RNA World research—still evolving

THOMAS R. CECH

Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Colorado BioFrontiers Institute, Boulder, Colorado 80309-0596, USA

One step forward, two steps back. That's how laboratory research seems to progress for most grad students and postdocs from week to week. Yet, when we step back and look at how far the RNA field has progressed in the 20 years since the *RNA* journal was born, we get a very different perspective. The international community of RNA scientists, through hard work, serendipity and building on each others' findings, has made remarkable discoveries. In this perspective, I will remind the reader of the state of RNA science then, in 1995, where the field is now, and what the next 20 years may bring.

Then

In 1995, the hot new science being presented at RNA conferences was concentrated in two areas—RNA splicing and ribozymes—and there was also a lot of action in RNA editing, snoRNAs, hnRNPs, tRNAs, translation, and HIV RNA. Although the intron-exon structure of eukaryotic genes had been discovered almost 20 years earlier, it took some time for the chemical mechanism to be worked out, spliceosome assembly and snRNP functions to be understood, and alternative splicing to be fully appreciated. These mechanistic investigations were in full swing by 1995. Ribozymes, having been discovered in 1982, were also well studied, and structural analysis was a major topic but still primitive. Most RNA scientists conceptualized the structures of ribozymes and other functional RNAs as flat base-paired secondary structures. The field was on the verge of jumping from 2D to 3D, with the crystal structure of one conformation of the hammerhead ribozyme having just been published by McKay's group. Crystal structures of a large ribozyme domain (1996) and a full active group I ribozyme (1998) were still in the future. Ribosome crystal structures were a dream in 1995, but by a few years later we had them.

Most striking is what we did not know in 1995:

- We did not know of RNA interference or siRNAs, miRNAs or piRNAs, the Fire and Mello paper still three years in the

future. Although the discovery of this large class of ncRNAs had been presaged in 1993 papers by the Ambros and Ruvkun labs, the generality of their initial example in nematode worms was unknown.

- We did not know of riboswitches, remarkable given their abundance. These RNA regulatory elements bind small-molecule metabolites and switch gene expression on and off at the level of transcription or translation or RNA splicing.
- We did not know much of bacterial sRNAs, small RNA regulators that act by base-pairing to mRNAs.
- We did not know that DNA elimination in ciliates was orchestrated by scnRNAs.
- We did not understand that bacteria had an RNA-based acquired immune system, stored in their DNA in the form of Clusters of Regularly Interspersed Short Palindromic Repeats (CRISPR), and we certainly did not know that this system would function in eukaryotes to provide a robust new tool for genome editing.
- We had a growing list of long noncoding RNAs (lncRNAs), but we did not appreciate the magnitude of the lncRNA universe or their incredibly diverse functions.

Now

RNA science has undergone a revolution in the past 20 years. We've moved from looking at one RNA-one protein systems (or small multiples thereof) to doing experiments genome-wide and transcriptome-wide. As has been the case for molecular biology since its birth, technology drives innovation, and rapid deep sequencing of nucleic acids and mass spectrometry proteomics now allow RNA scientists to do massively parallel experiments. Instead of measuring the level of a transcript relative to several control RNAs, we use RNAseq to measure the level of every transcript in the cell. Instead of assessing transcription from a gene with nuclear run-on assays, we use GROseq to capture instantaneous transcription genome-wide. Instead of testing whether RNA X binds to protein Y, we use RIPseq or CLIPseq or a related

Corresponding author: thomas.cech@colorado.edu

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.049965.115>. Freely available online through the RNA Open Access option.

© 2015 Cech This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

technique to catalog all of the RNAs in a particular cell type that are associated with protein Y. Instead of assessing translation by asking whether a particular RNA is “on polysomes,” we interrogate the coding potential of large genomes and evaluate translational mechanisms transcriptome-wide by “ribosome profiling.”

These high-throughput technologies have opened the door to a brave new world of RNA science. The technologies are disruptive—they transform the way we design experiments and the way we make discoveries. At the same time, they come associated with new limitations (and, dare I say, dangers) that are recognized by the research community but not always dealt with in a robust manner.

RNAseq and quantitative RT-PCR measurements can give distorted RNA abundances due to variable efficiencies of PCR, preferential degradation of certain RNAs, etc. RIPseq and CLIPseq protocols that involve antibody pull-down of a protein may reveal not only direct binding partners but also indirect interactors and RNAs that simply come along for the ride. In addition, there is the Tyranny of Large N. We demand that results be statistically significant, and in transcriptome-wide experiments the number of RNAs (N) is so large that papers commonly report *P* values of 10^{-30} or even 10^{-100} , giving the impression that the claimed relationship is unassailably proven. Although better procedures for controlling the false discovery rate are now commonly utilized, statistical significance may or may not equate to biological relevance.

Very importantly, the RNA community is not so enamored with transcriptome-wide experiments that we have forgotten the standards that we so laboriously and carefully established in past decades. First, an RNA secondary structure element is not believable just because sequence 1 is followed by a complementary sequence 1'. Instead, one must compare good alignments of sequences from different organisms and find enough covariations (e.g., A→C on sequence 1 is accompanied by U → G on sequence 1') to establish the existence of base-pairing. Alternatively, site-specific mutagenesis is used to engineer mutations and test whether second-site compensatory mutations are indeed compensatory for proper function in vivo. These same criteria, established for RNA secondary structure, are also applied to hypotheses of intermolecular base-pairing of RNA. Just because two RNAs have some stretch of complementarity and can be shown to form a hybrid in vitro does not mean that they undergo base pairing in vivo. Furthermore, when a given protein is

found to be associated with a cohort of RNAs using immunoprecipitation and deep sequencing, then the next steps are to identify the binding motif and to quantify binding affinity and specificity. If a hypothesis fails these tests, then it is not necessarily wrong, but perhaps just incomplete. For example, one or more partner proteins may be enhancing the binding specificity in vivo, and they need to be identified and incorporated into the biochemical assays before specificity emerges in vitro.

Tomorrow

A major challenge for RNA science in the next 20 years is to figure out how many of the tens of thousands of lncRNAs expressed in eukaryotic genomes are functional, and then to understand their mechanisms of action. This is a daunting task; given that 10–50 nucleotides of RNA are sufficient to bind a protein or bind a small molecule or base-pair with mRNA or even perform catalysis, clearly a 10 kb lncRNA could perform multiple functions, and perhaps different functions in different cell types or stages of development (see T.R. Cech & J.A. Steitz, *Cell* 157:77–94, 2014). Unraveling this complexity will require integrating the powerful transcriptome-wide studies with deep validation at the level of biochemistry, cell biology and structural biology, and computational biology will become even more central than it is now.

New experimental and computational tools will continue to lead the way. Genome editing with the CRISPR-Cas9 system promises to be very useful, allowing genetic alterations in mammalian cells as has been possible in the past with yeast and bacteria. High-throughput nucleic acid sequencing will continue to leap forward with new approaches, including single-molecule sequencing and identification of modified nucleotides. Real-time super-resolution microscopy will reveal the intricate movement of RNAs and RNPs in living systems. And electron microscopy at previously unforeseen near-atomic resolution will provide structures of the enormous RNP complexes that are central to RNA biology. These and other technological advances provide optimism that the next 20 years of RNA science will be just as discovery-filled as the last 20.

Acknowledgments

The author thanks Chen Davidovich for useful discussions. T.R.C. is an investigator of the Howard Hughes Medical Institute.

RNA

A PUBLICATION OF THE RNA SOCIETY

RNA World research—still evolving

Thomas R. Cech

RNA 2015 21: 474-475

Open Access

Freely available online through the *RNA* Open Access option.

Creative Commons License

This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:

<http://rnajournal.cshlp.org/subscriptions>