My adventure in tRNA biology, so far

TAO PAN
Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, USA

How I got into tRNA biology

When the RNA journal started in March 1995, I was a second year Assistant Professor still trying to establish my research program. My postdoctoral work with Olke Uhlenbeck at the University of Colorado at Boulder was on finding new ribozymes through in vitro selection. I was very excited to continue the research in RNA catalysis and RNA folding in my own lab. Our first paper in RNA, published in June 1996, described our finding of the folding domains in bacterial RNase P RNA. This work has been cited 189 times as of December 2014 and led to the eventual success of the first high resolution RNase P RNA structure in 2003.

After I received tenure in 2000, I had a period of “tenure blues,” when I tried to figure out whether to stay the course of what I had done or venture into new directions. The decisive moment for me to change my research into biology was the publication of the draft human genome sequence in February 2001. I remember reading the Nature article where Table 20 showed the entire known classes of human RNA at the time: this table covered less than one half of a page! I thought that there must still be ample opportunities to study RNA biology. Indeed, the field of micro-RNA and RNA interference leapt into prominence shortly thereafter, and RNA biology rapidly expanded into all areas of biological inquiries.

With no prior experience and training in biological work, I decided to stay with at least something I knew: transfer RNA. I decided early on that my entry to success in tRNA biology would have to come from developing new techniques that are specifically tailored for tRNA. I was very fortunate to recruit a very talented graduate student, Kimberly Dittmar, to initiate this work in my lab. In 2001, all ‘omics research on nucleic acids were done using microarrays. Although commercial microarrays were very powerful, they were very expensive, costing ~$1000 per chip. Further, because of the abundance of post-transcriptional modifications that interfere with quantitative cDNA synthesis and extensive secondary structure that interfere with hybridization, tRNA expression was difficult to quantify. Standard methods at the time required cDNA synthesis for fluorescent labeling, and the commercial microarrays typically used short antisense oligonucleotides that were not efficient for tRNA hybridization on the chip. We came up with an idea to selectively label only tRNA in the total RNA without cDNA synthesis (based on the universal 3′CCA in all tRNA), and obtained custom-made microarrays with probes covering the entire length of the tRNA. These developments not only made the quantitative assessment of tRNA in biological samples possible, it also reduced the cost to ~$100 per array. The drawback of our method was a reduction of resolution from single nucleotide level for mRNA chips to ~8 nucleotides for tRNA. This was not a problem for bacterial and yeast work where almost all individual tRNA isoacceptors can be distinguished at this resolution. However, as I will describe below, this resolution was not useful in addressing certain questions on mammalian tRNA biology. Nevertheless, we successfully applied our microarray method for many aspects of tRNA biology in bacteria, yeast and mammals, including the identification of new tRNA species in HIV virions and the demonstration of initiator tRNA overexpression leading to global reprogramming of tRNA expression and increased proliferation, published in RNA in 2010 and 2013, respectively.

Surprises and potential new tRNA biology in mammals

When the complete human genome sequence became available in 2003, analysis of tRNA genes revealed two major surprises. It was not unusual to find that the human genome contained ~500 tRNA genes, similar to the number of tRNA genes in Drosophila and C. elegans, and less than twice that in S. cerevisiae. The first surprise was that ~150 of human tRNA genes are clustered together in chromosome 6 right next to the class I major histocompatibility (MHC) genes. MHC proteins display peptide antigens on cell’s surface for
immune surveillance, and the MHC sequences have the highest diversity in the human population. This genomic co-location between tRNA cluster and MHC genes may either be a coincidence (this association is absent in rodents) or suggest a previously unknown, functional linkage between tRNA and immunology. Jonathan Yewdell at NIAID and I came up with a hypothesis on linking tRNA and antigen presentation which involves tRNA misacylation producing misfolded proteins that are rapidly degraded as a source of peptides for MHC presentation. Our collaboration on testing this hypothesis instead ended up discovering regulated tRNA misacylation by reactive oxygen species (ROS) that produces “mutant” proteins useful for stress response in 2009. To this day, the tRNA-MHC association in the human genome has no good explanation.

The second surprise was the high diversity of mammalian tRNA gene sequences. For example, the ratio of tRNA sequences/tRNA gene copy number in *S. cerevisiae* is 5.4, i.e., each tRNA sequence has 5–6 identical genes on average in the yeast genome (275 genes, 51 sequences). This ratio for the reference human genome (and other mammalian genomes) is ~1.65, so there are ~300 different tRNA sequences among the ~500 tRNA genes. Additional tRNA gene sequences can be found in ~1600 genome sequences in the human population. For simplicity, we termed all tRNAs with the same anticodon but different body sequences as isodecoders. On the surface, the presence of abundant tRNA isodecoder genes is not good news. All tRNAs contain identity elements for correct charging by their cognate synthetases. At the same time, they also need to avoid mischarging by the 19 non-cognate synthetases in cells. tRNA isodecoders therefore may be more prone for mischarging. Clearly, no one believes that mammals maintain this large collection of tRNA isodecoders to harm themselves. Just the opposite: a recent report from Susan Ackerman lab shows that one tRNA isodecoder (Arg-UCU) is specifically expressed in the mouse central nervous system, and decreased expression of this isodecoder leads to neurodegeneration. This result indicates that one reason for mammals to maintain a large number of tRNA isodecoder genes is to enable cell-type specific expression that tunes the translation of specific mRNAs. Other functions of tRNA isodecoders may go beyond translation and involve their use for interacting with specific mRNA or proteins.

**My personal view on current challenges in mammalian tRNA biology**

The first challenge to significantly advance tRNA biology is again technical: we need a high throughput sequencing method that truly quantifies tRNA expression. High throughput sequencing is commonly based on cDNA synthesis. The high abundance of tRNA modifications, in particular those that interfere with cDNA synthesis (such as N1-methyl-A and N1-methyl-G) and the extensive tRNA structure make quantitative tRNA-seq very difficult. High resolution tRNA-seq is required because the microarray method does not have the resolution needed to study the expression and function of isodecoders.

The second challenge is the difficulty of manipulating tRNA expression in mammalian cells for functional studies. tRNA isoacceptors are present in high copy numbers in the genome, so to significantly alter expression requires extra efforts as standard methods for mRNA overexpression using transfection or knockdown using siRNA are typically ineffective to affect tRNA expression. The Crispr/Cas9 system currently offers the best hope to knockout specific tRNA isodecoder genes for their functional analysis. However, because many tRNA isodecoder genes have very similar sequences, off-target effect may be more problematic for tRNA studies.

The third challenge is to understand the many potential functions of tRNA modifications. Although most tRNA modifications are not essential for life, they can affect translation in specific ways and act on specific mRNAs in specific cell types. Many tRNA modification enzymes have now been associated with human diseases. Functions of specific modifications can now be studied transcriptome-wide using the ribosome-profiling method. I expect this area to fully prosper in the coming years.

**Acknowledgments**

tRNA biology research in my lab is currently funded by NIH (DP1GM105386). I thank all past and current members of my lab in this area for joining me on this adventure: Wesley Clark, Kimberly Dittmar, Kamela Dorobek, Molly Evans, Cristina Gomes, Jeffrey Goodenbour, Renaud Geslain, Thomas Jones, Mariana Pavon Eternod, Marc Parisien, Mridusmita Saikia, Michael Schwartz, Anna Walenciak, Xiaoyun Wang, Elizabeth Wiltrout, John Zaborske, Donghui Zhou, and Guanqun Zheng.