# RNA–RNA base-pairing: theme and variations

**JOAN STEITZ**
Boyer Center for Molecular Medicine, Yale University, New Haven, Connecticut 06536-0812, USA

A row of small white volumes marches along one bookshelf in my cluttered office. It is flanked on both ends by slightly larger volumes. The first is labeled "Processing of RNA, *Brookhaven Symposia in Biology, Number 26*," the proceedings of a meeting I attended in May, 1974, as a relatively new assistant professor with my undergraduate student David Ginsburg. The other larger volume is "RNA '96, *The First Annual Meeting of the RNA Society*." Most of the volumes in between are labeled "RNA Processing" and are Cold Spring Harbor meeting abstracts dating from 1982 to 1993. I am a terrible pack rat. But these volumes remind me that I was invited, along with John Dunn and Mike Matthews, to organize the first CSH RNA Processing meeting in 1982. RNA Processing meetings at Cold Spring Harbor were to be a biannual event. But already by 1986, as the pace of discovery in RNA accelerated, they became annual. By 1990 the size of the RNA Processing meeting severely stretched the limits of the CSH facility.

The RNA Society itself predates its first Annual Meeting by several years. There had been talk of establishing a society and in May, 1992, Tom Cech, Walter Keller, Olke Uhlenbeck and Alan Weiner organized an RNA Processing meeting held not at CSH but at the higher-capacity facility in Keystone, Colorado. By the time I arrived in Boulder to do a sabbatical year with Tom and Olke in August, 1992, legal papers incorporating the RNA Society in the state of Colorado had already been drawn up. Tom was Secretary, Olke was Treasurer, and they dubbed me President by fiat. I therefore "presided" over the decision to expand the annual meeting of the RNA Society from a focus on processing to RNA in general.

Concurrently, the need for an *RNA* journal was obvious. With the splicing field (both self-splicing and snRNP-catalyzed) burgeoning, even very good papers were not finding their rightful places in the Big Three journals. I took it upon myself to send a letter around to the RNA community admonishing that we were being too harsh on each other as referees, thereby hurting the field by restricting publication in the most fashionable journals. But we also needed a venue for publishing solid RNA science without concern about current glitz. Tim Nilsen was the natural leader and became nascent editor. I recall trudging the streets of Manhattan with him in the summer of 1994, interviewing contending publishers. We finally settled on Cambridge University Press and the first issue of *RNA* appeared in March, 1995. What a tremendous success the journal has been—mostly due to our fearless leader, Tim.

My notions regarding where the RNA field has been and where it is headed are shaped by my monolithic fixation on RNA–RNA base-pairing, a concept pivotal to almost every important advance my lab has made over several decades. Although I worked on RNA phage and on ribosome binding sites in mRNA as a graduate student and postdoc, respectively, it was not until well into my postdoc (~1970) that the idea of RNA secondary structure as an important element shaping the function of all RNAs—not just tRNAs—took hold. The concept of intermolecular RNA–RNA pairing had not yet surfaced—except for the example of codon-anticodon interactions.

RNA–RNA pairing between molecules came onto my radar screen one wintry day in 1974, when the telephone in my lab rang and an Australian voice explained that he wished to visit me regarding a topic of great mutual interest. The caller was Lynn Dalgarno from Canberra. He and his student John Shine had just correctly assigned the sequence at the 3′-end of 16S ribosomal RNA and hypothesized in the preprint he gave me that pairing between this region and mRNA might explain how bacterial ribosomes accurately initiate protein synthesis. The approximately 30 nt-long fragments of mRNA that I and others had isolated by ribosome protection indeed all contained polypurine stretches upstream of the initiator AUGs—but they were conserved in neither position nor sequence. I was immediately captivated by what I called the Shine-Dalgarno hypothesis and thought seriously about how to prove it correct. Our experiments published in 1975 used colicin E3 to cleave the 16S rRNA and create an analyzable rRNA–mRNA hybrid complex. They comprised the first direct evidence for rRNA participating in intermolecular base-pairing during ribosome function.

It was on another wintry day four years later (1978) that our foray into snRNPs and splicing began. A new issue of the journal *Nature* arrived, containing a letter describing

**Corresponding author: joan.steitz@yale.edu**
Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.050427.115. Freely available online through the *RNA* Open Access option.

autoantibodies in the sera of patients with mixed connective tissue disease and their recognition of a nuclear substance containing RNA and protein. The article caught my eye because the previous year, when I had unsuccessfully tried to raise antibodies against hnRNP proteins, someone had mentioned that certain patients might make such antibodies. My MD/PhD student Michael Lerner crossed the street to visit Yale rheumatologists that very afternoon and began his quest for the autoantigenic targets of patient sera. After a year of struggle, other serendipitous events enabled Michael to show that autoantibodies from Lupus patients often react with snRNPs, which contain small U-rich nuclear (sn)RNAs (U1 and U2 were already in the literature; we described U4, U5 and U6). The potential for the 5′-end sequence of U1 snRNA to base-pair with the growing roster of consensus sequences at the 5′-ends of introns was obvious; Rogers and Wall published the same idea also in 1980. We initially hypothesized that U1 might extend its interactions to the 3′-splice site, but later experiments by my students Steve Mount, Doug Black, Benoit Chabot, David Wassarman and Erik Sontheimer revealed pre-mRNA interactions not only with U1, but also with U2, U5 and U6, confirming their involvement in splicing. Elegant genetic suppression experiments from Alan Weiner's and Christine Guthrie's labs were instrumental in confirming snRNA-pre-mRNA base-pairing.

Examining other abundant, small-sized nuclear RNAs proved to be a gold mine for unveiling additional intermolecular RNA–RNA interactions. Following the initial characterization of base-pairing interactions in the spliceosome, my student Kim Mowry sequenced the low-abundance human U7 snRNA and found its 5′-end sequence to be complementary to the conserved purine-rich downstream element beyond the 3′-cleavage sites in histone pre-mRNAs; Max Birnstiel's lab had just previously observed complementarity between sea urchin U1 and histone pre-mRNAs and had performed compensatory mutations supporting intermolecular hybrid formation in histone pre-mRNA processing. Next came the discovery of the minor (U12-dependent) spliceosome by my postdoc Woan-Yuh Tarn. The base-pairing interactions made by U11 and U12 snRNAs (discovered earlier by my student Karen Montzka Wassarman) are homologous to those of U1 and U2 but employ different consensus sequences; U6atac snRNA (discovered by Tarn), like U6, pairs with U4atac and the 5′-splice site. In 1996, Kazio Tycowski (research scientist) and Christine Smith (graduate student) contemporaneously with the labs of Bachellerie and Kiss realized that BoxC/D small nucleolar (sno)RNAs exhibit complementarity to conserved sequences in rRNA adjacent to sites of 2′-O-methylation. The concept of base-pairing guiding RNA modification was subsequently extended to pseudouridylation by the box H/ACA class of snoRNAs (Fournier and Kiss labs) and to the small Cajal body RNAs (scaRNAs) that introduce comparable modifications into snRNAs (Kiss lab and Tycowski).

Meanwhile, our efforts to assign functions to noncoding (nc)RNAs produced by gamma herpesviruses (initiated in 1981) languished. But recently (since 2010) great strides have been made, yielding molecular insights that further expand the catalog of diverse roles for RNA–RNA base-pairing in cell regulation. Demian Cazalla (postdoc) discerned complementarity between the seed sequences of several host micro (mi)RNAs and two of the HSURs (*Herpesvirus saimiri* U RNAs, which mimic the splicing snRNPs) produced in infected monkey T cells. One of the microRNAs is targeted for decay, revealing that viral evolution has repurposed an snRNA for a role in RNA degradation rather than RNA processing. Second, Nara Lee (postdoc) has discovered that EBER2 (a highly abundant nuclear ncRNA produced by EBV) base-pairs with nascent transcripts crossing the terminal repeat (TR) region at the termini of the linear EBV genome; in doing so EBER2 promotes the binding of PAX5, a master transcriptional regulator of B cell function, to its consensus sites in the TRs. Base-pairing between EBER2 and the TR transcripts is conserved in a divergent lymphocryptovirus and apparently contributes to EBV lytic DNA replication, with implications for EBER-induced tumorigenesis. Although not intermolecular, the structure determined by Rachel Mitten-Fry (postdoc) of an element that stabilizes the KSHV PAN ncRNA by forming a triple helix with its polyA tail has led to new insights into triple helix formation by Jessica Brown (postdoc) from studies of the cellular MALAT1 ncRNA.

The past 20 years have been marked by the discovery of whole new classes of regulatory RNAs. Most stunning are the 22 nt miRNAs and other tiny RNAs, which act by base-pairing with transcripts of protein-coding genes. MiRNAs were always visible on polyacrylamide gels used to fractionate metabolically-labeled RNAs ranging in size up to 300 nt, but were dismissed as uninteresting degradation products running at the front; it took the genetic insights of Victor Ambros and Gary Ruvkun to open our minds to their vast regulatory potential. An equally remarkable realization came from deep sequencing—evidence for pervasive transcription of genomes. Even though the products may be extremely low in abundance, such ncRNAs (many of which are quite long) could play important roles in cell signaling and regulation, as well as in chromatin architecture. Although investigating their functions has just begun, I expect that many more pivotal intermolecular RNA–RNA base-pairing interactions will emerge. There are at least two challenges that confront us in these efforts. One is what I call the "black hole" of RNA biology: RNAs of 50 to 300 nt have simply not been analyzed by deep sequencing. Second, the discard of repeated sequences in transcriptome analyses is likewise foolish: Since roughly half our DNA is comprised of SINES, LINES, pseudogenes, endogenous viruses and other repeats, ignoring this RNA landscape is like walking blindfolded into a beautiful wilderness.