BIOCHEMISTRY

Cut and paste for RNA structure

Proximity ligation allows for the global resolution of RNA secondary structures.

RNAs have diverse roles in the cell, and being able to determine their complex secondary and tertiary structures is a key to understanding these roles. Although computational approaches for predicting RNA secondary structure can be useful, experimental approaches are still needed for accurate RNA structure determination.

Several transcriptome-wide methods have been developed to probe RNA structure. Dimethyl sulfate (DMS)-seq and selective 2′-hydroxyl acylation and primer extension (SHAPE)-seq use chemical reactivity to distinguish paired and single-stranded regions, and other methods, such as parallel analysis of RNA structure (PARS)-seq and Frag-seq, use cleavage by structure-specific nucleases to probe secondary structure. Although powerful, these methods reveal only bases participating in secondary structure, and not the bases with which they interact. In addition, these methods are typically performed on RNA extracted from cells.

Jay Shendure at the University of Washington and Vijay Ramani, a graduate student in Shendure's lab, sought to develop a strategy complementary to these approaches that would reveal which bases are interacting with one another in cells. "One of the things that sparked our thinking was recent studies highlighting the differences between *in vivo* and *in vitro* RNA secondary structure," recalls Shendure. To develop a strategy to better map RNA secondary structure as it exists in cells, they turned to a method that had been used previously to study chromatin architecture as well as microRNA interactions: proximity ligation.

In proximity ligation, molecules of interest are digested by enzymes and then resealed by a ligase. Ligation occurs only between molecules that are close in physical space. Thus, analysis of which regions of the target molecules are ligated can provide a map of spatial arrangement. This method, RNA proximity ligation (RPL, pronounced "ripple"), involves digestion of RNA in cells, application of an enzyme that makes the ends competent for ligation, and enzymatic ligation of the RNA. Deep sequencing is then used to determine which bases interact to form the secondary structure.

CHEMISTRY

A FIX FOR FORMALDEHYDE FIXATION

Researchers use chemistry to solve a nagging problem that has challenged the analysis of RNA from formaldehyde-fixed tissue specimens.

In hospitals far and wide, tissue samples taken during routine biopsies and surgeries are prepared for histology analysis by being fixed in formaldehyde and embedded in paraffin. Hundreds of millions of these preserved samples are banked and indexed by disease, treatment and outcome. Collectively, they serve as a resource with enormous potential for yielding genetic links to disease and for identifying potential drug targets, says Eric Kool, a chemist at Stanford University.

Extracting molecular information about RNA, DNA and proteins from formaldehyde-fixed specimens, however, has been fraught with challenges. Formaldehyde causes extensive molecular cross-linking and the formation of aminal and hemiaminal adducts, hindering the analysis of fixed samples by PCR, sequencing, immunohistochemistry and mass spectrometry. A number of protocols exist for de-paraffinization and reversal of formaldehyde cross-links and adducts, but they require the application of biomolecule-degrading heat and remain far from efficient.

Kool and his team recently tackled the challenge of reversing formaldehyde cross-links in RNA using chemistry—and without applying high heat. "A few years ago a researcher at a local genomics company came to me to see if I had any ideas for solving their 'formaldehyde problem,'" Kool recounts. At that time, he didn't have a solution for the researcher, but the problem stayed lodged in the back of his mind until recently, when his lab was working on an unrelated project that involved developing catalysts to speed the formation of imine bonds. "It occurred to me that these catalysts could also be useful in breaking imine and imine-like bonds, and suddenly the formaldehyde problem came back to me," he says.
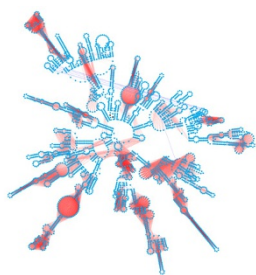
Secondary structure of yeast 25S RNA (black lines and blue dots) overlaid by RPL data (red lines).

The team used RPL in yeast and mammalian cells. In both cases, they obtained strong signals for abundant RNAs such as ribosomal RNAs, which were used to generate accurate secondary-structure maps. In analyzing their data, they found that the majority of reads mapped across loops in the RNA that were cleaved by the digestion enzymes, yielding information about the secondary structure. They also saw evidence of long-range, tertiary interactions, indicating that the method may yield tertiary-structure information.

On the road to implementing RPL, the team encountered a few obstacles. "Mapping algorithms aren't set up well to look at these chimeras, so we had to hijack software that was intended for other purposes," notes Shendure. They also had to deal with technical issues, such as low ligation efficiency, and the large numbers of reads, or high sequencing depth, necessary to perform the analysis. For these they used a 'brute force' approach by obtaining enough sequencing reads to map abundant RNA species.

The large number of reads needed for accurate secondary-structure mapping is perhaps the limiting factor for using RPL to study cellular RNAs. According to Ramani, "Quite a small fraction of library molecules we sequence actually harbor the events we're interested in… We're thinking about employing subtraction or hybrid enrichment to get at these RNAs." The team notes that future improvements to RPL could include methods to increase sensitivity, as well as better computational tools for merging these data with data from complementary approaches such as SHAPE-seq. The team hopes to use RPL to study biologically important RNAs such as long noncoding RNAs and mRNAs in the future.

Rita Strack

RESEARCH PAPERS
Ramani, V. *et al.* High-throughput determination of RNA structure by proximity ligation. *Nat. Biotechnol.* **33**, 980–984 (2015).

Using monomeric nucleotides as a model system, Kool's group tested the ability of a series of arylacid and amine catalysts to reverse formaldehyde adducts. The winning catalyst was a bifunctional phosphanilate compound that could reverse both aminal and hemiaminal adducts. They next tested an extensively cross-linked 16-mer RNA strand consisting of a central self-complementary sequence flanked by overhanging adenosines on both sides. Heating the sample to 60 °C resulted in the removal of formaldehyde cross-links and adducts, but it also caused RNA degradation. The phosphanilate catalyst, however, efficiently removed cross-links and adducts at the much milder temperature of 37 °C.

As a final test, the researchers prepared a model tissue specimen consisting of a pellet from a cultured cell line fixed with formaldehyde and embedded in paraffin. They used a standard protocol for de-paraffinization and then added their phosphanilate catalyst, optimizing the incubation time and temperature. Kool's team found that in comparison to a commercial kit and a literature protocol that required heating at higher temperatures (80 °C and 70 °C, respectively), the phosphanilate catalysts, applied at an optimal temperature of 55 °C, improved RNA yields from the model specimen 13-fold on average.

Kool expects that a modified version of his team's mild approach to reversing cross-links and adducts in RNA could also be applied to DNA, and to proteins, which could potentially be used for proteomic analysis of formaldehyde-fixed specimens. His group has also begun to apply the method to real clinical specimens, with promising results. "We hope that catalytic methods [such as this] will be applied broadly and will help large numbers of cancer patients as their doctors make medical decisions about treatment," he says.

Allison Doerr

RESEARCH PAPERS
Karmakar, S. *et al.* Organocatalytic removal of formaldehyde adducts from RNA and DNA bases. *Nat. Chem.* **7**, 752–758 (2015).