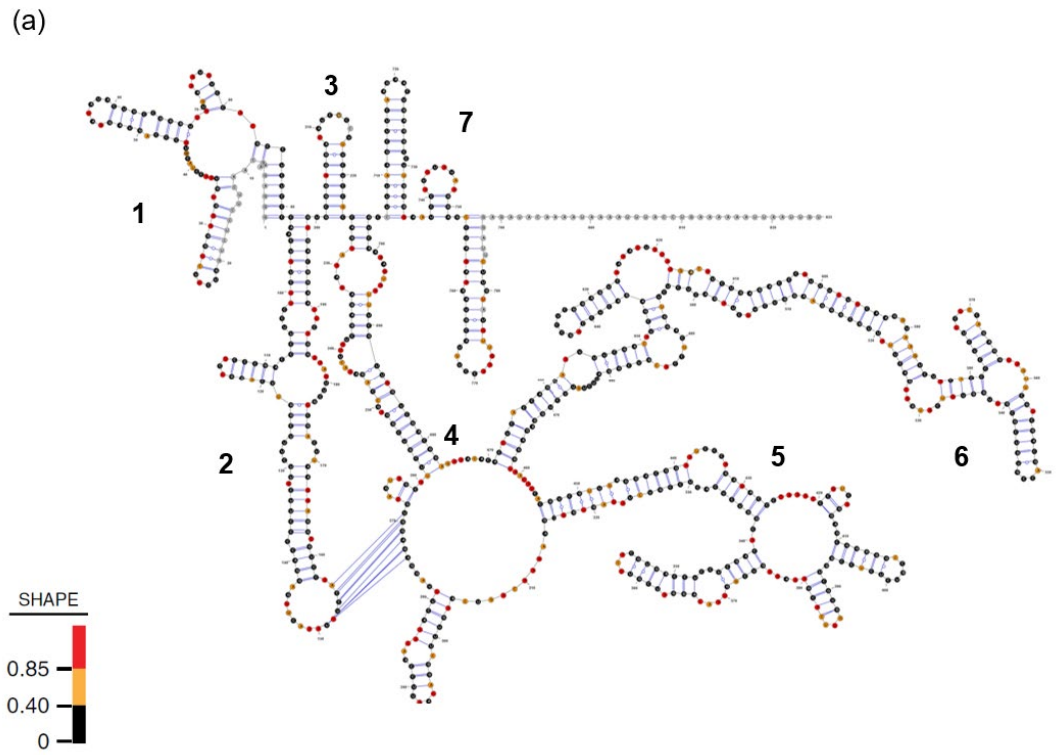

Figure 4.5. Secondary structure of *in-cellulo* Pnky.

(a) *In-cellulo* secondary structural map of Pnky colored by SHAPE reactivities. Blue lines indicate predicted pseudoknot. Modules are numbered 1-7.

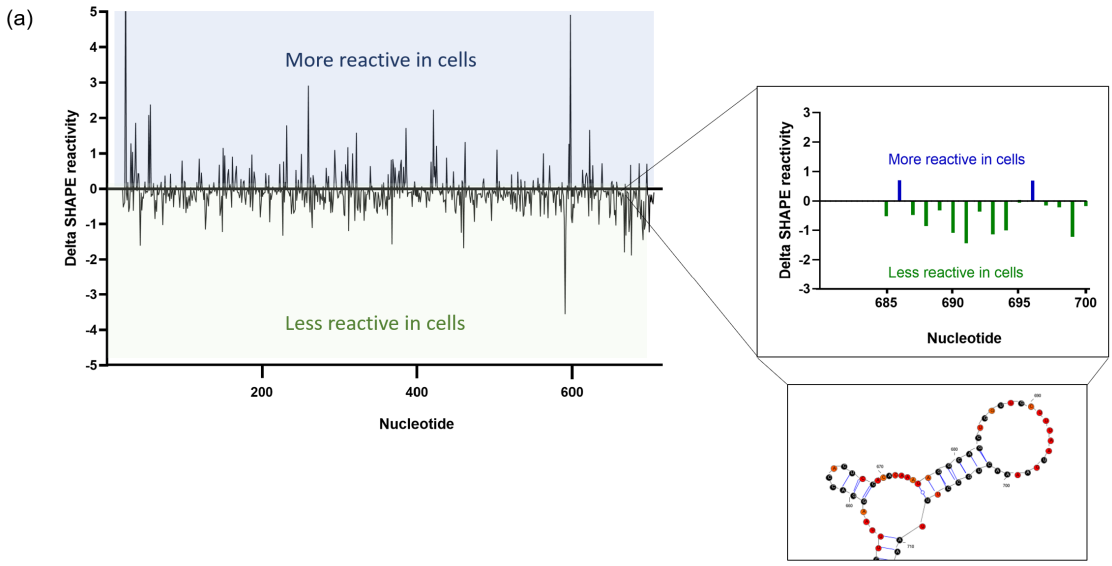

Figure 4.6. ΔSHAPE analysis.

(a) ΔSHAPE reactivity profile. The top insert displays the ΔSHAPE profile between nucleotides 685 and 700. The bottom insert displays the corresponding region in the *in-vitro* structural map.

4.2.6 Tb-seq analysis shows *in-vitro* Pnky tightly coordinating with metal ions forming local, compact regions of higher order RNA structure.

The secondary structural mapping provided an initial start in determining the structuredness of this RNA, but often the formation of higher-order RNA structures confer functionality. Therefore, I used two complementary approaches, Tb-seq and electron microscopy, to understand the tertiary architecture of Pnky.

One hallmark of tertiary structure formation is the ability of RNAs to coordinate with divalent ions such as Mg^{2+}, allowing them to compact and form a complex three-dimensional architecture. Previous work with AUC provided an initial indication that Mg^{2+} induces compaction of Pnky on a global level. I employed Tb-seq to further elucidate stable tertiary motifs and aid in designing more targeted mutants for functional assays.

Overall, Pnky displayed numerous sites of strong Tb-seq signal located throughout the identified modules (Figure 4.7). This indicated that Pnky compacted into a globular-like structure (similar to what we observe with folding group II introns). This was quite surprising as my previous work with lncRNA RepA revealed no strong Tb-seq signal profile (unpublished). Upon closer inspection, I observed many of the sites associated with strong Tb-seq signal localizing to stem-loop or bulge-loop regions in the conserved mouse Pnky and human PNKY sequences (Figure 4.7), thus implicating the importance of these f these structural motifs in neurogenesis. Importantly, Tb-seq signal was present near the predicted kissing loop in modules 2 and 4 (Figure 4.7). This suggested a 3-D organization consisting of sharp bends in the RNA backbone that orient the pseudoknot base pairing interaction and further supported this pseudoknot to play a critical functional role in neurogenesis. Current work is underway to produce a second replicate. Together, these

secondary and tertiary probing strategies provided a roadmap to pinpoint and prioritize regions to test functional significance.

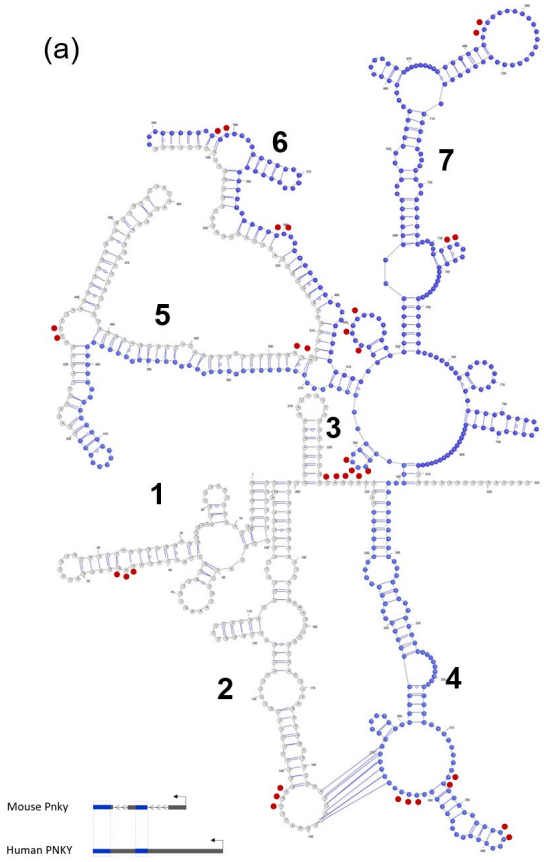

Figure 4.7. Tb-seq probing *in-vitro* Pnky.

(a) *In-vitro* secondary structural map of Pnky colored according to conserved regions (blue). Surrounding red dots indicate regions of Tb^{3+} cleavage.

4.2.7 Negative stain of Pnky

To directly visualize the three-dimensional architecture of the RNA, Ananth and I employed negative stain electron microscopy. Negative stain EM is a simple and rapid way to assess the size and homogeneity of an RNA and evaluate low-resolution features. I acquired images of Pnky folded at 0mM, 10mM, and 15mM $MgCl_2$. However, during imaging, it was difficult to find regions of stain and thus particles on the grid, suggesting that the RNA did not like the acidic nature of the stain. Subsequent attempts to change the stain from the standard 2% uranyl acetate to 2% uranyl formate resulted in a similar staining pattern.

Upon closer inspection of the micrograph of "unfolded" Pnky (Figure 4.8a), I identified well-separated particles, with some having a characteristic V-shape bend. As expected, in the absence of $MgCl_2$ there was a heterogenous population of RNA. Surprisingly, in the micrographs where Pnky was folded at 10mM and 15mM $MgCl_2$, some globular-like particles were present, but many particles were elongated/aggregated (Figure 4.8b and 4.8c). It was possible that the addition of the acidic stain or adherence to the carbon grid caused the RNA to unfold. However, given the strong biophysical and probing data highly suggesting the presence of a homogenous population of compact RNAs, I decided to try and visualize these particles by cryo-EM.

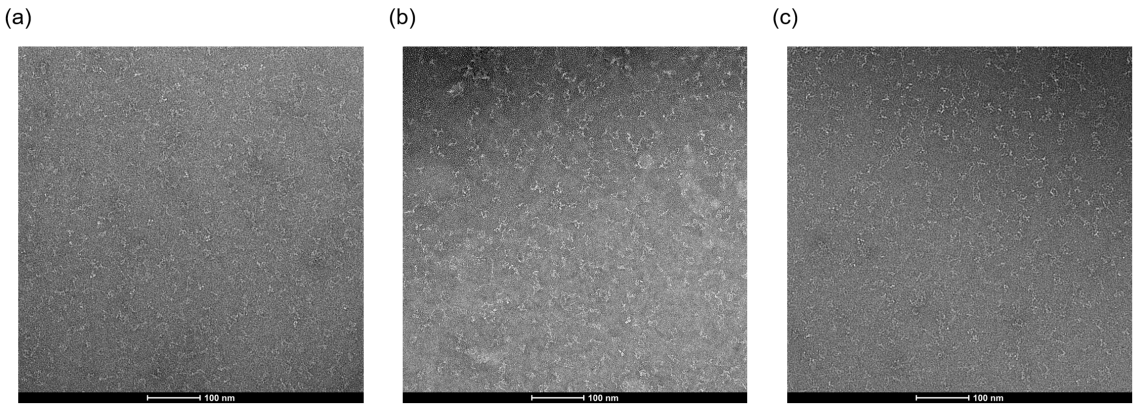

Figure 4.8. Negative stain EM of *in-vitro* Pnky.

(a) Representative micrograph of Pnky folded in 0mM $MgCl_2$.

(b) Representative micrograph of Pnky folded in 10mM $MgCl_2$.

(c) Representative micrograph of Pnky folded in 15mM $MgCl_2$.

4.2.8 Cryo grid freezing optimization

Implementing the workflow previously established with the group II intron, we first optimized freezing conditions to obtain a consistent, even distribution of particles in a thin layer of vitrified ice. The use of UltrAuFoil grids resulted in little to no particles present (Figure 4.8a). The reverse side blotting technique that was previously successful resulted in a few single particles with approximate sizes of ~12-16 nm (Figure 4.9b). Therefore, I further tried to increase the number of particles in these holes by increasing the concentration of the RNA. However, all efforts resulted in the presence of large aggregates. Finally, I decided to use commercial Quantifoil grids containing a thin layer of carbon over the holes. However, these grids also showed the presence of large aggregates consisting of unfolded, elongated RNA particles (Figure 4.9c). Upon discussion with Marc Llaguno, we decided to manually add the carbon film over the holes. These micrographs revealed the presence of globular, single particles of ~ 12-16 nm in size (Figure 4.9d). While the micrographs contained <50 particles and some aggregates were present, they were a promising start. Therefore, I performed an initial dataset collection to assess the 2-D alignment of the particles.

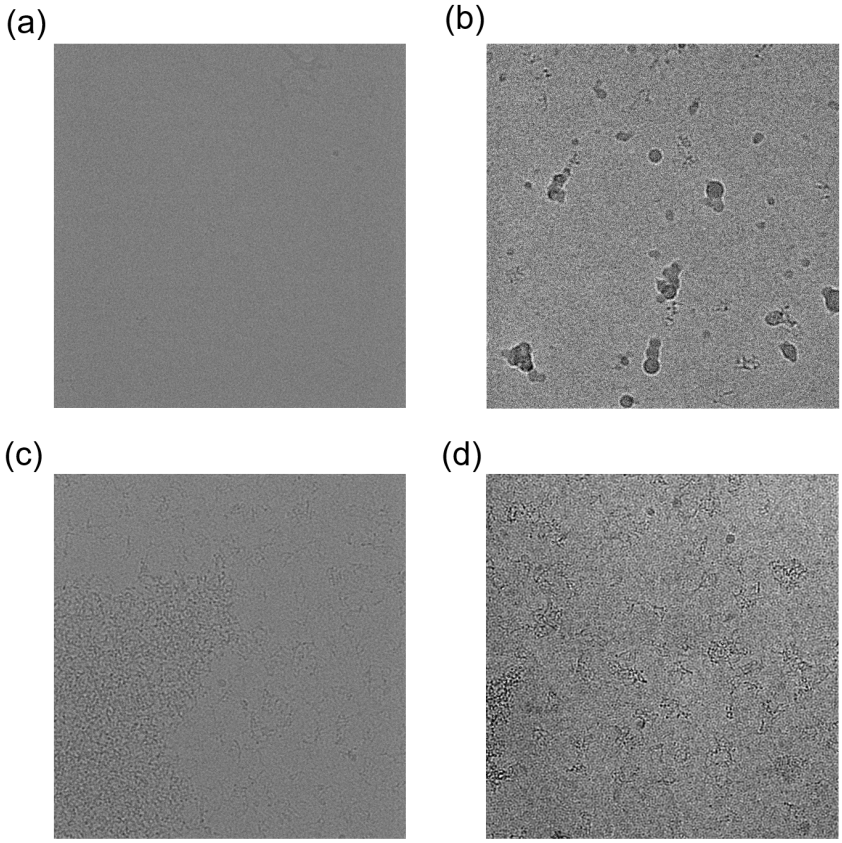

Figure 4.9. Optimizing cryo-freezing conditions.

(a) Representative particle distribution using UltrAuFoil holey carbon grids R 1/2/1.3 µm, 300 mesh, Au support.
(b) Representative particle distribution using reverse side blotting technique on Quantifoil grids R 1.2/1.3 µm, 300 mesh, Au support.
(c) Representative particle distribution using Quantifoil with 2 nm ultrathin carbon R 2/2 µm, 300 mesh, Cu support.
(d) Representative particle distribution using Quantifoil with 4 nm ultrathin carbon (added in house) R 1.2/1.3 µm, 300 mesh, Au support.

4.2.9 Preliminary cryo-EM reconstruction reveals Pnky adopting a globular, compact architecture

I collected data on a 300 keV FEI Titan Krios electron microscope equipped with a K3 camera (Gatan). A general overview of the data processing is outlined in Figure 4.9. All micrographs were processed through RELION4.0 and corrected for beam induced motion using MOTIONCORR2,[120] followed by CTF estimation and correction using CTFFIND[121]. Thereafter, I manually picked ~2000 particles and performed 2D class averaging. To my surprise, these class averages revealed well-aligned particles with unique structural characteristics and diverse distribution of particle orientations. These classes were then used to optimize auto-picking parameters across the ~4000 micrographs. Given the mixture of well-folded, globular particles and aggregate particles on the micrographs, the auto-picking software had substantial difficulties in picking the "correct" particle and would often pick elongated, aggregated particles. This resulted in poor 2D averages consisting of mostly noise and little well-aligned particles. Upon discussion with my colleagues, I realized this is a common issue in RELION's auto-picking algorithm and needed to use alternative approaches.

To further optimize the picking, I resorted to using a neuronal net-based algorithm, TOPAZ.[124] I first trained the program to select the "good" particles away from the aggregates and then implemented it to perform auto-picking across all ~4000 micrographs. These efforts resulted in promising 2D class averages with features indicative of a tertiary structural architecture. A repeating "butterfly"-like feature (Figure 4.10) suggested that this lncRNA contained sharp turns in the RNA backbone, and furthermore, in 3D space, these regions encapsulated around a central helix.

Thereafter, an initial 3-D model was generated and low pass filtered to 60 Å for subsequent 3-D classification. Upon inspection, class 2 showed promising low-resolution structural features (Figure 4.9, blue circle). This was subject to one round of 3-D refinement and post processing resulting in a ~15 Å structure (Figure 4.10). Importantly this reconstruction contained size features comparable to that predicted by AUC. Additionally, Tb-seq probing indicated a globally compacted RNA containing local regions of sharp turns in the backbone, which are distinctly visible in the reconstruction. This initial dataset collection and processing showed a promising path to obtaining a high-resolution reconstruction. However, my work with the group II intron indicated that ~500,000 particles were needed to obtain a high-resolution map. Therefore, current work is underway to enhance the number of globular Pnky particles on the micrographs.

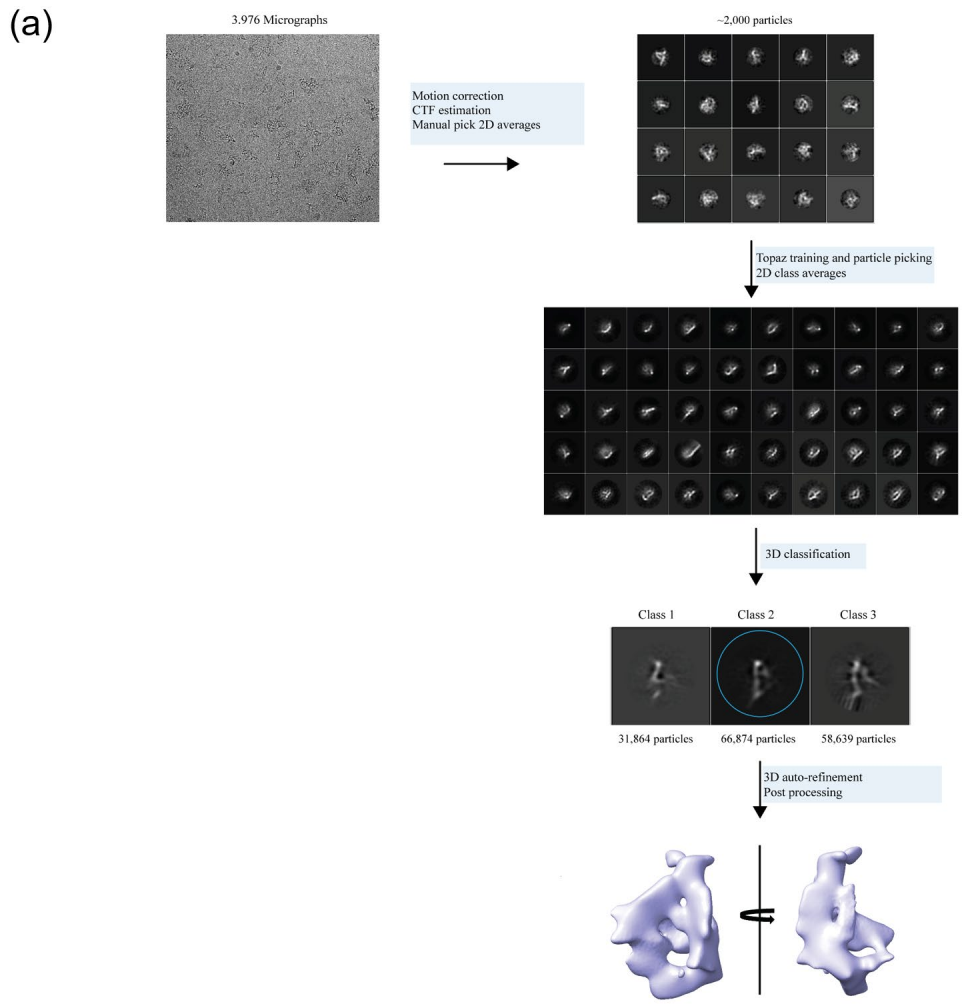

Figure 4.10. Cryo-EM data processing workflow.

(a) One data set was collected on a 300 keV Titan Krios equipped with a Falcon-II detector. Class 2 was used for the final ~15 Å reconstruction.

4.3 Discussion and outlook

As we expand our understanding of lncRNA roles in biological processes, systematic approaches are needed to assess their function. RNA structure plays a critical role in the classically studied RNAs, giving rise to the likelihood that many lncRNAs function through their structural architecture. Here, we present an integrative pipeline utilizing chemical probes and microscopy to study Pnky's structure and demonstrate its underlying importance in function.

Systematic characterization of Pnky *in-vitro*, revealed the presence of seven discrete structural modules and a pseudoknotted helix. Our efforts to characterize the secondary structure of this low abundant RNA in cells, demonstrated a remarkably similar structural organization in the modules and support for the predicted pseudoknot. Follow-up work is underway to assay the role of these structures in NSC differentiation. Mutational analysis disrupting the pseudoknot and candidate protein recognition sites will identify structure or sequence-specific motifs that mediate Pnky's role in neurogenesis. Our *in-cellulo* probing work with overexpressed BAC Pnky now allows us to this system and perform in-cell cross-linking and pull-down coupled with mass spectroscopy to identify Pnky's full protein interactome. Once a riboregulatory element is identified, complementary e-CLIP studies can be performed.

The tertiary structure analysis by Tb-seq and electron microscopy revealed Pnky to contain a globular, compact architecture with many regions of sharp backbone turns forming local structural motifs. Importantly, this data underscores that not all lncRNAs are flexible strings of nucleotides, some can form higher-order structures just like what we

observe in ribozymes. Future high-resolution structure determination work will unravel the 3-D architecture of Pnky and provide molecular-level details on how these structures may act as scaffolds to recruit cellular partners and stabilize complex formation.

Most of the work here focused on characterizing mouse Pnky secondary and tertiary structure. Applying this pipeline to characterize human PNKY secondary and tertiary structures will greatly enhance our understanding of Pnky's role in human biology. Furthermore, developing a functional assay for PNKY will allow for systematic investigation of candidate functional tertiary structures present in both species. Overall, studying the structure-function relationship in both systems will advance our understanding of lncRNA roles in biology and disease.

4.4 Materials and Methods

In-vitro transcription and purification

The *in-vitro* transcription and semi-native purification of Pnky was performed as previously described.[159] In brief, RNA was transcribed by runoff transcription using T7 RNA polymerase[93] in a buffer containing 12mM $MgCl_2$, 40mM Tris-Cl pH8, 2mM Spermidine, 10mM NaCl, 0.01% Triton X-100, 10mM DTT, 5µl SUPERase-In and 3.6mM of each NTP and incubated at 37˚C for 2 hours. Thereafter, 2U of TURBO DNase (2U/ul, Life Tech) was added, and the mixture was incubated at 37˚C for 30 min. To chelate excess divalent ions, 5µl of 0.1M EDTA was added. The RNA was exchanged into filtration buffer (50mM MOPS-KOH pH 7, 150mM KCl) using a 100-kDa Amicon Ultra filtration column at room temperature. The RNA was purified using a self-packed 24ml Sephacryl S-400 gel filtration column equilibrated in filtration buffer. RNA from the peak fraction was folded by incubating with the indicated $MgCl_2$ concentrations at 37˚C for 40 min.

Sedimentation velocity-analytical ultracentrifugation (SV-AUC)

SV-AUC experiments were performed with a Beckman XL-1 centrifuge and an AN-50 Ti Analytical Rotor. For each condition, 500µl of the RNA at an A260 of 0.6 was used. RNA was folded as described above. For each condition, 500µl of buffer was also prepared. AUC cells were cleaned thoroughly with RNase Zap and 100% ETOH, dried, and assembled according to the manufacturer's instructions. Using a round gel loading tip, 2x 210µl of buffer was added to the left side of the cell chamber. Similarly, 2x 195µl of RNA was added to the right side of the cell chamber. Cells were sealed according to the manufacturer's instructions, placed into the rotor, and equilibrated to 20˚C for 1.5-2 hours under vacuum. Thereafter, cells were spun at 25,000 rpm for 16 hours for a total of 120

scans. Data was analyzed in Sedfit using the continuous distribution model.[125,126] Hydrodynamic radii (R_h) were calculated in Sedfit using a partial specific volume of 0.53 cm^3/g and a hydration of 0.59 g/g. [127] Buffer viscosity and density were estimated using Sednterp and figures were made using Gussi c(s).[128]

In-vitro SHAPE probing

For probing, 2µg of RNA was incubated with 15mM $MgCl_2$ in a final volume of 20µl at 37˚C for 40 min. Thereafter, 18µl of folded RNA was transferred to a tube containing either 2µl of 100mM 1M7 (final concentration 10mM) or pure DMSO as a control and incubated at 37˚C for 10 min, followed by quenching with 2µl of 1M DTT. All RNA samples were purified using a Zymo RNA clean and concentrator kit according to the manufacturer's protocol and eluted in 15µl of nuclease-free water.

Cell culture and *in-cell* probing

SVZ NSCs (Pnky +/+ WT or Pnky +/+; BAC+) were grown in N5 growth medium: (DMEM/F-12 with GlutaMAX, supplemented with 5% fetal bovine serum, N2 supplement, 35 mg/mL bovine pituitary extract, antibiotic-antimycotic, 20 ng/mL epidermal growth factor, and 20 ng/mL basic fibroblast growth factor). Approximately 70-90 million cells were used for each probing condition.

For all probing conditions media was aspirated, cells were washed once with DPBS, and dislodged using Accutase. The cells were collected and centrifuged at 300g x 5min at 25˚C. The supernatant was removed, and to every ~10 million cells 900µl of DPBS was added. Cells were resuspended by pipetting up and down to ensure a homogenous cell suspension. A final concertation of 50mM 1M7 (freshly dissolved in DMSO) or an

equivalent volume of DMSO was added to the cells. The reaction mixture was incubated at 37˚C for 5 minutes and pelleted at 300g x 5min at 25˚C.

The resulting cell pellet was resuspended in nuclear fractionation buffer (10mM Tris-HCl, 140mM NaCl, 1.5mM $MgCl_2$, 0.5% (w/v) Igepal and 10mM Ribonucleoside vanadyl complex) and incubated on ice for 5-10 minutes. The mixture was centrifuged 1000g x 5min at 4˚C. The supernatant (cytoplasmic fraction) was discarded, and the procedure was repeated twice. In the final addition of the fractionation buffer, 0.5% (w/v) sodium-deoxycholate was added. The nuclear pellet was resuspended in Trizol and extracted using a QIAGEN RNeasy kit according to the manufacturer's protocol. The RNA was eluted in ME buffer (8mM MOPS pH 6.5, 1mM EDTA pH 8.5) and was ribosome depleted using a Ribominus kit according to the manufacturer's protocol.

MAP RT

For each probing condition, ~2000ng of IVT RNA or ~800-1000ng of nuclear enriched, ribodepleted RNA was mixed with 2 pmol of RT primer (Appendix, Table 1) in a final volume of 10µl and annealed at 90˚C for 1 min, followed by 30˚C for 2 min. Thereafter, 20U/µl of Superscript II, 2U/µl of SUPERase-In RNase inhibitor, and SHAPE-Map Buffer (50mM 1M Tris-HCl pH 7.5, 75mM KCl, 10mM DTT, 0.5mM dNTPs, 6mM Mn^{2+}) were added, and the RT reactions were incubated at 42°C for 3 hours. The resulting cDNA was purified using a 1.8:1 ratio of AMPure XP beads to sample according to the manufacturer's protocol. Purified cDNA was eluted in 12µl of nuclease-free water.

PCR and amplicon generation

Next, amplicons were generated by using 0.5mM of gene-specific forward and reverse PCR primers (Appendix, Table 1), 1x Q5 reaction buffer, 0.2mM dNTPs, and 0.02U/ul Q5 HF Hot start DNA Polymerase and 10µl of purified cDNA in a final volume of 50uµ. Touch down PCR cycles were used to enhance specificity with a final annealing temperature of 63˚C for amplicon 1 or 60˚C for amplicon 2. PCR samples were purified using 1:1 AMPure XP beads to sample ratio and eluted in 12µl of nuclease-free water.

NexteraXT library prep

Amplicon concentrations were determined using a Qubit dsDNA kit according to manufacturer protocol and diluted to 0.2ng/ul. To generate sequencing libraries the NexteraXT DNA library prep kit was used with 1/5 of the manufacturer's recommended volumes. Libraries were once more cleaned using a 1:1 AMPure beads to sample ratio.

Sequencing

Library concentrations were determined using a Qubit dsDNA HS Assay Kit and a BioAnalyzer High Sensitivity DNA Analysis. Libraries were diluted, pooled, and sequenced using a NextSeq 500/550 or NextSeq 2000 platform.

SHAPE-MaP data analysis

The sequencing data was analyzed using ShapeMapper2,[168] aligning reads to Pnky transcript. SHAPE reactivities were computed by the ShapeMapper2 script by calculating mutation rate differences between SHAPE-modified and unmodified samples for each nucleotide of Pnky and then normalizing the values into a final ~0-2 reactivity scale. For

the ΔSHAPE analysis, the reactivity obtained from *in-vitro* probing was subtracted from the reactivity obtained from *in-cell* probing.

Secondary structure prediction

ShapeKnots was first used in 600-nt windows to scan the Pnky transcript for the presence of pseudoknots using *in-vitro* derived SHAPE reactivities. A pseudoknot was judged plausible if both strands of pseudoknotted helix had low (<0.4) SHAPE reactivities. A plausible pseudoknot was then used as constraints for the final secondary structure computation. A consensus secondary structure of the full-length Pnky was obtained using SuperFold 1.0[21] with default settings along with the *in-vitro* SHAPE reactivities and the identified pseudoknot constraint. This consensus secondary structure map was used for comparison between *in-cell* (see below) and *in-vitro* SHAPE results by mapping the SHAPE reactivity values from each experiment to every nucleotide position of the region of interest.

Tb^{3+} probing

Terbium probing was performed by incubating 18µl of the folded RNA with 2µl of 5mM $TbCl_3$ stocks prepared in monovalent buffers (final concentration 0.5mM) or 2µl of monovalent buffer (negative control) for 10 min at 25 °C. All reactions were quenched with 3µl of 50mM EDTA pH 8 and cleaned using a Zymo RNA clean and concentrator.

Tb^{3+} library preparation

For each probing condition, the RNA was mixed with 0.5pmol of gene-specific primers (Appendix, Table 1) and brought to a volume of 7.5µl. To anneal primers, the mixture was heated at 90˚C for 1 min, followed by 30˚C for 2 min. To initiate reverse transcription, 0.5ul of Marathon RT, [96] 10µl of 2x Marathon buffer (100 mM Tris-HCl pH

8.3, 400 mM $(NH_4)_2SO_4$, 4 mM MgCl2, 10 mM DTT and 40% glycerol), 1µl of 10mM dNTP mix (NEB), and 1ul of 10mM DTT were added and incubated at 42˚C for 15 min. RNA was degraded with the addition of 1µl of 3M KOH, heated to 95˚C for 5 min, and snap cooled to 4˚C for 5 min. Thereafter, 1µl of 3M HCl was added to neutralize the reaction.

cDNA products from RT were purified using AMPure XP beads by adding 1.2x bead to sample ratio and eluted in 12µl of nuclease-free water. Thereafter, 3' adaptor ligation was performed by mixing 8µl of purified cDNA with 0.2µl of 50µM 3' adaptor (Appendix, Table 1), 1µl of 10mM ATP, 2µl of T4 RNA Ligase buffer, 8µl of 50% PEG 8000. The mixture was incubated at 25˚C for 16 hours, followed by enzyme deactivation at 65˚C for 15 min. Ligated products were purified with AMPure XP beads using a 1.2x bead to sample ratio. The products were PCR amplified 9 cycles with Q5 HF DNA polymerase using Illumina TruSeq forward primer and indexed reverse primers (Appendix, Table 1), with cycle times of 98˚C for 10 sec, 62˚C for 45 sec, and 72˚C for 60 sec. PCR products were purified with 1.2x volume of AMPure XP beads.

Tb^{3+} data analysis

All FASTQ files were processed using Cutadapt to remove Illumina adaptor sequences and aligned to the respective RNA sequence using HISAT2 (v2.10). Stop information was extracted using the RTEventsCounter.py script.[74] The probability of stop per nucleotide was calculated as the number of stops divided by the sum of the total number of read-through events plus the number of stops. Probabilities were background subtracted against a no probe control.

Negative stain EM

For negative stain, EMS continuous carbon grids (300 mesh, CU) were glow discharged for 30 sec at 26mAmps, 0.36 mBar using a PELCO easiGlow Discharge Cleaning System. Immediately after, 4µl of 100nM RNA was added to the glow discharged surface of the grid and incubated for 40 seconds. The RNA was wicked away using Whatman paper, and the grid was immersed into a 30µl drop of 2% uranyl acetate (UA) stain. After 10 seconds, the stain was wicked away, and the grid was immersed in a second 30µl drop of 2% UA. After 30 seconds, the excess stain was once more wicked away and the grid was air dried for 2 min. Grids were imaged using a 120 keV Talos L120C (Thermo Fisher Scientific) at a magnification of ~80,000x and a defocus of -2.5µM.

Optimizing cryo-freezing conditions

Cryo-EM grids were prepared using a Vitrobot Mark IV (Thermo Fisher Scientific). Grids were glow discharged for 90 seconds at 26 mAmps and 0.36 mBar using a PELCO easiGlow Discharge Cleaning System. Aliquots of 3.5µl of RNA (concentrations varying from 100nM-60nM) were applied to the surface of the grid inside the Vitrobot chamber equilibrated at 25˚C and 100% humidity. Blotting was done for 1.5-3 sec with a force of -4 and using filter paper (Ted Pella) equilibrated in the chamber 20 minutes prior to starting the session. Grids were plunge frozen in liquid ethane chilled to -170˚C and stored in liquid nitrogen until further use. Preliminary screening was done using a 200keV Glacios (Thermo Fisher Scientific).

The following grids were tested: UltrAuFoil (R 1.2/1.3 µm, 300 mesh, Au support), Quantifoil with 2nm ultrathin carbon (hole size R 2/2 µm, 300 mesh, Cu support), and Quantifoil with 4nM of carbon made in house (R1.2/1.3 µm, 300 mesh, Cu support). For reverse side blotting, Quantifoil grids with a hole size of 1.2/1.3 µm, Au support, and 300

mesh were glow discharged carbon side facing up. However, the RNA was added to the Au support size, blotted, and plunge frozen, as described above.

Cryo-EM data acquisition and processing

All micrograph data were collected using a 300keV Titan Krios microscope (Thermo Fisher Scientific) equipped with a BioQuantum filter (Gatan) and a spherical aberration of 2.7mm. The micrographs were collected with a Post-GIF K3 camera (Gatan) in super-resolution mode at a magnification of 130,000x and a pixel size of 0.4125 with a total dose of 50 e^- per $Å^2$ per sec fractionated over 35 frames. Data were collected using EPU with a defocus range of -0.8 μM-3 μM. A total of 3,976 micrographs were collected. The data were processed as described above and ~67,000 particles were subject to auto-refinement and post processing resulting in a ~15 Å reconstruction.

Structure and graphical display

All secondary structures were visualized and drawn using StructureEditor.[98] All three-dimensional structure renderings were done using USCF Chimera.[131] Graphical displays were made using GraphPad Prism 8.

5. Conclusions and Future Outlook

In recent years, there has been intense interest in therapeutically targeting RNAs to regulate their function. The key to further developing and advancing this field is RNA structure. When a crystal structure of a protein is solved, we learn new information about how its amino acids form alpha helices, beta sheets, and loops, which are arranged together into a 3-D architecture crucial for molecular function. RNA acts similarly, as its nucleotides are organized into complex helical turns and twists, forming intra-and inter-domain interactions. Even single-stranded regions in stem-loops and bulges, which are often thought to be "unstructured," form sharp backbone turns and participate in complex higher-order structures such as multi-helix base stacks, base triples, and protein interactions that orient adjacent helices. Just like protein structures, RNA structures are dynamic and can adopt different conformations upon complexing with other molecules or as the RNP complex proceeds through stages of its biological processes. While significant progress has been made in understanding RNA structure, there are many classes of RNAs, such as lncRNAs or mRNAs, where this level of detail is limited. This underscores the need to update our toolbox of techniques to understand the tertiary structure in large RNAs.

When analyzing multi-kilobase RNAs, we need ways to pinpoint and focus on specific regions for functional assessment. Tb-seq is a powerful addition to the chemical probing toolbox. When working with a new RNA system, it is essential to use a combination of orthogonal probes, both *in-vitro* and *in-cellulo,* to reveal regulatory RNP interfaces. To understand the proteome interactions with these regions, one can perform targeted CLIP, pull down/mass spectroscopy with a bait RNA, or even recombinant

purification and *in-vitro* biochemistry. To understand if the structure is functional, one can establish a phenotypic assay and disrupt candidate regions with LNAs or ASOs or transfect mutant RNAs into cells to observe their effects.

Complementary, cryo-EM and cryo-ET will allow us to visualize molecular level details of these impressive molecules. With every structure we will learn more about RNA folds, even revealing never before seen architectures. The development of computational tools such as software for model building into lower density maps or denoising backgrounds in micrographs will further drive this field forward, ultimately leading us to visualize an RNA and its associated protein partners inside cells using cryo-ET.

The further exploration of RNA structure will result in new insights into biological processes and promote the development of novel tools to study structure in an even greater depth. An upcoming structural revolution combined with small molecule discovery approaches will lead us into a new era of therapeutically targeting RNA structures.

6. Appendix

6.1 RNA Sequences

aI5γ D135 intron

GAGCGGUCUGAAAGUUAUCAUAAAUAAUAUUUACCAUAUAAUAAUGGAUAAAUU
AUAUUUUUAUCAAUAUAAGUCUAAUUACAAGUGUAUUAAAAUGGUAACAUAAAU
AUGCUAAGCUGUAAUGACAAAAGUAUCCAUAUUCUUGACAGUUAUUUUAUAUUA
UAAAAAAAAGAUGAAGGAACUUUGACUGAUCUAAUAUGCUCAACGAAAGUGAAU
CAAAUGUUAUAAAAUUACUUACACCACUAAUUGAAAACCUGUCUGAUAUUCAAUU
AUUAUUUAUUAUUAUAUAAUUAUAUAAUAAUAAAUAAAAUGGUUGAUGUUAUGU
AUUGGAAAUGAGCAUACGAUAAAUCAUAUAACCAUUAGUAAUAUAAUUUGAGAG
CUAAGUUAGAUAUUUACGUAUUUAUGAUAAAACAGAAUAAACCCUCUUCGGAGG
GUAAAAGAUUGUAUAAAAAGCUAAUGCCAUAUUGUAAUGAUAUGGAUAAGAAUU
AUUAUUCUAAAGAUGAAAAUCUGCUAACUUAUACUAUAGGUGAUACUUCGGUAU
UGCGUGAGCCGUAUGCGAUGAAAGUCGCACGUACGGUUCCUGGAUC

O.i. intron

GUGUGCCCGGCAUGGGUGCAGUCUAUAGGGUGAGAGUCCCGAACUGUGAA
GGCAGAAGUAACAGUUAGCCUAACGCAAGGGUGUCCGUGGCGACAUGGAA
UCUGAAGGAAGCGGACGGCAAACCUUCGGUCUGAGGAACACGAACUUCAU
AUGAGGCUAGGUAUCAAUGGAUGAGUUUGCAUAACAAAACAAAGUCCUUU
CUGCCAAAGUUGGUACAGAGUAAAUGAAGCAGAUUGAUGAAGGGAAAGAC
UGCAUUCUUACCCGGGGAGGUCUGGAAACAGAAGUCAGCAGAAGUCAUAG
UACCCUGUUCGCAGGGGAAGGACGGAACAAGUAUGGCGUUCGCGCCUAAG
CUUGAACCGCCGUAUACCGAACGGUACGUACGGUGGUGGAU

HCV 5’ terminus

ACCUGCCCCUAAUAGGGGCGACACUCCGCCAUGAAUCACUCCCCUGUGAGG
AACUACUGUCUUCACGCAGAAAGCGCCUAGCCAUGGCGUUAGUAUGAGUG
UCGUACAGCCUCCAGGCCCCCCCCUCCCGGGAGAGCCAUAGUGGUCUGCGG
AACCGGUGAGUACACCGGAAUUGCCGGGAAGACUGGGUCCUUUCUUGGAU
AAACCCACUCUAUGCCCGGCCAUUUGGGCGUGCCCCCGCAAGACUGCUAGC
CGAGUAGCGUUGGGUUGCGAAAGGCCUUGUGGUACUGCCUGAUAGGGCGC
UUGCGAGUGCCCCGGGAGGUCUCGUAGACCGUGCACCAUGAGCACAAAUCC
UAAACCUCAAAGAAAAACCAAAAGAAACACCAACCGUCGCCCACAAGAC

RNase P

AUAGGGCGGAGGGAAGCUCAUCAGUGGGGCCACGAGCUGAGUGCGUCCUG
UCACUCCACUCCCAUGUCCCUUGGGAAGGUCUGAGACUAGGGCCAGAGGCG
GCCCUAACAGGGCUCUCCCUGAGCUUCGGGGAGGUGAGUUCCCAGAGAAC
GGGGCUCCGCGCGAGGUCAGACUGGGCAGGAGAUGCCGUGGACCCCGCCCU
UCGGGGAGGGGCCCGGCGGAUGCCUCCUUUGCCGGAGCUUGGAACAGACU
CACGGCCAGCGAAGUGAGUUCAAUGGCUGAGGUGAGGUACCCCGCAGGGG
ACCUCAUAACCCAAUUCAGACUACUCUCCUCCGCCCAUU

SARS-CoV2 5’ terminal 1400 nucleotides

AUUAAAGGUUUAUACCUUCCCAGGUAACAAACCAACCAACUUUCGAUCUC
UUGUAGAUCUGUUCUCUAAACGAACUUUAAAAUCUGUGUGGCUGUCACUC
GGCUGCAUGCUUAGUGCACUCACGCAGUAUAAUUAAUAACUAAUUACUGU
CGUUGACAGGACACGAGUAACUCGUCUAUCUUCUGCAGGCUGCUUACGGU
UUCGUCCGUGUUGCAGCCGAUCAUCAGCACAUCUAGGUUUCGUCCGGGUG
UGACCGAAAGGUAAGAUGGAGAGCCUUGUCCCUGGUUUCAACGAGAAAAC
ACACGUCCAACUCAGUUUGCCUGUUUUACAGGUUCGCGACGUGCUCGUAC
GUGGCUUUGGAGACUCCGUGGAGGAGGUCUUAUCAGAGGCACGUCAACAU
CUUAAAGAUGGCACUUGUGGCUUAGUAGAAGUUGAAAAAGGCGUUUUGCC
UCAACUUGAACAGCCCUAUGUGUUCAUCAAACGUUCGGAUGCUCGAACUG
CACCUCAUGGUCAUGUUAUGGUUGAGCUGGUAGCAGAACUCGAAGGCAUU
CAGUACGGUCGUAGUGGUGAGACACUUGGUGUCCUUGUCCCUCAUGUGGG
CGAAAUACCAGUGGCUUACCGCAAGGUUCUUCUUCGUAAGAACGGUAAUA
AAGGAGCUGGUGGCCAUAGUUACGGCGCCGAUCUAAAGUCAUUUGACUUA
GGCGACGAGCUUGGCACUGAUCCUUAUGAAGAUUUUCAAGAAAACUGGAA
CACUAAACAUAGCAGUGGUGUUACCCGUGAACUCAUGCGUGAGCUUAACG
GAGGGGCAUACACUCGCUAUGUCGAUAACAACUUCUGUGGCCCUGAUGGC
UACCCUCUUGAGUGCAUUAAAGACCUUCUAGCACGUGCUGGUAAAGCUUC
AUGCACUUUGUCCGAACAACUGGACUUUAUUGACACUAAGAGGGGUGUAU
ACUGCUGCCGUGAACAUGAGCAUGAAAUUGCUUGGUACACGGAACGUUCU
GAAAAGAGCUAUGAAUUGCAGACACCUUUUGAAAUUAAAUUGGCAAAGAA
AUUUGACACCUUCAAUGGGGAAUGUCCAAAUUUUGUAUUUCCCUUAAAUU
CCAUAAUCAAGACUAUUCAACCAAGGGUUGAAAAGAAAAAGCUUGAUGGC
UUUAUGGGUAGAAUUCGAUCUGUCUAUCCAGUUGCGUCACCAAAUGAAUG
CAACCAAAUGUGCCUUUCAACUCUCAUGAAGUGUGAUCAUUGUGGUGAAA
CUUCAUGGCAGACGGGCGAUUUUGUUAAAGCCACUUGCGAAUUUUGUGGC
ACUGAGAAUUUGACUAAAGAAGGUGCCACUACUUGUGGUUACUUACCCCA
AAAUGCUGUUGUUAAAAUUUAUUGUCCAGCAUGUCACAAUUCAGAAGUAG
GACCUGAGCAUAGUCUUGC

Pnky

GAGAAGCAACUUCCUCUGGUCUUCUGGAGGUGUAACCUACGUGCCCAGUA
GGAUAUACUGCCGGGUUGUGAAAUGUCCACGCCUCUCCCCAACUGUCUUCC
UUCCCAGCCCUCCGGGCUUGGCUUUCUUGCUUCCCCAGGAGUUCAAAUCUC
CAACCUGCGGAAGAAUUCAGCUGCUUGAAAGGACUUAAGGCAGUGUGCGG
AGGACAUCUCCUUUCUCCGCCAGUAAAGAGAGCUGUUCAAAGACCGAGGC
UGCUUACGAUGACGUGGAGAGGAUUUCAAACAACCUUAACAGGCUUUGAA
CUGACAAGAAGCGGUGAUAUCUGCCUCACUCUGGCACAGCUCCUCCAGUGC
ACUUGCUAGGACAAUGGCUGAGAAAGCACUUGGUGCUGGCCUCUCUGCCG
GGGGCGAGGACUGCCCAAACCCCAGAGACCCUAAGGCAGGAGUUGCUGCAC
UACAAUGGAGAGGUCUUGUCUUAGGCUGCACCUCAGGUUCUCUCAGAUCU
UCUGGAAGGGCGUUAUUCAGCGGUACUGUGGUUGCGCUGUGCCAGCAGCU
GCUUGAUGGAGACCUCUUCUCCCCACAUCUGAAUGGAACGUCUUUGCCCAG
AGUCUACAGAAUGUCAAAACUGAGGCUCCGACCUCAGAGCUACAGCUUUA
AGGACCACUCCACACAAAGAGGCAGCUGGUUGCUUUAAUGAAAACUGCCU
UUAAGCUUCAAGAACUGAGGCCUUGGGGAAUCCAUUUAUAAGGAGCCUAG
AAAAUGCAUUUCCAAGUUGUAUGUUCUUAGGAGAUACAAAAUAAAAUUUA
CCUGAAAAAAUUAAUUGU

6.2 Table of primers and DNA oligos

Table 1

Sequence	Description
5’-CAGACGTGTGCTCTTCCGATCTTATCACCTATAGTATAAGT-3’	Gene specific RT primer for D135 intron. Contains TruSeq overhang
5’-CAGACGTGTGCTCTTCCGATCTATACGGCGGTTCAAGCTTAGG-3’	Gene specific RT primer for Oi intron. Contains TruSeq overhang
5’CAGACGTGTGCTCTTCCGATCTGGAGGAGAGTAGTCTGAATTGGG 3’	Gene specific RT primer for human RNase P. Contains TruSeq overhang
5’-CAGACGTGTGCTCTTCCGATCTTTTTCTTTGAGGTTTAGG-3’	Gene specific RT primer for HCV RNA. Contains TruSeq overhang
5’-CAGACGTGTGCTCTTCCGATCTCCTGTAAAACAGGCAAACTGAGTTG-3’	1 of 5 gene specific RT primers containing a TruSeq overhang for SARS-COV2.
5’-CAGACGTGTGCTCTTCCGATCTCCGTACTGAATGCCTTCGAG-3’	2 of 5 gene specific RT primers containing a TruSeq overhang for SARS-COV2.
5’-CAGACGTGTGCTCTTCCGATCTTAATGCACTCAAGAGGGTAGC-3’	3 of 5 gene specific RT primers containing a TruSeq overhang for SARS-COV2.
5’-CAGACGTGTGCTCTTCCGATCTGGTGTCAAATTTCTTTGCC-3’	4 of 5 gene specific RT primers containing a TruSeq overhang for SARS-COV2.
5’-CAGACGTGTGCTCTTCCGATCTCAAGACTATGCTCAGGTCC-3’	5 of 5 gene specific RT primers containing a TruSeq overhang for SARS-COV2.
5’Phos-NNNNNNAGATCGGAAGAGCGTCGTGTAG-3’Bio	DNA adapter used to adapter ligate cDNA in TruSeq library prep
5’- GCGCAACCACAGTACCGCTGAATA-3’	1 of 2 gene specific Pnky RT primers for MaP RT
5’- CAGGTAAATTTTATTTTGTATCTCCTAAGAAC-3’	2 of 2 gene specific Pnky RT primer for MaP RT

5’-CAGACGTGTGCTCTTCCGATCTACAATTAATTTTTTCAGGTAAATT-3’	Gene specific Pnky RT primer with TruSeq overhang
5’- GGGAGAAGCAACTTCCTCTG-3’	Amplicon 1 Forward primer
5’- GCGCAACCACAGTACCGCTGAATA-3’	Amplicon 1 Reverse primer
5’- CCAAACCCCAGAGACCCTA-3’	Amplicon 2 Forward primer
5’- CAGGTAAATTTTATTTTGTATCTCCTAAGAAC-3’	Amplicon 2 Reverse primer

7. References

1 Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561-563, doi:10.1038/227561a0 (1970).

2 Fedor, M. J. & Williamson, J. R. The catalytic diversity of RNAs. *Nature Reviews Molecular Cell Biology* **6**, 399-412, doi:10.1038/nrm1647 (2005).

3 Shimoni, Y. *et al.* Regulation of gene expression by small non-coding RNAs: a quantitative view. *Molecular Systems Biology* **3**, 138, doi:10.1038/msb4100181 (2007).

4 Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology* **22**, 96-118, doi:10.1038/s41580-020-00315-9 (2021).

5 Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nature Review Genetics* **15**, 469-479, doi:10.1038/nrg3681 (2014).

6 Wilkinson, M. E., Charenton, C. & Nagai, K. RNA splicing by the spliceosome. *Annual Review in Biochemistry* **89**, 359-388 (2020).

7 Zhao, C. & Pyle, A. M. Structural Insights into the Mechanism of Group II Intron Splicing. *Trends in Biochemical Sciences* **42**, 470-482, doi:10.1016/j.tibs.2017.03.007 (2017).

8 Wu, J. *et al.* Cryo-EM structure of the human ribonuclease P holoenzyme. *Cell* **175**, 1393-1404.e1311 (2018).

9 O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology* **9**, doi:10.3389/fendo.2018.00402 (2018).

10 Vicens, Q. & Kieft, J. S. Thoughts on how to think (and talk) about RNA structure. *Proceedings of the National Academy of Sciences* **119**, e2112677119, doi:doi:10.1073/pnas.2112677119 (2022).

11 Mauger, D. M. *et al.* mRNA structure regulates protein expression through changes in functional half-life. *Proceedings of the National Academy of Sciences* **116**, 24075-24083, doi:doi:10.1073/pnas.1908052116 (2019).

12 Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* **152**, 17-24 (2013).

13 Kortmann, J. & Narberhaus, F. Bacterial RNA thermometers: molecular zippers and switches. *Nature Reviews Microbiology* **10**, 255-265, doi:10.1038/nrmicro2730 (2012).

14 Lambert, D. & Draper, D. E. Effects of osmolytes on RNA secondary and tertiary structure stabilities and RNA-Mg2+ interactions. *Journal of Molecular Biology* **370**, 993-1005, doi:10.1016/j.jmb.2007.03.080 (2007).

15 Corley, M., Burns, M. C. & Yeo, G. W. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Molecular Cell* **78**, 9-29, doi:https://doi.org/10.1016/j.molcel.2020.03.011 (2020).

16 Fedorova, O. *et al.* Small molecules that target group II introns are potent antifungal agents. *Nature Chemical Biology* **14**, 1073-1078, doi:10.1038/s41589-018-0142-0 (2018).

17 Zhang, K. *et al.* Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nature Communications* **10**, 5511, doi:10.1038/s41467-019-13494-7 (2019).

18 Marcia, M. & Pyle, Anna M. Visualizing Group II Intron Catalysis through the Stages of Splicing. *Cell* **151**, 497-507 (2012).

19 Underwood, J. G. *et al.* FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **7**, 995-1001 (2010).

20 Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103-107, doi:10.1038/nature09322 (2010).

21 Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A. & Weeks, K. M. Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nature Protocols* **10**, 1643-1669, doi:10.1038/nprot.2015.103 (2015).

22 Wang, P. Y., Sexton, A. N., Culligan, W. J. & Simon, M. D. Carbodiimide reagents for the chemical probing of RNA structure in cells. *RNA* **25**, 135-146 (2019).

23 Tijerina, P., Mohr, S. & Russell, R. DMS footprinting of structured RNAs and RNA-protein complexes. *Nature Protocols* **2**, 2608-2623, doi:10.1038/nprot.2007.380 (2007).

24 Smola, M. J., Calabrese, J. M. & Weeks, K. M. Detection of RNA–Protein interactions in living cells with SHAPE. *Biochemistry* **54**, 6867-6875 (2015).

25 Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nature Methods* **14**, 75-82, doi:10.1038/nmeth.4057 (2017).

26 Smola, M. J. *et al.* SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proceedings of the National Academy of Sciences.***113**, 10322-10327, doi:10.1073/pnas.1600008113 (2016).

27 Draper, D. E. A guide to ions and RNA structure. *RNA* **10**, 335-343, doi:10.1261/rna.5205404 (2004).

28 Harris, D. A. & Walter, N. G. in *Handbook of RNA Biochemistry* 205-213 (Weinheim, 2005).

29 CIiesiloka, J., Marciniec, T. & Krzyzosiak, W. J. Probing the environment of lanthanide binding sites in yeast tRNAPhe by specific metal-ion-promoted cleavages. *European Journal of Biochemistry.* **182**, 445-450 (1989).

30 Regulski, E. E. & Breaker, R. R. In-line probing analysis of riboswtiches. *Methods in Molecular Biology* 53-67 (Humana Press, 2008).

31 Forconi, M. & Herschlag, D. Metal ion-based RNA cleavage as a structural probe Ch. 5. *Methods Enzymology* **468**, 91-106 (2009).

32 Zinshteyn, B. *et al.* Assaying RNA structure with LASER-seq. *Nucleic Acids Research* **47**, 43-55 (2019).

33 Tullius, T. D. & Greenbaum, J. A. Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.* **9**, 127-134 (2005).

34 Adilakshmi, T., Soper, S. F. C. & Woodson, S. A. in *Methods in Enzymology* Vol. 468 239-258 (Academic Press, 2009).

35 Feng, C. *et al.* Light-activated chemical probing of nucleobase solvent accessibility inside cells. *Nature Chemical Biology* **14**, 276 (2018).

36 Hulscher, R. M. *et al.* Probing the structure of ribosome assembly intermediates in vivo using DMS and hydroxyl radical footprinting. *Methods* **103**, 49-56, doi:10.1016/j.ymeth.2016.03.012 (2016).

37 Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**, 1267-1279 (2016).

38 Aw, J. G. *et al.* In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Molecular Cell* **62**, 603-617 (2016).

39 Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**, 26, doi:10.1186/1748-7188-6-26 (2011).

40 Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406-3415, doi:10.1093/nar/gkg595 (2003).

41 Townshend, R. J. L. *et al.* Geometric deep learning of RNA structure. *Science* **373**, 1047-1051, doi:doi:10.1126/science.abe5650 (2021).

42 Yu, H., Qi, Y. & Ding, Y. Deep Learning in RNA Structure Studies. *Frontiers in Molecular Biosciences* **9**, doi:10.3389/fmolb.2022.869601 (2022).

43 Tomezsko, P. J. *et al.* Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* **582**, 438-442 (2020).

44 Homan, P. J. *et al.* Single-molecule correlated chemical probing of RNA. *Proceedings of the National Academy of Sciences* **111**, 13858-13863 (2014).

45 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

46 Fürtig, B., Richter, C., Wöhnert, J. & Schwalbe, H. NMR spectroscopy of RNA. *Chembiochem* **4**, 936-962, doi:10.1002/cbic.200300700 (2003).

47 Ferré-D'Amaré, A. R. & Doudna, J. A. RNA Folds: Insights from Recent Crystal Structures. *Annual Review of Biophysics and Biomolecular Structure* **28**, 57-73, doi:10.1146/annurev.biophys.28.1.57 (1999).

48 Mustoe, A. M., Brooks, C. L. & Al-Hashimi, H. M. Hierarchy of RNA functional dynamics. *Annual Review of Biochemistry* **83**, 441-466, doi:10.1146/annurev-biochem-060713-035524 (2014).

49 Kim, D. N. *et al.* Zinc-finger protein CNBP alters the 3-D structure of lncRNA Braveheart in solution. *Nature Communications* **11**, 148, doi:10.1038/s41467-019-13942-4 (2020).

50 Patel, T. R., Chojnowski, G., Koul, A., McKenna, S. A. & Bujnicki, J. M. Structural studies of RNA-protein complexes: A hybrid approach involving hydrodynamics, scattering, and computational methods. *Methods* **118**, 146-162 (2017).

51 Kühlbrandt, W. The Resolution Revolution. *Science* **343**, 1443-1444, doi:doi:10.1126/science.1251652 (2014).

52 Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. W. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861, doi:10.7554/eLife.36861 (2018).

53 Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R. & Johnson, R. Towards a complete map of the human long non-coding RNA transcriptome. *Nature Reviews Genetics* **19**, 535-548 (2018).

54 Andersen, R. E. *et al.* The Long Noncoding RNA Pnky Is a Trans-acting Regulator of Cortical Development In Vivo. *Developmental cell* **49**, 632-642.e637 (2019).

55 Brunel, C., Marquet, R., Romby, P. & Ehresmann, C. RNA loop–loop interactions as dynamic functional motifs. *Biochimie* **84**, 925-944 (2002).

56 Staple, D. W. & Butcher, S. E. Pseudoknots: RNA structures with diverse functions. *PLoS Biology.* **3**, e213 (2005).

57 Zhao, C. & Pyle, A. M. Structural insights into the mechanism of group II Intron splicing. *Trends in Biochemical Sciences* **42**, 470-482 (2017).

58 Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339-346 (2012).

59 Cech, Thomas R. & Steitz, Joan A. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell* **157**, 77-94 (2014).

60 Šponer, J. *et al.* RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chemical Reviews* **118**, 4177-4338 (2018).

61 Fürtig, B., Richter, C., Wöhnert, J. & Schwalbe, H. NMR spectroscopy of RNA. *ChemBioChem* **4**, 936-962 (2003).

62 Reyes, F. E., Garst, A. D. & Batey, R. T. Strategies in RNA crystallography. *Methods Enzymology* **469**, 119-139 (2009).

63 Fernandez-Leiro, R. & Scheres, S. H. Unravelling biological macromolecules with cryo-electron microscopy. *Nature* **537**, 339-346 (2016).

64 Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2‘-hydroxyl acylation and

primer extension (SHAPE). *Journal of the American Chemical Society* **127**, 4223-4231 (2005).

65 Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103-107 (2010).

66 Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696-700 (2014).

67 Pyle, A. M. Metal ions in the structure and function of RNA. *Journal of Biological Inorganic Chemistry* **7**, 679-690 (2002).

68 Soukup, G. A. & Breaker, R. R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5**, 1308-1325 (1999).

69 Forconi, M. & Herschlag, D. in *Methods in Enzymology* Vol. 468 91-106 (Academic Press, 2009).

70 Harris, D. A., Tinsley, R. A. & Walter, N. G. Terbium-mediated footprinting probes a catalytic conformational switch in the antigenomic hepatitis delta virus ribozyme. *Journal of Molecular Biology* **341**, 389-403 (2004).

71 Sigel, R. K. O., Vaidya, A. & Pyle, A. M. Metal ion binding sites in a group II intron core. *Nature Structural Biology* **7**, 1111 (2000).

72 Hargittai, M. R. & Musier-Forsyth, K. Use of terbium as a probe of tRNA tertiary structure and folding. *RNA* **6**, 1672-1680 (2000).

73 Choudhary, P. K., Gallo, S. & Sigel, R. K. O. in *RNA Folding: Methods and Protocols* (ed Christina Waldsich) 143-158 (Humana Press, Totowa, NJ, 2014).

74 Sexton, A. N., Wang, P. Y., Rutenberg-Schoenberg, M. & Simon, M. D. Interpreting reverse transcriptase termination and mutation events for greater insight into the chemical probing of RNA. *Biochemistry* **56**, 4713-4721 (2017).

75 Su, L. J., Waldsich, C. & Pyle, A. M. An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic Acids Research* **33**, 6674-6687 (2005).

76 Pyle, A. M. The tertiary structure of group II introns: implications for biological function and evolution. *Critical Reviews in Biochemistry and Molecular Biology* **45**, 215-232 (2010).

77 Quade, N., Boehringer, D., Leibundgut, M., Van Den Heuvel, J. & Ban, N. Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution. *Nature Communication* **6**, 1-9 (2015).

78 Lukavsky, P. J., Kim, I., Otto, G. A. & Puglisi, J. D. Structure of HCV IRES domain II determined by NMR. *Nature Structural Moleulcar Biology* **10**, 1033-1038 (2003).

79 Yokoyama, T. *et al.* HCV IRES captures an actively translating 80S ribosome. *Molecular Cell* **74**, 1205-1214.e1208 (2019).

80 Paulsen, R. B. *et al.* Inhibitor-induced structural change in the HCV IRES domain IIa RNA. *Proceedings of the National Academy of Sciences* **107**, 7263-7268 (2010).

81 Dibrov, S. M. *et al.* Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches. *Proceedings of the National Academy of Sciences* **109**, 5223-5228 (2012).

82 Huston, N. C. *et al.* Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Molecular cell* **81**, 584-598. e585 (2021).

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

84 Ziv, O. *et al.* The short- and long-range RNA-RNA Interactome of SARS-CoV-2. *Molecular Cell* **80**, 1067-1077.e1065 (2020).

85 McHugh, C. A. & Guttman, M. RAP-MS: a method to identify proteins that interact directly with a specific RNA molecule in cells. *Methods in Molecular Biology.* **1649**, 473-488 (2018).

86 Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464-469 (2008).

87 Ma, H., Jia, X., Zhang, K. & Su, Z. Cryo-EM advances in RNA structure determination. *Signal Transduction and Targeted Therapy.* **7**, 58 (2022).

88 Spokoini-Stern, R. *et al.* Visualizing the structure and motion of the long noncoding RNA HOTAIR. *RNA* **26**, 629-636 (2020).

89 Uroda, T. *et al.* Conserved pseudoknots in lncRNA MEG3 are essential for stimulation of the p53 pathway. *Molecular Cell* **75**, 982-995 e989 (2019).

90 Aguilar, R. *et al.* Targeting Xist with compounds that disrupt RNA structure and X inactivation. *Nature* **604**, 160-166 (2022).

91 Swisher, J., Duarte, C. M., Su, L. J. & Pyle, A. M. Visualizing the solvent-inaccessible core of a group II intron ribozyme. *The EMBO Journal.* **20**, 2051 (2001).

92 Qin, P. Z. & Pyle, A. M. Stopped-Flow Fluorescence Spectroscopy of a Group II Intron Ribozyme Reveals that Domain 1 Is an Independent Folding Unit with a Requirement for Specific Mg2+ Ions in the Tertiary Structure. *Biochemistry* **36**, 4718-4730 (1997).

93 Tang, G.-Q. *et al.* Relaxed rotational and scrunching changes in P266L mutant of T7 RNA polymerase reduce short abortive RNAs while delaying transition into elongation. *PloS one* **9**, e91859 (2014).

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

95 Wan, H., Adams, R. L., Lindenbach, B. D. & Pyle, A. M. The in vivo and in vitro architecture of the Hepatitis C Virus RNA genome uncovers functional RNA secondary and tertiary structures. *Journal of Virology.* **96**, e0194621-e0194621 (2022).

96 Guo, L.-T. *et al.* Sequencing and structure probing of long RNAs using MarathonRT: A next-generation reverse transcriptase. *Journal of Molecular Biology.* **432**, 3338-3352 (2020).

97 Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling

(SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nature Protocols.* **10**, 1643-1669 (2015).

98 Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics.* **11**, 129 (2010).

99 Kleinjung, J. & Fraternali, F. POPSCOMP: an automated interaction analysis of biomolecular complexes. *Nucleic Acids Research.* **33**, W342-W346 (2005).

100 Daniels, D. L., Michels Jr, W. J. & Pyle, A. M. Two Competing Pathways for Self-splicing by Group II Introns: A Quantitative Analysis of in-vitro Reaction Rates and Products. *Journal of Molecular Biology* **256**, 31-49, doi:https://doi.org/10.1006/jmbi.1996.0066 (1996).

101 Costa, M., Walbott, H., Monachello, D., Westhof, E. & Michel, F. Crystal structures of a group II intron lariat primed for reverse splicing. *Science* **354**, aaf9258, doi:10.1126/science.aaf9258 (2016).

102 Haack, D. B. *et al.* Cryo-EM Structures of a Group II Intron Reverse Splicing into DNA. *Cell* **178**, 612-623.e612, doi:https://doi.org/10.1016/j.cell.2019.06.035 (2019).

103 Michel, F., Kazuhiko, U. & Haruo, O. Comparative and functional anatomy of group II catalytic introns — a review. *Gene* **82**, 5-30, doi:https://doi.org/10.1016/0378-1119(89)90026-7 (1989).

104 Lehmann, K. & Schmidt, U. Group II introns: structure and catalytic versatility of large natural ribozymes. *Critical Reviews in Biochemistry and Molecular Biology* **38**, 249-303, doi:10.1080/713609236 (2003).

105 Qu, G. *et al.* Structure of a group II intron in complex with its reverse transcriptase. *Nature structural & molecular biology* **23**, 549-557, doi:10.1038/nsmb.3220 (2016).

106 Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77-82, doi:10.1126/science.1153803 (2008).

107 Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514**, 193, doi:10.1038/nature13790 (2014).

108 Peebles, C. L. *et al.* A self-splicing RNA excises an intron lariat. *Cell* **44**, 213-223, doi:10.1016/0092-8674(86)90755-5 (1986).

109 Swisher, J., Duarte, C. M., Su, L. J. & Pyle, A. M. Visualizing the solvent-inaccessible core of a group II intron ribozyme. *The EMBO Journal* **20**, 2051-2061, doi:10.1093/emboj/20.8.2051 (2001).

110 Su, L. J., Brenowitz, M. & Pyle, A. M. An alternative route for the folding of large RNAs: apparent two-state folding by a group II intron ribozyme. *Journal of Molecular Biology* **334**, 639-652 (2003).

111 Fedorova, O. & Pyle, A. M. A conserved element that stabilizes the group II intron active site. *RNA* **14**, 1048-1056, doi:10.1261/rna.942308 (2008).

112 Reyes, F. E., Garst, A. D. & Batey, R. T. Strategies in RNA crystallography. *Methods in enzymology* **469**, 119-139 (2009).

113 Wang, H.-W. & Wang, J.-W. How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Science* **26**, 32-39, doi:https://doi.org/10.1002/pro.3022 (2017).

114 Cole, J. L., Lary, J. W., T, P. M. & Laue, T. M. Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods in Cell Biology* **84**, 143-179, doi:10.1016/s0091-679x(07)84006-4 (2008).

115 Somarowthu, S., Legiewicz, M., Keating, K. S. & Pyle, A. M. Visualizing the ai5γ group IIB intron. *Nucleic acids research* **42**, 1947-1958, doi:10.1093/nar/gkt1051 (2014).

116 Drulyte, I. *et al.* Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallographica: D Structural Biology* **74**, 560-571, doi:10.1107/s2059798318006496 (2018).

117 Kampjut, D., Steiner, J. & Sazanov, L. A. Cryo-EM grid optimization for membrane proteins. *Iscience* **24**, 102139 (2021).

118 Drulyte, I. *et al.* Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallographica Section D* **74**, 560-571, doi:doi:10.1107/S2059798318006496 (2018).

119 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166, doi:10.7554/eLife.42166 (2018).

120 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).

121 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).

122 Boudvillain, M. & Pyle, A. M. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J* **17**, 7091-7104, doi:10.1093/emboj/17.23.7091 (1998).

123 Fedorova, O., Mitros, T. & Pyle, A. M. Domains 2 and 3 Interact to Form Critical Elements of the Group II Intron Active Site. *Journal of Molecular Biology* **330**, 197-209, doi:https://doi.org/10.1016/S0022-2836(03)00594-1 (2003).

124 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods* **16**, 1153-1160, doi:10.1038/s41592-019-0575-8 (2019).

125 Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysics Journal* **78**, 1606-1619, doi:10.1016/s0006-3495(00)76713-0 (2000).

126 Brown, P. H. & Schuck, P. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophysical Journal* **90**, 4651-4661, doi:10.1529/biophysj.106.081372 (2006).

127 Mitra, S. in *Methods in Enzymology* Vol. 469 209-236 (Academic Press, 2009).

128 Brautigam, C. A. Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. *Methods Enzymology* **562**, 109-133, doi:10.1016/bs.mie.2015.05.001 (2015).

129 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Structural Biology* **66**, 486-501, doi:10.1107/s0907444910007493 (2010).

130 Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology* **75**, 861-877, doi:10.1107/s2059798319011471 (2019).

131 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70-82, doi:10.1002/pro.3943 (2021).

132 Rinn, J. L. & Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry* **81**, 145-166, doi:10.1146/annurev-biochem-051410-092902 (2012).

133 DiStefano, J. K. The Emerging Role of Long Noncoding RNAs in Human Disease. *Methods in Moleuclar Biology* **1706**, 91-110, doi:10.1007/978-1-4939-7471-9_6 (2018).

134 Sparber, P., Filatova, A., Khantemirova, M. & Skoblov, M. The role of long non-coding RNAs in the pathogenesis of hereditary diseases. *BMC Medical Genomics* **12**, 42, doi:10.1186/s12920-019-0487-6 (2019).

135 Romero-Barrios, N., Legascue, M. F., Benhamed, M., Ariel, F. & Crespi, M. Splicing regulation by long noncoding RNAs. *Nucleic Acids Research* **46**, 2169-2184, doi:10.1093/nar/gky095 (2018).

136 Yamazaki, T. *et al.* Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Molecular Cell* **70**, 1038-1053 e1037, doi:10.1016/j.molcel.2018.05.019 (2018).

137 Xiang, J. F. *et al.* Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Research* **24**, 513-531, doi:10.1038/cr.2014.35 (2014).

138 Yang, F. *et al.* The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biology* **16**, 52, doi:10.1186/s13059-015-0618-0 (2015).

139 Chen, C.-K. *et al.* Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* **354**, 468-472, doi:doi:10.1126/science.aae0047 (2016).

140 Lee, S. *et al.* Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* **164**, 69-80, doi:10.1016/j.cell.2015.12.017 (2016).

141 Wu, H. *et al.* Unusual Processing Generates SPA LncRNAs that Sequester Multiple RNA Binding Proteins. *Molecular Cell* **64**, 534-548, doi:10.1016/j.molcel.2016.10.007 (2016).

142 Yap, K. *et al.* A Short Tandem Repeat-Enriched RNA Assembles a Nuclear Compartment to Control Alternative Splicing and Promote Cell Survival. *Mol Cell* **72**, 525-540 e513, doi:10.1016/j.molcel.2018.08.041 (2018).

143 Fang, R., Moss, W. N., Rutenberg-Schoenberg, M. & Simon, M. D. Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. *PLoS Genetics* **11**, e1005668-e1005668, doi:10.1371/journal.pgen.1005668 (2015).

144 Liu, F., Somarowthu, S. & Pyle, A. M. Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nature Chemical Biology* **13**, 282 (2017).

145 Somarowthu, S. *et al.* HOTAIR Forms an Intricate and Modular Secondary Structure. *Molecular Cell* **58**, 353-361, doi:10.1016/j.molcel.2015.03.006 (2015).

146 Xue, Z. *et al.* A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. *Molecular Cell* **64**, 37-50, doi:https://doi.org/10.1016/j.molcel.2016.08.010 (2016).

147 Hawkes, E. J. *et al.* COOLAIR antisense RNAs form evolutionarily conserved elaborate secondary structures. *Cell reports* **16**, 3087-3096 (2016).

148 Ilik, I. A. *et al.* Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in Drosophila. *Molecular cell* **51**, 156-173 (2013).

149 Novikova, I. V., Hennelly, S. P. & Sanbonmatsu, K. Y. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic acids research* **40**, 5034-5051 (2012).

150 Zhang, B. *et al.* Identification and characterization of a class of MALAT1-like genomic loci. *Cell reports* **19**, 1723-1738 (2017).

151 Aguilar, R. *et al.* Targeting Xist with compounds that disrupt RNA structure and X inactivation. *Nature* **604**, 160-166 (2022).

152 Briggs, J. A., Wolvetang, E. J., Mattick, J. S., Rinn, J. L. & Barry, G. Mechanisms of long non-coding RNAs in mammalian nervous system development, plasticity, disease, and evolution. *Neuron* **88**, 861-877 (2015).

153 Ng, S.-Y., Lin, L., Soh, B. S. & Stanton, L. W. Long noncoding RNAs in development and disease of the central nervous system. *Trends in Genetics* **29**, 461-468, doi:https://doi.org/10.1016/j.tig.2013.03.002 (2013).

154 Ramos, Alexander D. *et al.* Integration of Genome-wide Approaches Identifies lncRNAs of Adult Neural Stem Cells and Their Progeny In Vivo. *Cell Stem Cell* **12**, 616-628, doi:https://doi.org/10.1016/j.stem.2013.03.003 (2013).

155 Bond, A. M. *et al.* Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nature neuroscience* **12**, 1020-1027 (2009).

156 Sauvageau, M. *et al.* Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *elife* **2**, e01749 (2013).

157 Andersen, R. E. *et al.* The Long Noncoding RNA Pnky Is a Trans-acting Regulator of Cortical Development In Vivo. *Developmental Cell* **49**, 632-642 e637, doi:10.1016/j.devcel.2019.04.032 (2019).

158 Ramos, A. D. *et al.* The long noncoding RNA Pnky regulates neuronal differentiation of embryonic and postnatal neural stem cells. *Cell Stem Cell* **16**, 439-447, doi:10.1016/j.stem.2015.02.007 (2015).

159 Chillón, I. *et al.* Native Purification and Analysis of Long RNAs. *Methods in enzymology* **558**, 3-37, doi:10.1016/bs.mie.2015.01.008 (2015).

160 Sosnick, T. R. & Pan, T. RNA folding: models and perspectives. *Current opinion in structural biology* **13**, 309-316 (2003).

161 Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences* **101**, 7287-7292 (2004).

162 Haas, E. S., Morse, D. P., Brown, J. W., Schmidt, F. J. & Pace, N. R. Long-range structure in ribonuclease P RNA. *Science* **254**, 853-856, doi:10.1126/science.1719634 (1991).

163 Zaug, A. J. & Cech, T. R. Analysis of the structure of Tetrahymena nuclear RNAs in vivo: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. *RNA* **1**, 363-374 (1995).

164 Brierley, I., Pennell, S. & Gilbert, R. J. C. Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nature Reviews Microbiology* **5**, 598-610, doi:10.1038/nrmicro1704 (2007).

165 Staple, D. W. & Butcher, S. E. Pseudoknots: RNA Structures with Diverse Functions. *PLOS Biology* **3**, e213, doi:10.1371/journal.pbio.0030213 (2005).

166 Petrova, O. E., Garcia-Alcalde, F., Zampaloni, C. & Sauer, K. Comparative evaluation of rRNA depletion procedures for the improved analysis of bacterial biofilm and mixed pathogen culture transcriptomes. *Scientific Reports* **7**, 41114, doi:10.1038/srep41114 (2017).

167 Uroda, T. *et al.* Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway. *Molecular Cell* **75**, 982-995.e989, doi:https://doi.org/10.1016/j.molcel.2019.07.025 (2019).

168 Busan, S. & Weeks, K. M. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA* **24**, 143-148, doi:10.1261/rna.061945.117 (2018).