

Reflections on the RNA world

MIHAELA ZAVOLAN and WALTER KELLER

Biozentrum, University of Basel, CH-4056 Basel, Switzerland

The authors of these reflections are a generation apart, yet their career paths have much in common. After training in medicine, they both followed the call of “real science” to discover a passion for the world of RNAs, a little corner of which they are now exploring jointly. This is a brief account of these meanderings.

Walter Keller: For me the turning point came when, as a resident in human genetics, I read Jim Watson’s 1965 “Molecular Biology of the Gene,” which was a thrilling revelation. After a circuitous postdoctoral journey via the Johns Hopkins University Medical School and the National Institutes of Health I was appointed a Senior Staff Investigator in the Tumor Virus Group of the Cold Spring Harbor Laboratory, which was directed by Jim Watson. I worked with such illustrious colleagues as Bob Crouch, Ashley Dunn, Joe Sambrook, Phil Sharp, John Arrand, Rich Roberts, Bill Sugden and others. By purifying topoisomerase I and using it to generate circular topoisomers of SV40 DNA, in 1975 I managed to determine the number of superhelical turns of native SV40 DNA. When showing these results to Francis Crick, I was extremely surprised by his pointing out that I had shown, in fact, that the Watson–Crick structure of DNA must be correct. Ironically, I was not aware that there was a group of heretic X-ray crystallographers that had proposed an alternative “side by side model” of the DNA structure. My fame did not last long, because soon afterwards all doubts were resolved by X-ray crystallography of synthetic double-helices.

The 1970’s saw the characterization of fundamental steps of gene expression. I concentrated on the purification of RNA polymerase II and of other enzymes, such as poly(A) polymerase and RNase H from HeLa cells. However, in returning to the University of Heidelberg in the summer of 1976, I missed the dramatic events that surrounded the discovery of pre-mRNA splicing during the following winter by Phil Sharp (then at MIT) and Louise Chow, Tom Broker and Rich Roberts at Cold Spring Harbor. In an era preceding the ubiquitous presence of mobile devices I was taken entirely by surprise by the reports that were presented at the 1977 Cold Spring Harbor Symposium.

As group leader at the German Cancer Research Center in Heidelberg I entered the race for a system with which splicing could be reproduced *in vitro*. In 1983, Nouria Hernandez accomplished this by uncoupling transcription from splicing. David Frendewey studied the stepwise assembly of the mammalian spliceosome and Angela Krämer and Reinhard Lührmann showed in 1984 that the U1 snRNP recognizes the 5’ splice site by base pairing with the 5’ splice site of introns. In the years that followed, the space between genes and cellular processes started to be filled in increasing detail. The world of RNAs became larger and larger, RNA was found at the core of the ribosome, the RNA Society was established in 1993, and the *RNA* journal followed shortly after. I chose to pursue projects on pre-mRNA splicing, 3’-end processing and RNA editing, since 1983 at the Biozentrum in Basel, with the help of a large number of very competent students and postdocs, without whom I would have accomplished very little and to whom I am very grateful. The key players were Angela Krämer, Elmar Wahle, Joachim Lingner, Ursula Rügsegger, Pascal Preker, Isabelle Kaufmann, Andrea Kyburz, Lionel Minvielle-Sebastia, Silvia Barabino, Mary O’Connell, André Gerber, Jeannette Wolf, Stepanka Vanacova, Bernhard Dichtl, Christiane Rammelt and Georges Martin. After retiring in 2008 I had the good fortune to be able to continue the adventure with the support of grants from the Swiss National Science foundation. I have collaborated with Mihaela on further characterizing pre-mRNA 3’ end processing taking advantage of various high-throughput approaches that have emerged in the recent years.

Mihaela Zavolan: As a student in computational biology at the University of New Mexico, the Los Alamos National Laboratory and the Santa Fé Institute I was initially grappling with the mystery of adaptive immune systems, simulating the dynamics of immune responses to understand their speed and efficiency. At some point I found that the Genbank database of the National Center of Biotechnology Information (NCBI) already catalogued hundreds of sequences of immune receptor genes, annotated with functional domains. Analyzing these genes from the perspective of empirically inferred sequence-dependent rates of somatic hypermutation, Tom Kepler (then at North Carolina State University) and

Corresponding author: mihaela.zavolan@unibas.ch

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.049999.115>. Freely available online through the RNA Open Access option.

© 2015 Zavolan and Keller 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/>.

I found that evolution has shaped the codon bias of immune receptor genes such that somatic hypermutation preferentially targets the antigen-binding regions and avoids the structural parts of these molecules. Furthermore, by comparing the sequence-specific bias of somatic hypermutation with that of mutations that occurred during evolution, we inferred that somatic hypermutation appears to be the superposition of a mutational process similar to that taking place during evolution and a specific process that targets G/C nucleotides. The specific component is now known to be the activation-induced deaminase (AID), while the general component involves mismatch repair proteins and DNA polymerase ϵ . Then and now I find it astounding that computational analyses of nucleic acid sequence can bring us so far to understanding how organisms function.

As a postdoctoral fellow at the Rockefeller University I started to look at RNAs. I had the chance to work with hot-off-the-press mouse genome and transcript sequences, searching for functionally important splicing isoforms and rather finding that much of the observed transcript variation comes from stochasticity in the splicing process. At the Rockefeller University, I serendipitously joined Tom Tuschl in the race for finding miRNA genes in flies, fishes, viruses, humans and mice. MiRNAs burst on the scene in the year 2000, when two papers from the Ruvkun lab reported on the let-7 miRNA, which is perfectly conserved between worms and humans. The founder of this family of molecules was, however, discovered much earlier, when worm developmental phenotypes were traced to a gene encoding the very small non-coding RNA *lin-4* by the Ruvkun and Ambros labs. Also in the early 1990's studies from the Jorgensen lab in plants and from the Macino group in fungi found that introduction of transgenes led to the silencing of homologous genes in *trans*. All of these findings converged in the years that followed onto the RNA interference (RNAi) pathway, which probably originated as a mechanism by which organisms defended themselves against foreign genetic material, but has been co-opted in many guises in a variety of cellular processes, ranging from transcriptional silencing and silencing of repeat elements to the repression of protein production. Because RNAi can also be triggered by exogenous small interfering RNA (siRNA), it has become part of the basic molecular biology toolbox.

miRNAs are only one of many classes of RNAs that have been discovered in the recent years. They are quasi-ubiquitous, being encoded by genomes as small as those of the polyomaviruses, which are ~5 kilobases in length, to the human genome, which is ~3 gigabases and that of the Norway spruce, which is larger than 20 gigabases. Their roles in gene silencing are well established. Many other categories of small RNAs have been recently uncovered and are intensely studied. For example, the piRNAs regulate the expression primarily of repeat elements in germ cells, embryos and stem cell gametogenesis in animals. Furthermore, short RNA by-products are generated from transcription start sites

and splice sites and during other steps of RNA processing. Whether these molecules carry out independent functions is currently unknown. Intriguingly, some molecules that have been viewed as aberrant products of RNA processing reactions such as the circular RNAs, reported by the Bailleul group twenty years ago, were recently found by the Rajewsky and Kjems labs to have important biological functions.

A technology that is currently taking the world by storm makes use of the RNA-guided Cas enzymes, that bacteria have evolved as an anti-viral defence mechanism to target DNA and the Charpentier and Doudna groups have explored as biotechnology tools. The range of applications of this CRISPR (clustered regularly interspaced short palindromic repeats)/Cas system is continuously expanding and currently includes processes from transcription regulation to genome editing. The fantastic versatility of nucleic acids comes to a large extent from the simple but powerful base-pairing rules that not only enable transmission of genetic information but also allow us to interrogate a variety of aspects of cellular functions by sequencing RNAs or DNAs.

The quest for fast, cost-effective and high-throughput sequencing technologies gained momentum after the initial draft of the human genome was published in 2001. A variety of sequencing platforms emerged and their applications range from the more "mundane" sequencing of transcripts to estimate gene expression to the global analysis of mRNA translation in whole cells as well as specific subcellular compartments with approaches developed by the Weissmann lab. The sensitivity of sequencing methods is increasing as well, to the point that mRNA expression profiles can currently be obtained from single cells, and the van Oudenaarden group further improved the accuracy of mRNA expression level estimation by using molecular bar codes. Image-based transcriptomics approaches such as that developed by the Pelkmans lab have also been used for parallel quantification of RNA species in a large number of cells. These methods enable investigations into the sources and consequences of stochasticity in gene expression, particularly for cell differentiation and development.

Experiments that probe the accessibility of nucleotides to either Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE) or to endonucleases that are specific for double-stranded RNA (RNase V1) or single-stranded RNA (RNase S1) coupled with deep sequencing have provided the first insights into the RNA secondary structure in vivo. Surprisingly, the Weissmann group showed that mRNAs are actively unfolded in vivo. Some RNAs such as the ribosomal and spliceosomal RNAs are known to undergo specific modifications such as methylation and pseudo-uridylation that are presumably important for their stability and function. Most likely, only a very small fraction of the modification sites is currently known, and here again, methods for mapping RNA modifications genome-wide that are based on RNA sequencing and analysis have started to emerge. For example,

the sites of ribose 2'-O-methylation can be mapped taking advantage of the fact that these bases are resistant to alkaline hydrolysis, in contrast to their non-modified variants. This approach has been employed by the Nielsen lab to map modification sites on yeast rRNAs and we are awaiting the scaling of the method to the entire transcriptome.

Synthetic biology applications that make use of RNAs have also started to emerge and there is no doubt that the kaleidoscopic world of RNAs will continue to expand and complexify. Current challenges include increasing the resolution of the measurement technologies to the extent that we will be able to perform single molecule sequencing, localize individual long and small RNAs within individual cells and follow

the dynamics of RNA-protein interactions in vivo. Notably, the challenge that has been brought by the high-throughput measurement technologies, namely how to integrate large amounts of data into a conceptual understanding and translate them into predictive models, remains. Examples of approaches that have been proposed include the “splicing code,” developed by the Frey group for the prediction of tissue-specific splicing patterns and the “motif activity-response analysis” developed by the van Nimwegen group for the identification of key regulators of mRNA levels. It will be extremely interesting to learn what the application of such approaches to the emerging personalized high-throughput data will tell us about human diseases.



RNA

A PUBLICATION OF THE RNA SOCIETY

Reflections on the RNA world

Mihaela Zavolan and Walter Keller

RNA 2015 21: 531-533

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>
