On the occasion of the 20th anniversary of the *RNA* journal

GIDEON DREYFUSS

Howard Hughes Medical Institute, Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA

On the wonderful occasion of the *RNA* journal's 20th anniversary, I think one can say without a doubt that the RNA field has been an amazing source of excitement and remarkable advances that have inseminated multiple fields across biology and medicine. The *RNA* journal reflects the vibrant and collegial RNA community, but the journal's success is also a tribute to the superb leadership, integrity, and scholarship of its Editor, Tim Nilsen, and dedicated Editorial Board. I feel very fortunate to have been working in this exciting field no less because of the many friendships I have made with wonderful colleagues and mostly the truly exceptional individuals that I have had in my laboratory. The journey, so far, has been phenomenal and full of surprises. Looking back, I would not have been able to foresee the course that my laboratory took since the journal's inception. I will briefly describe some highlights that stand out in my mind from our own research and how we got there. I regret that the concise format precludes mentioning the major contributions and influence that others have had on our work—they are many!

My laboratory has a long-standing interest in RNA-binding proteins, RNA-protein complexes (RNPs), and their roles in gene regulation and disease. Most gene regulation in complex eukaryotes occurs post-transcriptionally and is mediated by RNA-binding proteins and small noncoding RNAs. To produce mRNA, the primary gene transcripts of the majority of protein coding genes (pre-mRNAs; historically hnRNAs) are extensively processed to remove translation open reading frame-disrupting introns by splicing, and by 5′-end capping and 3′-end cleavage and polyadenylation (CPA). Splicing is mediated by the spliceosome, comprised of non-coding small nuclear RNPs (major: U1, U2, U4, U5, U6; minor: U11, U12, U4atac, U6atac) and protein factors. The majority of pre-mRNAs have multiple introns, each with 5′- and 3′-splice sites (ss) and multiple polyadenylation signals (PASs) that can be utilized in various alternative combinations to produce diverse mRNA and protein isoforms from the same gene. Pre-mRNA processing initiates co-transcriptionally and completes in the nucleus. The mRNAs are then transported to the cytoplasm where they can be translated and are subsequently degraded. Each of these events is regulated in a cell type and cell cycle stage-dependent manner and in response to external cues. All these processes, including recognition of constitutive and alternative pre-mRNA processing signals, depend on RNA-binding proteins (RBPs or RNP proteins).

RNA-binding proteins, hnRNPs, and a pathway from transcription sites to ribosomes

As they are transcribed, nascent gene transcripts become densely decorated with proteins (hnRNP proteins). The bound proteins make chromatin-attached RNP fibrils noted by 19th century cytologists as lampbrush-like chromosomes in amphibian oocytes and as the "puffs" of dipteran polytene chromosomes. By the mid-late 20th century it became clear that mRNAs are processed within these fibrils/hnRNP complexes by splicing of hnRNAs (later re-named as pre-mRNAs). However, for over a century, the fragility of these large macromolecular complexes and shortcomings in experimental methods forced the identity of hnRNP proteins (and cytoplasmic mRNPs) to remain elusive and their functions unknown.

Beginning in the early 1980s my laboratory developed a powerful experimental approach to identify and characterize the proteins that interact with pre-mRNAs and mRNAs. Using UV light to photoactivate RNAs and to crosslink them to bound proteins in living cells, we could purify hnRNAs and mRNAs under protein denaturing conditions, and thus isolate only proteins that were crosslinked to RNA. Since this predated sensitive mass spectrometry, we immunized mice with the resulting material and generated monoclonal antibodies to RBPs—in turn allowing us to identify and clone the principal hnRNP and mRNP proteins (>20), and leading to the discovery of the major RNA-binding motifs (RNP consensus [RBD/RRM], KH-, and RGG-domain). It was clear that this was just the tip of the iceberg as even with 2D gel electrophoresis of hnRNP complexes and RNA affinity chromatography showed many more, albeit less abundant RBPs. Over the last twenty plus years, I nevertheless watched with awe as an ever-increasing compendium

of genomic sequences located these motifs in an enormous assortment of RBPs (>5% of human genes) involved in every aspect of RNA processing and function, and connecting their perturbations to many diseases. Combining UV crosslinking with mass spectrometry and deep sequencing, great progress has been more recently made toward comprehensive identification of numerous additional RBPs and genome wide mapping of their binding sites.

Counter to the view that there was only a small number of "core" hnRNP proteins, "packaging" RNA in nucleosome-like, non-specific "beads-on-a-string," we went on to show that each hnRNP and mRNP protein has a distinct RNA-binding specificity, which drives assembly of a specific constellation of hnRNP proteins on each pre-mRNA, sculpting its presentation to the processing machineries and determining the mRNA it produces and its fate.

Converging biochemical and cytological experiments, we visualized hnRNP proteins' binding to nascent pre-mRNAs at chromosomal transcription sites and then made the surprising observation of nucleo-cytoplasmic shuttling, showing that many of them remain bound through splicing and transport to the cytoplasm, where, as we proposed, they also have functions in mRNA transport, translation, localization, and stability. Our studies of nucleo-cytoplasmic hnRNP shuttling, later generalized to numerous proteins, traced a coordinated mRNA biogenesis path, from gene to mRNA translation on ribosomes. We further delineated novel nuclear import and export signals, as well as the transport receptors (transportins) that choreograph RNP trafficking.

Establishing perhaps the first connection between RNA-binding proteins and disease, my laboratory identified the Fragile X mental retardation syndrome (FMR1) protein as a (KH-domain) RNA-binding protein, an activity we showed is impaired by a patient's mutation. In parallel, other observations, also disease-related, began to unfold that opened up new and irresistable areas, eventually becoming some of our main interests in recent years. I will focus here on two of these.

The SMN complex: RNA-protein chaperone, snRNP assembly, spinal muscular atrophy (SMA)

SMN (survival of motor neurons) was identified as the SMA disease gene by Melki and colleagues in 1995. Independently, as an offshoot of our studies on hnRNP proteins, we discovered SMN and the multi-protein complex it forms, with proteins we termed Gemins. We established that the SMN complex is critical for biogenesis of snRNPs. We further defined the function of the SMN complex in a key step: chaperoning assembly of a heptameric Sm protein ring (Sm core) on each spliceosomal snRNA. This was unexpected, as RNPs were previously believed to form by self-assembly. Indeed, Sm proteins have the propensity to nonspecifically assemble Sm cores on RNAs, which undoubtedly would be deleterious. We demonstrated that the SMN complex is crucial for preventing illicit assembly and identified the specificity determi-

nants it recognizes in Sm proteins and snRNAs. Using biochemical assays and inhibitors we identified by high throughput screening, we dissected subunits and intermediates in a stepwise snRNP biogenesis pathway. The structure of a key intermediate provided important insights into the mechanism of this first-of-its-kind RNP assembly device at atomic resolution, revealing how cells distinguish the abundant non-coding snRNAs from all other RNA classes.

Our interest in the SMN complex derived both from its fundamental role in RNA metabolism and gene expression, and from its role in SMA, which is caused by SMN deficiency. As expected from its function, the SMN complex is expressed in and required for viability of all cells across eukaryotes, but the basis for the clinical manifestation, primarily in motor neurons, was not understood. Towards this we profiled the transcriptome in motor neurons of SMA mice, showing that SMN deficiency causes widespread, tissue-specific RNA metabolism perturbations. These studies advanced understanding of SMA pathogenesis and the prospects of therapy for this leading hereditary cause of infant mortality.

U1 snRNP protects pre-mRNAs from premature termination and determines mRNA length

Stimulated by the observations of snRNP changes in SMA, we asked what effect modulating snRNP repertoire might have. Even at baseline, despite their 1:1 stoichiometry in the spliceosome, snRNP are not equi-molar in cells. In fact it was noted but unexplained since snRNAs were first discovered, in the 1960s, that U1 snRNA is much more abundant than other snRNAs in human cells. We therefore systematically inactivated individual snRNPs and probed for its transcriptome effects. This led to the surprising observation that U1 snRNP (U1) inhibition caused pre-mature termination in the majority of pre-mRNAs. Our experiments revealed that in addition to and separate from its function in splicing, U1 is a suppressor of premature cleavage and polyadenylation from cryptic PASs in introns. We termed this activity telescripting, as it allows transcription to go farther. Like splicing, telescripting depends on U1 base-pairing to pre-mRNA. U1's limited telescripting range (~1 kb) suggests that U1 base-paired at 5'ss would be insufficient to suppress PASs in large introns. Because U1 telescripting is effective even when base-paired to sequences that cannot function as 5'ss, we proposed that additional U1s bind in introns—to protect them, providing a plausible explanation for U1 overabundance.

In addition to its essential role in protecting pre-mRNAs, telescripting also plays a role in regulating mRNA length, by affecting alternative PAS selection in 3'-untranslated regions (3' UTRs). We are studying the potential role of U1 in the widespread mRNA shortening of 3' UTRs occurring in cancer, proliferating cells, and activated immune cells and neurons. Telescripting is a major step in gene expression regulation; however, its mechanism is unknown, leaving a major challenge that we are pursuing.