Reflections on the RNA journal at 20 and its role in the question of splice site choice

DOUG BLACK
Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, California 90095-1662, USA

My first response to hearing that the RNA journal was turning 20 was one of some alarm—how can that much time have gone by and so little progress made? But the task of commemorating this anniversary brings back mostly pleasurable thoughts and of course a lot of progress has been made, just not always in the directions one might have anticipated. My own contribution to RNA Volume 1 was a somewhat florid review of questions regarding how splice sites are chosen for pairing within the spliceosome and some then-recent concepts in the understanding splice site recognition.

The motivation for the 1995 review was to discuss the concept of exon definition recently proposed by Berget. There had been several studies by the Berget, Grabowski, and Reed labs among others that provided experimental support for the idea that exons are recognized as a unit prior to splice site pairing across the intron. It was shown that exon definition was not essential if an intron was very short (as in yeast), but the requirements for spliceosome assembly changed as introns were extended. Exon definition nicely explained the exon skipping and cryptic splice site activation phenotypes of single splice site mutations, as well as the relatively narrow range of lengths for internal exons. Specialized interactions between the Cap and the 5′ splice site and between the 3′ splice site and the poly-A site were proposed, with some support, to extend exon definition to 5′ and 3′ terminal exons. Unusually short and unusually long exons were found to have different properties from typical exons presumably due to different mechanisms of exon recognition. Our own studies had centered on a very short neuron specific exon in the Src transcript that I had proposed was skipped due to interference between complexes assembling at each end. We were coming to view this hypothesis as likely incorrect, as the data could also be explained by the short exon length disallowing the newly-defined exon definition interactions, or by the presence of a newly identified splicing repressor. Interestingly, these exons of nonstandard length were often highly conserved alternative exons implying that their regulation and their unusual structure were linked. Classic examples of this were unusually long and unusually short exons found in the NCAM and cardiac Troponin T transcripts, and other early systems where the tissue specific regulation of alternative splicing was studied.

In parallel with these studies of exon/intron architecture, several groups had discovered the SR proteins. These proteins contained one or more RRM-type RNA binding domains linked to an SR domain carrying repeated Arg-Ser dipeptides. These domains, which were found to be highly phosphorylated, are now called low complexity or intrinsically disordered domains. It was quickly recognized that many proteins contained SR domains, in addition to the original SR proteins, similar serine-arginine-rich sequences were seen in core spliceosomal components such as U2AF and U1-70K, as well as in the emerging class of splicing regulators first identified in the Drosophila sex determination pathway. The classical SR proteins were shown to bind to sequence elements in exons (eventually called exonic splicing enhancers), and in 1995 the role of the SR domain in stabilizing exon definition complexes was being actively studied. These proteins, as well as members of the hnRNP group of proteins and additional proteins in the Drosophila sex determination pathway, proved to be initial examples of the very large group of pre-mRNA binding proteins that serve to regulate spliceosome assembly and thus alter splice site choices. Both the concept of exon definition and the key role of SR proteins in exon complex assembly remain fundamental to our understanding of how splice sites are chosen.

In addition to the SR proteins and exon definition, a number of ideas were outlined in the 1995 review to explain how a 5′ splice site could ultimately find its correct 3′ splice site partner across an intron that might be 100,000 nucleotides long. Proposed mechanisms included coupling the search for splice sites with transcription, an idea that I believe originated with Charles Weissmann. There were also possible mechanisms for shortening the effective distance of the intron bridge and thereby simplifying the search. These included the recursive splicing discovered by Lopez in the Ubx transcript that allowed a very long intron to be shortened
by initial shorter splicing events. Other mechanisms to bring the splice sites closer together were the formation of RNA secondary structures and the packaging of introns into large RNP structures, presumably by the hnRNP proteins. All of these ideas offered the potential to shorten the effective length of the intron allowing the 5′ and 3′ splice sites to find each other more easily. In the ensuing years evidence has accumulated that each of these processes—transcription coupling, recursive and other forms of repeated splicing, intronic secondary structure, and RNP packaging—all contribute to proper splice site pairing in particular transcripts. However, it is not clear that any of them is universally important and/or able to provide a general mechanism for spliceosome assembly. It may be that they all make modest contributions to an as yet undefined unifying mechanism, or that different solutions have arisen for each exon-exon joining event and that each of these processes varies in significance for each splicing choice to be made. The genome-wide methods of analysis that were not yet on the horizon in 1995 have illuminated the highly variable nature of regulation between different genes and individual splicing events.

Thinking about our field since the first RNA Journal, one tends to draw a line demarking before and after the genome. As with other fields, the completion of the whole genome sequences and the emergence of genome-wide methods of analysis enabled the examination of different kinds of questions. New methods such as comparative genomics, RNAseq, and in vivo crosslinking-immunoprecipitation have become the first level of analysis for almost any study. Among the benefits of the early whole genome/whole transcriptome analyses of gene function was that they brought the problem defining pre-mRNA splicing patterns to the attention of the rest of the biology community. Prior to the genome sequences, between 500 and 1000 papers were being published each year describing alternatively spliced isoforms arising from individual genes. To those in the field, this made it seem obvious that splicing regulation was pervasive in metazoan genes. However, it was the ability to quantify isoform numbers starting with EST data and more thoroughly by RNAseq that allowed the definite statement that splicing choices were being regulated for nearly all genes.

Similarly, the ability to define in vivo the genome-wide binding sites of splicing regulatory proteins has not yet contributed many new conceptual principles to our understanding of the mechanisms behind splice site choices. For example, ideas and experiments going back to the Drosophila sex determination pathway had made clear that individual splicing choices were being controlled by complex combinations of protein factors, and that the position of binding for individual regulators could determine the positive or negative valence of their effect. However, these methods have again allowed the breadth of RNA-level regulation to be quantified. One example where the genome-wide analysis has yielded new mechanistic understanding is the phenomenon of tele- scripting. It was known for many years that the U1A protein of the U1 snRNP could inhibit polyadenylation in certain transcripts, but the genome-wide characterization of U1 function uncovered that this serves as a widespread mechanism to prevent cryptic polyadenylation in the very long primary transcripts of eukaryotic genes. With RNAseq, the various forms of CLIPseq, and other new methods, the complexity of the regulatory networks that determine the expression of individual spliced isoforms can finally be measured. The biological significance of these splicing regulatory programs has been powerfully demonstrated through genetics. And a whole new field of statistical analysis of these networks promises to allow a systems level understanding of splicing regulation and its integration into larger cellular function and organismal development. From the point of view of understanding mechanisms, the high-throughput methods have also allowed important in vivo confirmation of interactions and mechanisms studied in vitro, where detailed results made possible by in vitro analysis can now often be definitively described in their cellular context. Indeed, the choice of individual system for study in vitro often now starts with an in vivo data set.

Our ability to study the mechanistic questions of how splice sites are chosen to be assembled into a spliceosome has also changed dramatically in 20 years. Sophisticated and comprehensive genetic screens in yeast, RNAi screens in metazoans, and the analysis of RNA/Protein assemblies by mass spectrometry have again brought the complexity of the splicing process into clear view. The identification of the hundreds of proteins involved in even the most simple of intron excision events, the regulation of spliceosome assembly by DEAD/DEAH box helicases and other proteins, and the conformational changes in the spliceosome that are driven by these proteins have all been major advances toward the goal of a detailed understanding of spliceosome assembly. The development of single molecule methods and their application to the spliceosome have allowed the definitive analysis of its assembly kinetics and proven that even a simple spliceosome can assemble by multiple pathways. Our understanding of the catalytic mechanism of the spliceosome, a separate problem from how splice sites get to the active center, has made particular strides. Important questions from 1995 regarding the roles of metal ions and the Prp8 protein in the catalytic center, and the relationship of the RNA-catalyzed Group II intron mechanism to the spliceosomal mechanism, have been resolved through heroic structural, biochemical, and genetic studies. Nevertheless, there are enormous hurdles to overcome—notably the structure of the larger spliceosome and the conformational changes leading to the assembly of its active site.

Happily, I find that many of the questions that we discussed in 1995 regarding how splice sites are chosen still fascinate me. Even with all the progress and new technologies there are many questions regarding the mechanisms of splicing regulation that are essential to understand. What is the relationship between spliceosome assembly and transcription? We still need a detailed understanding of the rates of RNA
synthesis, spliceosome assembly, and catalysis to address this question. How does the nuclear environment affect splice site recognition and spliceosome assembly rates? How does the coalescence of splicing regulators and other proteins through their low complexity domains affect intranuclear structure and activity? How do the splicing regulators binding to the pre-mRNA actually interact with the spliceosome? What is the result of these interactions in changing the assembly or subsequent contacts of a spliceosomal component? How does the combinatorial control of an exon work? What interactions allow some regulators to synergize and some to antagonize each other? What does an exon definition complex look like? How is it different from an early spliceosome assembled across an intron? How do two exon definition complexes interact to form a spliceosome? These questions about the nature of exon definition complexes were being posed in 1995 and are still mysterious. I think that very recent advances in electron microscopy hold exciting promise for finally making them tractable.

I find that after 20 years of RNA, these processes are still my greatest interest and I think that they are key to understanding gene function. Many of our motivating questions are still unanswered but are now more easily studied. In discussing our field since the first RNA Journal, one must mention the role of its editor, whose interest in splicing and deep knowledge of spliceosome function has led to the publication of many of the most interesting papers in our field. Over the years, many of my most interesting discussions about these questions have been with Dr. Nilsen. Thanks Tim!
Reflections on the RNA journal at 20 and its role in the question of splice site choice

Doug Black

RNA 2015 21: 497-499

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

© 2015 Black; Published by Cold Spring Harbor Laboratory Press for the RNA Society