

The Journey to RNA Splicing and the Spliceosome

Phillip A. Sharp, Institute Professor, Department of Biology and the Koch Institute for Integrative Cancer Research at MIT

In keeping with the objective of this meeting “Forty Years of mRNA Splicing: From Discovery to Therapeutics” to provide historical perspective, I have outlined recollections of some of the events surrounding the discovery of RNA splicing. The focus will be on contributions from the MIT group. My account is clearly self-serving and omits many other, highly important, contributions to the field; hopefully, others will add their recollections of events to the meeting’s web site.

THE TRANSITION TO MOLECULAR AND CELL BIOLOGY.

During the last year of my graduate studies under Victor Bloomfield at the University of Illinois, I read the annual Cold Spring Harbor Symposium volume of 1966 with the title *The Genetic Code* where the structures of chromosomes were reviewed (1). At the time, I faced two major steps in my career, writing a thesis about the physical chemistry of stiff polymers—such as DNA—and deciding on the type of position to target for a job search. The research outlined in the Genetic Code symposium summarized the current status of measuring and characterizing chromosomes. Even for chromosomes as simple as that of a bacteriophage, this presented a challenge in 1969. Chromosomes seemed clearly to be units containing the triplet genetic code organized in genes and aligned in a linear array. They were seen to vary enormously in length but whether each chromosome consisted of a continuous segment of DNA was unknown at the time. To a physical chemist, this frontier in biology seemed to be ripe for new discoveries and I was fortunate to be accepted by Norman Davidson at Caltech as a postdoctoral fellow.

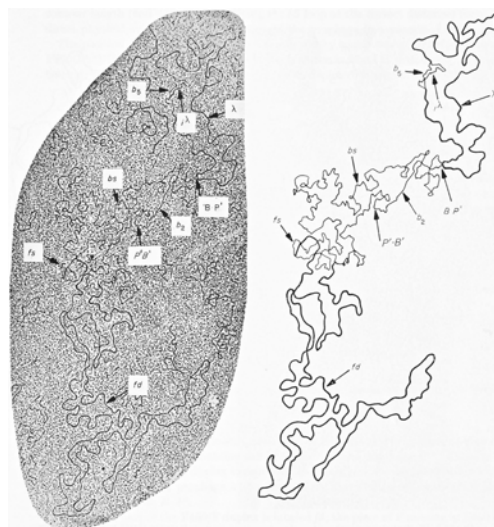


Figure 1. Electron micrograph of a three-part hybrid showing the position of integrated λ phage in relation to the F' sex factor plasmid in the F' factor. The Campbell model for integrated λ phage is apparent at the top right as a double strand and the F factor is at the bottom as a double strand. The two single strand arms are E. coli genome DNA. See reference (3) for details.

CALTECH WITH NORMAN DAVIDSON.

Norman had a highly distinguished career as a physical and inorganic chemist before expanding his research to biological science about a decade before my arrival. Also, just before I arrived, Ron Davis, a graduate student in his lab, who is now famous for his science at Stanford University, had developed methods to map genes on small chromosomes, in the first case λ phage, using heteroduplexes visualized by electron microscopy (Davis and Davidson, 1968) (2). My objective at Caltech was to map genes on the larger *E. coli* bacteria genome, which was impossible to isolate intact. Thus, I developed methods to isolate bacterial

chromosome segments as part of F-prime sex factor plasmids as shown in **Figure 1 (3)**, where integrated λ phage is visualized and mapped relative to the boundaries of the sex factor (3). While writing the papers describing this research, we introduced the term “kb” for kilobase, because the DNA segments were hundreds of thousands of base pairs in length. These studies by Davidson were an early harbinger of the transformative genome science that arose following recombinant DNA and DNA sequencing technology.

While at Caltech 1969–71, I followed research about the structure and expression of RNAs from mammalian genes. This included the early description of heterogeneous nuclear RNA (hnRNA), the presence of poly(A) on the 3' end of mRNAs in 1971. These experiments primarily used technology of labeling of RNA with radioactivity and measuring its length by sedimentation and hybridization to DNA sequences. Even at this early stage in my research, I recognized that the question of the relationship between nuclear RNA and cytoplasmic mRNA was important, but was difficult to determine with the available technology. For example, after struggling to characterize genes on bacterial genomes, the challenge of studying by methods such as electron microscopy RNA transcribed from the much larger and poorly defined mammalian genome seemed to be impossible at the time. However, having also collaborated during my time at Caltech with Jerome Vinograd to investigate the DNA tumor Simian virus SV40 using electron microscopy, characterizing RNA from these types of small genomes seemed feasible. Thus, DNA tumor viruses whose genes are expressed in the nucleus and mRNA translated in the cytoplasm appeared to be the most accessible biological system to investigate the relationship of hnRNA to mRNA.

COLD SPRING HARBOR LABORATORY AND ADENOVIRUSES.

As I completed my postdoctoral fellowship at Caltech, I considered a few independent positions, but decided that I wanted to switch my research from bacterial systems to mammalian cells and needed research experience in these systems. Luckily, Jim Watson offered a postdoctoral slot at CSHL, and my family moved there during the summer of 1971. It was a busy and fun summer as I attended the virology course and we shared a house with Janet Defoe and Ron Davis. When research began in the fall, a collaboration with Joe Sambrook and Walter Keller led to a transcriptional map of early and late SV40 mRNAs using separated strands (Sambrook et al., 1972) (4). As expected, the early region was expressed in rodent tumors caused by SV40 infections. More relevant to the future discovery of RNA splicing, my officemate at CSHL was Ulf Pettersson, who had just completed a Ph.D. studying human adenovirus with Lennart Philipson in Uppsala, Sweden. Ulf taught the lab how to work with adenovirus. He also provided a source of deep knowledge about the virology of adenovirus, and we spent many pleasant evenings discussing early research in virology. Ulf was interested primarily in DNA replication, while I focused on transcription. This was the time, 1972, when Herb Boyer, Dan Nathans, and Ham Smith reported the biochemical activities of restriction endonucleases. I developed a convenient assay for these activities using ethidium–bromide-stained agarose gel electrophoresis (Sharp et al., 1973) (5). Then, in collaboration with others including Ulf, we determined the cleavage maps of several adenovirus serotypes (Pettersson et al., 1973) (6). We then discovered that only one region of the virus was retained in tumor cells, the E1A region (Sharp et al., 1974) (7) and used restriction length polymorphisms to map the position on the

genome of several viral proteins (Sambrook et al., 1975) (8). Sarah Jane Flint, a new postdoctoral fellow at CSHL, was a special collaborator on transcriptional mapping of adenovirus. She moved to MIT in 1974 with me to continue research in this area (Flint et al., 1976) (9).

MIT AS A MEMBER OF THE CENTER FOR
CANCER RESEARCH.

I was recruited to the newly established Center for Cancer Research at MIT by its director, Salvador Luria, and was fortunate to be assigned lab space next door to David Baltimore, Robert Weinberg, and Nancy Hopkins. As the lab opened, Jane Flint further investigated the transcription of adenovirus, including a characterization of the copies per cell of both nuclear and cytoplasmic RNA during the early and late stages of infection, **Figure 2** (9). These results showed that during the late stage very significant levels of nuclear RNA accumulated from certain regions of the viral genome with almost no corresponding cytoplasmic RNA. They presented clear questions about the proposed relationship between nuclear hnRNA and cytoplasmic mRNA, and motivated a comparison of the two populations.

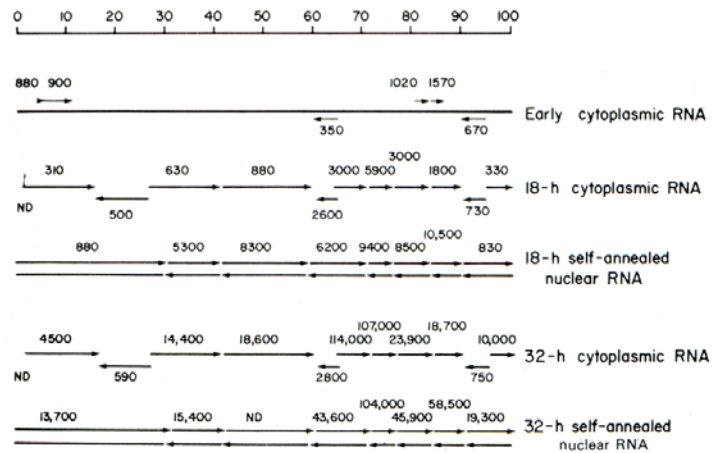


Figure 2. Quantitation of adenovirus RNA from cytoplasm and nucleus of early and late infected cells. Fragments of the viral genome were hybridized as separated ^{32}P labeled strands and copies per cell were calculated as compared to DNA standards. The viral genome is shown at the top in units of one hundred. There are four regions expressed early E1 through E4. Note the presence of abundant nuclear RNA during late stages of infection, particularly from the (r) strand transcribed to the right. See reference (9) for details.

As I mentioned above, previous papers, primarily from Jim Darnell's lab, described the puzzle of long nuclear RNA and shorter cytoplasmic mRNA. These two populations shared RNA sequences coding for proteins but their relationship was not clear. While at Caltech, I began to follow these papers and those of other scientists, about heterogeneous nuclear RNA. As a new-comer to the field of molecular biology, it was puzzling that more scientists were not investigating the question of the role of hnRNA. Without knowing the biochemistry of how genes were expressed, questions such as development and cancer seemed beyond serious science. In parallel with Ulf's joining CSHL, his mentor Lennart Philipson joined Jim Darnell's lab for a sabbatical. With Randy Wall as first author, they reported the presence of long nuclear adenoviral RNA as compared to shorter cytoplasmic mRNAs during the late stage of infections (Wall et al., 1972) (10). Also in 1975, the 5' end of mRNA was discovered to be modified by the addition of a cap structure (see Darnell's book, *RNA: Life's Indispensable Molecule*, (11)). The same modification was found on hnRNA and, when linked to the previous finding that 3' poly-A segments were common with both RNA populations, presented a complex picture of their possible relationship. It seemed reasonable that characterization of nuclear and cytoplasmic

viral RNAs from the late stage of infections could reveal their biochemical relationship and possibly that of hnRNA/mRNA transcribed from cellular genes.

Mapping by electron microscopy of heteroduplexes, such as that done at Caltech, and RNA/DNA hybrids of cytoplasmic viral RNA on the viral genome, required identifying novel features relative to known markers on each observed structure. In 1976, several labs reported electron microscopy characterization of R-loops where the RNA displaced a DNA strand permitting the visualization of the length and position of the RNA (Thomas White and Davis, 1976) (12). Why some RNA loops were stable in linear DNA fragments remained a puzzle until a preprint arrived from my postdoctoral mentor, Norman Davidson, reporting that in the denaturant solvent formamide, used to suppress secondary structure of single-strand DNA in preparations for electron microscopy, RNA/DNA base pairs were thermodynamically more stable than DNA/DNA base pairs, particularly for GC base pairs (Casey and Davidson, 1977) (13). Since the Adenovirus 2 genome was known to be about 60% GC, formation of viral RNA/DNA hybrids would be expected to be strongly favored and relatively straightforward to characterize by electron microscopy.

THE DISCOVERY OF RNA SPLICING AT MIT.

Susan Berget joined my laboratory at MIT during the fall of 1974 and in collaboration with Jane Flint began to map specific viral mRNAs on the viral genome using hybridization (14). The next step was to turn to the higher resolution methods of electron microscopy to define the specific location of the abundant late mRNAs. The structures of the cytoplasmic mRNAs were to be compared with the structures of nuclear RNAs as a surrogate model for the hnRNA/mRNA paradox but as events unfolded, this step was anticlimactic. Groups at CSHL had already mapped several viral mRNAs to different regions of the genome (Lewis et al., 1975) (15) and Gelinas and Roberts had reported what I considered to be an interesting, puzzling, and non-conclusive finding: that many late-viral RNAs had a common 5' terminal oligonucleotide (Gelinas and Roberts 1977) (16). Sue purified the most abundant late-viral RNA, that for the hexon protein, and began to map its position on the viral genome using RNA loop formation with specific restriction fragments. Throughout these experiments Claire Moore, who at that time was in charge of the electron microscopy facility under my supervision but later a postdoctoral fellow in my lab, collaborated with Sue.

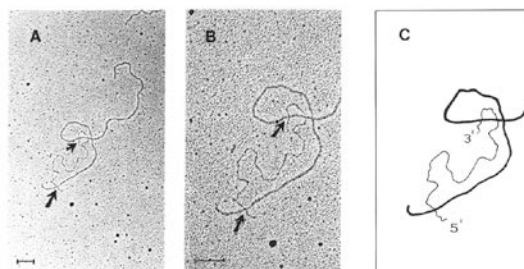


Figure 3. Electron micrograph of RNA/DNA hybrid of hexon mRNA and fragment of adenovirus genome. The 5' to 3' orientation of the mRNA is indicated. The displaced single strand viral DNA is shown as the thin light line.

During the series of R-looping experiments, Sue made the puzzling observation that at the 5' end of the mRNA, a tail of 150–200 nucleotides of RNA, did not anneal to the viral DNA sequences, **Figure 3** (17). We were surprised by this but were concerned that it was generated by branch migration where the RNA tail would be displaced by more stable DNA/DNA base pairs. This possibility was eliminated by forming the hexon mRNA/DNA hybrid with a single-strand fragment using conditions defined by the Casey and Davidson preprint. Under these conditions, there was no possible competing DNA strand. Remarkably, the 5' single strand tail of hexon mRNA persisted, indicating that it was not complementary to the adjacent viral DNA.

At this stage, we discussed these results with David Baltimore and shortly thereafter Sue presented the experimental results to weekly multigroup meeting, which included the Baltimore, Weinberg, Housman, and Hopkins labs (see photo of the 5th floor group meeting, 1974 and 1975, **Figures 4 and 5**, Courtesy of Robert Weinberg). There were four options for the source of the tail at the 5' end of the hexon mRNA sequences: sequences upstream on the r-strand of the viral genome, sequences from some other region of the viral genome, sequences joined from the cellular genome, or sequences generated by an unknown mechanism such as non-templated synthesis. The most straightforward experiment was to test whether the sequences were complementary to the upstream region of the virus. Sue hybridized the purified hexon mRNA to the EcoR1 fragment of Adenovirus 2 that contained all of the sequences upstream from the hexon mRNA and we observed the sample with electron microscopy, **Figure 6** (17).

As recorded in an article by Arnold Berk, a contemporary fellow in my lab and inventor of the S1 nuclease method of mapping RNA, we observed a remarkable and surprising structure of three reproducible loops in the RNA/DNA hybrids (Berk AJ, *PNAS* 2016) (18). This indicated that sequences in the 5' tail of the hexon mRNA were transcribed from sequences located at 16.8, 19.8, and 26.9 percentage units from the left end



Figure 4. Picture of members of multigroup (Baltimore, Weinberg, Housman, and Hopkins labs) meeting held weekly in the Center for Cancer Research at MIT. David Baltimore and Robert Weinberg are at lower left. Phillip Sharp is at upper right. From the top left- third person- Jane Flint and then Sue Berget. Behind Sue is Dan Donaghue and beside him is Claire Moore. There are many other important scientists in this picture.



Figure 5. Same as figure 4, 5th floor group meeting in 1975. Again, David Baltimore and Robert Weinberg are at lower left. Phillip Sharp, Nancy Hopkins, and David Housman are not present. Sue Berget is at the immediate right of Robert Weinberg. Claire Moore, second row, middle. Jane Flint, third row behind Claire's left shoulder, Dan Donahue immediately behind Jane. Tim Harrison in back row, fifth from left. Many other prominent scientists are featured in this

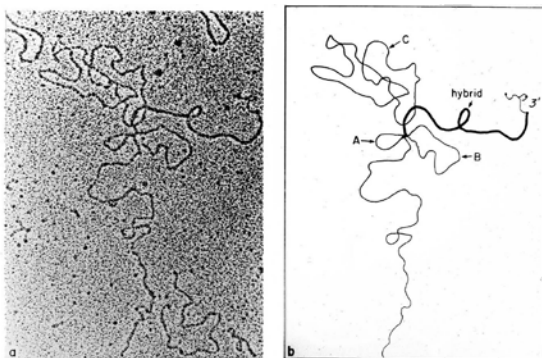


Figure 6. Electron micrograph of hybrid of the hexon mRNA and a single strand adenovirus fragment from 0-58. The three loops indicate three segments of RNA (exons) are spliced on to the body of the hexon mRNA. The loops are single strand DNA and their lengths determine positions of the three leader segments. See reference (17) for details.

of the genome, **Figure 7** (Courtesy of Phillip Sharp). We had observed the same length short 5' tails on other late viral mRNAs suggesting this was a common tail on many late adenovirus mRNAs. This conclusion was consistent with the proposal of Gelinas and Roberts that many different late mRNAs had a common 5' capped terminal oligonucleotide. We made the proposal that all late viral mRNAs from the r-strand were processed from a nuclear precursor that initiated at 16.8 units, was polyadenylated at different sites, and had three segments joined to the body of the mRNA before transport to the cytoplasm, **Figure 8** (19).

THE TERM "RNA SPLICING."

In writing the paper that David Baltimore communicated to the Proceedings of the National

Academy of Sciences, we wanted to select a name for the process of joining the short segments onto the 5' end of the body of the viral mRNAs. We introduced the term as the first word in the title "Spliced segments at the 5' terminus of adenovirus2 late mRNAs" (17). We called the process "RNA splicing" in all presentations. My first seminar outside of MIT about RNA splicing was presented at Sherbrooke, Canada where I had previously accepted an invitation to speak during the spring. We asked David Baltimore to communicate the paper early in 1977, because of a concern that reviewers and editors would find the results difficult to accept. However, because the *Proceedings* moves at its own place, the paper did not appear until August of 1977. The paper was dedicated to Jerome Vinograd, the discoverer of superhelical DNA and collaborator at Caltech, who passed away shortly before this.

1977 COLD SPRING HARBOR SYMPOSIUM MEETING.

The 1977 symposium at CSHL was a remarkable event. The symposium was organized to explore progress in the structure and biology of chromatin highlighting the discovery of the nucleosome. [As a side note, this was the symposium where Francis Crick dramatically called a presenter an "amateur" during his presentation. Later in the meeting Charles Weissmann midway through his presentation theatrically removed his jacket and turned to point to a slide. There across the back of his shirt was the word "Amateur". The audience laughed hysterically.] The new science of RNA splicing was presented on Thursday evening. Wally Gilbert chaired the session which opened with a presentation on transcription initiation sites on the adenovirus genome from the Darnell Lab, then my presentation was followed by that of Roberts from CSHL. As far as I recall, there were no presentations from

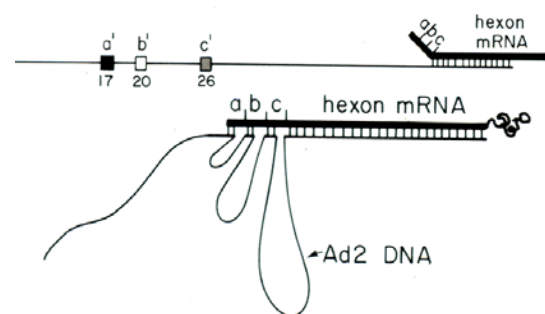


Figure 7. Diagram schematically interpreting the RNA/DNA hybrid shown in Figure 6.

other labs about split genes or RNA splicing at the meeting. However, there was a lot of conversation between sessions about observations such as restriction endonuclease sites in genome DNA for a gene that was not present in the cellular DNA (c-DNA) and loops observed by electron microscopy following hybridization in RNA/DNA preparations with cloned c-DNA segments. A flurry of papers appeared during the summer showing that split genes were common in mammalian genomes. Since submissions to the Symposium book were accepted for months after the meeting, many descriptions of split genes are present in the published volumes.

Three articles related to RNA Splicing from this early period stand out in my memory, 1) Joe Sambrook of CSHL summarized the results presented about adenovirus at the symposium entitled “Adenovirus amazes at Cold Spring Harbor” in a *Nature* “News and Views” July, 1977 (20). Although he describes a number of possible mechanisms to explain amazing electron microscopy structures and ends with the RNA splicing model, the word “splicing” is not used. 2) Walter Gilbert wrote another News and Views for *Nature* in February 1978 entitled “Why genes in pieces” (21). The “why” question is still not answered well, but this article introduced the terminology “intron” and “exon.” 3) The “Dean” of molecular biology, Francis Crick, summarized the subject in 1979 in a *Science* article entitled “Split Genes and RNA Splicing” (22). I use the term “Dean,” because others and myself viewed Francis as one of the deepest thinkers in molecular biology. By this time, “RNA splicing,” “intron/exon” and even the presence of common sequences at the splice sites were accepted. After this early period, many excellent reviews appeared recording the progress of the field.

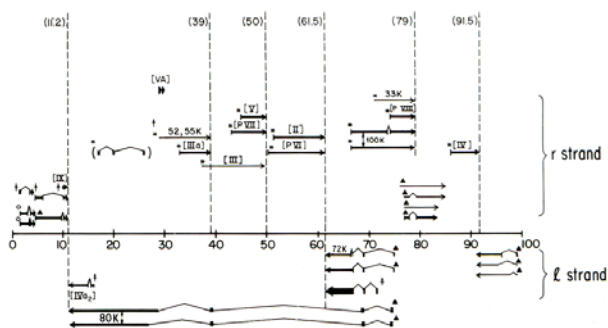


Figure 8. Diagram of the extensive splicing and alternative splicing of adenovirus 2 mRNAs. mRNAs from all four early regions, EI, EII, EIII, and EIV are alternatively spliced. Note the late mRNAs from the r strand are divided into five families defined by a common polyA site but have a common set of three leader RNA segments spliced at their 5' terminus. Each of these mRNAs are processed from a precursor RNA initiated at the major late promoter MLP at position 17. See reference (44) for details.

In parallel with the electron microscopy splicing results studying adenovirus, Arnold Berk, a fellow, combined selective hybridization conditions, the specificity of S1 nuclease for single-strand DNA, and high-resolution gel electrophoresis to map the position of introns in viral and cellular genes. In important papers in 1977 and 1978, he characterized the intron structures of the early regions of Adenovirus (23) and SV40 (24). During the same period, Sue, Arnie, and Tim Harrison collaborated to characterize the long nuclear RNA in late adenovirus infected cells, identifying the anticipated nuclear splicing intermediates generating mature cytoplasmic mRNAs (25). In collaboration with the Baltimore lab, Dan Donoghue, my first graduate student, characterized RNAs related to Moloney retroviruses (26).

BIOCHEMISTRY OF RNA SPLICING.

The years from 1977 to 1983 were difficult for a scientist with a laboratory focused virology, but exciting for someone interested in gene regulation. The revolution of recombinant DNA technology exploded during this period, allowing characterization of the intronic structure of cellular genes and their expression. Large and small groups around the world turned their efforts to this subject, with startling results about conserved sequences at the 5' and 3' splice sites, branch sites, variation in alternative splicing, and intron structure. Although many similar results were also reported for viral systems (Figure 8), particularly related to alternative splicing, the broader community was enthralled with cellular genes, their mutations, and regulation. The Crick 1979 article mentioned above summarizes some of the early results from cellular genes. In 1980, Joan Steitz's laboratory's proposal that U1snRNP recognized the 5' splice site during splicing dominated the thinking in the field (Lerner et al., *Nature* 283, 220–4, 1980) (27). During this period, we collaborated with Jim Manley, a fellow with Malcolm Gefter at MIT (Manley, Sharp, and Gefter, *PNAS* 1979, (28)), in biochemical studies of initiation of transcription at the major late promoter of Adenovirus. These transcription oriented experiments led to decades of work in my lab on the biochemistry of transcription that I will not attempt to summarize here, introduced the lab to mammalian cell biochemistry, and set the stage for our future studies of the biochemistry of RNA splicing.

Rick Padgett and Steve Hardy made a breakthrough when they detected, in a biochemical reaction, the splicing of RNA transcribed from the adenovirus major late promoter in 1983 (29). The system consisted of coupled transcription and splicing. A fragment of viral DNA was added to a mammalian transcription extract, developed during the biochemical transcription experiments with Manley and Gefter, then the newly labeled RNA was analyzed by RNAase digestion after hybridization to a cDNA fragment. The assay was very sensitive and detected a small level of spliced RNA. These results opened a long-sought opportunity to focus on biochemical characterization of RNA splicing. One of the first observations in this coupled system was that recognition of the cap structure was essential for splicing of the RNA (Konarska and Padgett, 1984) (30). In 1983 and 1984, Hernandez and Keller (31), and Krainer, Maniatis, Ruskin and Green (32), respectively, reported biochemical systems where externally added labeled RNA was spliced, and we quickly confirmed these results with our own preparation.

LARIAT STRUCTURE OF INTERMEDIATE AND EXCISED INTRON RNAs.

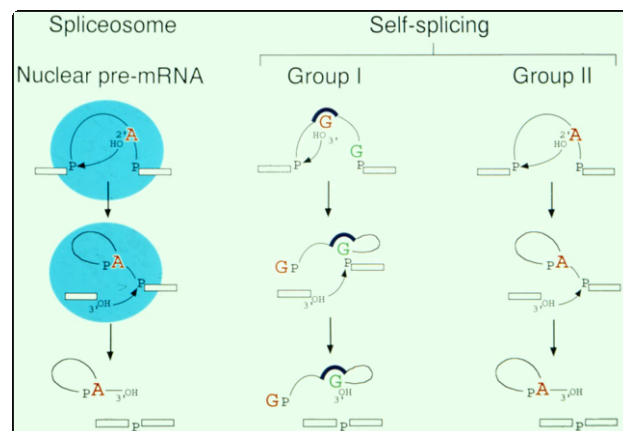


Figure 9. Schematic diagram of the lariat intermediate and excised intron of the spliceosome process shown in parallel with the group I and group II self-splicing processes. The blue circle represents the spliceosome that contain catalytic small RNAs while the two self-splicing processes depend on catalytic RNA structures formed by intron sequences. Note similar lariat structures formed during group II self-splicing. This similarity to the spliceosome chemistry was an early indication of the RNA catalytic core of the spliceosome, see reference (39) for further discussion.

Several major findings emerged shortly after the development of active biochemical systems. Paula Grabowski and Richard Padgett in the lab characterized a time course of the splicing reaction using a series of electrophoretic gels of different porosity to resolve labeled RNAs (33). Two RNAs, one intermediate and one product, *dramatically* shifted in mobility relative to linear markers when run on the different porosity gels indicating a non-linear structure. In our case, the circular nature of these RNAs was shown by nicking with endonuclease. We also detected labeled-oligonucleotides from these two RNAs reminiscent of a branched nucleotide following the first description of a branched nucleotide by Wallace and Edmonds in 1983 (34), **Figure 9** (diagram courtesy of Phillip Sharp). These were the intron intermediate and excise intron with the now well-known lariat structure. Definitive characterization of the branched adenosine was reported by Konarska, Grabowski, and Padgett (35). Related results were obtained in parallel by Ruskin, Krainer, Maniatis and Green (1984) (32) who analyzed the lariat structure by reverse transcription.

SPLICEOSOME COMPLEXES.

The lariat intermediate structure with branch formation concordant with cleavage at the 5' splice site meant that the two RNAs, the 5' exon, and the lariat intermediate, were

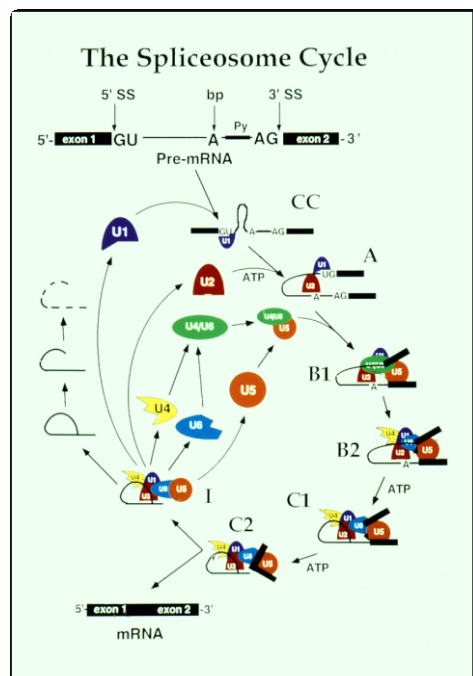


Figure 11. Schematic representation of the involvements of small ribonucleoprotein (snRNP) particles in pre-mRNA splicing. The schematic from review in 1993 (40). The cycle shows the specific snRNPs in the various spliceosome complexes resolved from mammalian cell extracts. The cycle is drawn similarly to that of Stephanie Ruby and John Abelson (1991), where results from spliceosome complexes identified in yeast extracts were summarized in (44).

probably retained in a complex. Paula Grabowski and Sharon Seiler identified in 1985 the complex as sedimenting at about 60S, **Figure 10** (36). In the last sentence of the paper we termed it the "spliceosome". Brody and Abelson at Caltech, using a yeast biochemical splicing reaction, reported co-sedimentation of the intermediate as a 40S "spliceosome" in a *Science* paper in May of 1985 (37). We submitted the above paper a week after their paper appeared, and the paper

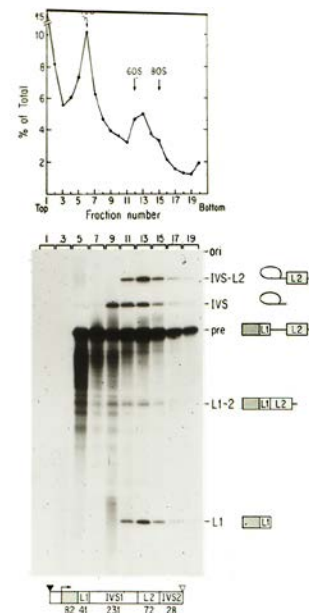


Figure 10. Sedimentation profile identifying the spliceosome at 60S. An *in vitro* splicing reaction was first fractionated by sedimentation and RNA from each fraction was resolved by gel electrophoresis. Note two RNAs defining the intermediate in splicing, 5' exon and lariat intermediate co-sedimented in fraction 13, indicating a spliceosome complex. See reference (36) for details.

was published June 11. We had been using the spliceosome terminology in the lab for months.

Many other findings emerged from the study of the biochemistry of splicing but here I will highlight only results that have been expanded in contemporary research on the atomic structure of spliceosome complexes. Magda Konarska introduced in the lab resolution of spliceosome-related complexes by gel electrophoresis in 1985 (Konarska and Sharp 1985) (38). Magda first resolved splicing complexes by sedimentation and then in a second dimension resolved fractions by non-denaturing gel electrophoresis. The content of the gel was transferred to paper and analyzed by hybridization with probes specific to individual snRNAs. In this matter, specific snRNAs could be assigned to resolved complexes. When combined with substrate RNAs with mutations in splice and branch sites, these experiments resolved 1) complex A of U2snRNA and substrate, 2) complex B with U2, U4, U5, and U6 snRNPs, 3) complex C with U2, U6 and U5 snRNPs, 4) the U4/U6 snRNP complex and 5) the U4/U6-U5 snRNP complex (38, 39,40). Michael Rosbash's lab also began to resolve complexes using biochemical reactions prepared from yeast in parallel with those in my lab (Pikielny et al., 1986, Nature) (41). We realized that the two labs were working in parallel on gel electrophoresis of spliceosome complexes when we met one Saturday at soccer field watching our daughters play and began discussing RNA splicing.

Tom Cech's early experiments characterizing group 1 self-splicing introns were consistent with the two reactions excising the intron occurring at a single catalytic site. His lab showed this was the case chemically by following the chirality of the two reactions using thiophosphates. Melissa Moore developed the splint ligation process for joining two long RNAs (Moore and Sharp 1992, Science) (42) and then using this technology showed that the active chirality of the 5' Splice site transesterification reaction forming the branch converted a thiophosphate from Rp to Sp while the second step at the 3' splice site also converted the Sp chirality to an Rp. This is consistent with the now well-established single two metal ion catalytic center recently visualized by cryo-electron microscopy.

The above text focuses on high points in the journey from the discovery of RNA splicing to characterization of the snRNPs and precursor RNA components of spliceosome complexes. There are several other important advances in my laboratory and that of others in topics related to RNA splicing not mentioned here. These range from regulation of alternative splicing, trans-splicing, RNA binding proteins, U11-U12 spliceosome, and



Figure 12A

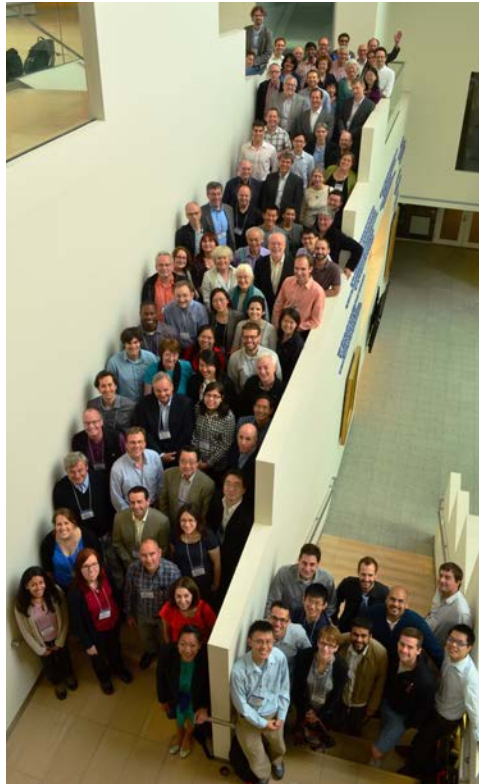


Figure 12B

more recently detained introns (43). The following list of reviews covers some of these topics.

Finally, the students, fellows, and visitors in my lab at MIT have been wonderful colleagues. I cannot adequately express my gratitude for their creative work and friendship. This is doubly the case for my longtime assistant, Margarita Siafaca. Figure 12 A&B are photographs of these colleagues at two reunions at MIT in 1994 and 2014. Note the presence of Norman Davidson and David Baltimore at the lower left of Figure 12A. They kindly chaired sessions of the reunion in 1994.

FURTHER COMMENTS THE DISCOVERY OF SPLIT GENES AND RNA SPLICING.

There have been speculations by others about the exchange of information between the group at CSHL and MIT during the spring of 1977. I will only address events at MIT because I do not, and did not, know about events at CSH during this period.

Sue Berget began experiments to purify the adenovirus hexon mRNA and map it on the viral genome in September 1976. These experiments require preparations, exploration of many conditions for hybridization, and production of the grids to be examined in the EM and then interpretation of the images in the EM. Berget and Moore participated in these steps over the 9 months of experimentation for the PNAS paper and I participated in examining the grids in the EM as well.

By the end of 1976, we had established that a short sequence at the 5' end of the hexon mRNA would not hybridize to the genome sequence adjacent to the body of the hexon mRNA. In January 1977, Sue presented this research at a weekly seminar group that included the groups of David Baltimore, Robert Weinberg, Nancy Hopkins and David Housman. At the same time, Sue and I met with David Baltimore about communicating a paper to PNAS about the results. He agreed to be editor of the paper for the Proceedings and needed to have the paper reviewed by an independent and knowledgeable scientist. He did not reveal the name of this reviewer, I never asked.

The original manuscript submitted for review in February, as I recall, contained data showing that an RNA segment several 100 nucleotides in length at the 5' end of the hexon mRNA would not hybridize to adjacent viral DNA sequences and thus must have originated elsewhere. There were three options for the origin of these sequences: 1) They were random in sequences being created by a non-specific polymerase, 2) They were transcribed from cellular sequences and added to the viral mRNA, or 3) They were transcribed from viral sequences and added to the body of the mRNA. At the meeting, David and I discussed these three possibilities and both agree that the third option was the one to be tested first. In fact, at the time it was the only option we had the tools to test because cDNA cloning was just being developed. The sentence in the Acknowledgement of the paper recognized this conversation with David in early February as we were finalizing the first manuscript for review.

While this draft of the manuscript was under review, the experiments were completed where we examined if the 5' end sequences would anneal to the sequences upstream of the hexon mRNA. We observed the three loops defining the three leader sequences that were "spliced" to the body of the mRNA. The expletive statement Arnold Berk described in the recent PNAS article about the discovery was in reaction to seeing this three-loop structure. We measured the lengths of these loops on multiple examples to determine where they mapped on the genome.

When the reviews of the first manuscript was given to us by David, we had the EM results that mapped the three leaders and added this to the final manuscript that was returned to David for submission. This paper was further reviewed and then submitted to PNAS for publication. The date for communication of the final manuscript is May 9th, 1977. This is after the second review. I traveled to Sherbrooke, Canada on April 15th for a seminar, and presented the complete results that appeared in the published paper. I had been invited during the fall of 1976 by Professor Joe Weber who studied Adenovirus. Just for the record, I note the “amazing” paper from CSH was received by Cell on June 9th and revised on July 5th, 1977.

As far as I am aware, there was no communication to anyone in my group of the results from experiments at CSHL concerning the leader structures of the viral mRNA before our final manuscript was accepted by PNAS. In fact, I did not see their results until the CSH Symposium in June 1977. We clearly were aware of all of the CSH papers published before this date and considered them “competitive but friendly” colleagues in the field.

Reviews of the field of RNA splicing:

Sharp PA. [Speculations on RNA splicing](#). Cell. 1981 Mar;23(3):643-6. PubMed PMID: 7226224.

[The Adenoviruses](#) (H. S. Ginsberg, ed.), Chapter 5, pp. 173-204, New York: Plenum Press Corp. (1984).

Sharp PA. [On the origin of RNA splicing and introns](#). Cell. 1985 Sep;42(2):397-400. Review. PubMed PMID: 2411416.

Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA. [Splicing of messenger RNA precursors](#). Annu Rev Biochem. 1986;55:1119-50. Review. PubMed PMID: 2943217.

Sharp PA. [Splicing of messenger RNA precursors](#). Science. 1987 Feb 13;235(4790):766-71. Erratum in: Science 1987 Aug 28;237(4818):964. PubMed PMID: 3544217.

Sharp PA. [Trans splicing: variation on a familiar theme?](#) Cell. 1987 Jul 17;50(2):147-8. PubMed PMID: 3594568.

Sharp PA, Eisenberg D. [The evolution of catalytic function](#). Science. 1987 Nov 6;238(4828):729-30, 807. PubMed PMID: 2445035.

*Sharp PA, Konarska MM, Grabowski PJ, Lamond AI, Marciniak R, Seiler SR. [Splicing of messenger RNA precursors](#). Cold Spring Harb Symp Quant Biol. 1987;52:277-85. PubMed PMID: 3331341.

Sharp PA. ["Five easy pieces"](#). Science. 1991 Nov 1;254(5032):663. PubMed PMID: 1948046.

Moore, M. J., Query, C. C., and Sharp, P. A. Splicing of precursors to messenger RNAs by the spliceosome. In [The RNA World](#), R. Gesteland and J. Atkins, eds., Cold Spring Harbor Laboratory Press, pp. 303-357 (1993).

Sharp PA. [Split genes and RNA splicing](#). Cell. 1994 Jun 17;77(6):805-15. Review. PubMed PMID: 7516265.

Sharp PA, Burge CB. [Classification of introns: U2-type or U12-type](#). Cell. 1997 Dec 26;91(7):875-9. Review. PubMed PMID: 9428511.

Burge, C., Tuschl, T. and Sharp, P.A. Splicing of precursors to mRNAs by the spliceosomes. In [RNA World II](#), R. Gesteland, T. Cech, and J. Atkins, eds., Cold Spring Harbor Laboratory Press, NY. pp. 525-560 (1999).

Jangi M, Sharp PA. [Building robust transcriptomes with master splicing factors](#). Cell. 2014 Oct 23;159(3):487-98. doi: 10.1016/j.cell.2014.09.054. Review. PubMed PMID: 25417102; PubMed Central PMCID: PMC4243530

References Cited

1. Cairns J, Crick F, Nirenberg M, Speyer J, Watson J, organizer and advisors. The Genetic Code. Cold Spring Harbor Symposia on Quantitative Biology; 1966 2-9 June, Vol 31; Cold Spring Harbor. New York: Cold Spring Harbor Laboratory Press; 1966 xxii-762
2. Davis RW, Davidson N. Electron-microscopic visualization of deletion mutations. Proceedings of the National Academy of Sciences of the United States of America. 1968;60(1):243-50. Epub 1968/05/01. PubMed PMID: 5241526; PMCID: PMC539109.
3. Sharp PA, Hsu M, Davidson N. Note on the structure of prophage lambda. Journal of molecular biology. 1972;71(2):499-501. Epub 1972/11/14. PubMed PMID: 4635993.
4. Sambrook J, Sharp PA, and Keller W. Transcription of Simian virus 40. I. Separation of the strands of SV40 DNA and hybridization of the separated strands to RNA extracted from lytically infected and transformed cells. J Mol Biol. 1972; 70(1):57-71.
5. Sharp PA, Sugden B, Sambrook J. Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose--ethidium bromide electrophoresis. Biochemistry. 1973;12(16):3055-63. Epub 1973/07/31. PubMed PMID: 4354250.
6. Pettersson U, Mulder C, Deluis H, Sharp PA. Cleavage of adenovirus type 2 DNA into six unique fragments by endonuclease R-R1. Proceedings of the National Academy of Sciences of the United States of America. 1973;70(1):200-4. Epub 1973/01/01. PubMed PMID: 4567334; PMCID: PMC433215.
7. Sharp PA, Pettersson U, Sambrook J. Viral DNA in transformed cells. I. A study of the sequences of adenovirus 2 DNA in a line of transformed rat cells using specific fragments of the viral genome. Journal of molecular biology. 1974;86(4):709-26. Epub 1974/07/15. PubMed PMID: 4473553.
8. Sambrook J, Williams J, Sharp PA, Grodzicker T. Physical mapping of temperature-sensitive mutations of adenoviruses. J. Mol Biol. 1975; 97(1) 369-390.
9. Flint SJ, Sharp PA. Adenovirus transcription. V. Quantitation of viral RNA sequences in adenovirus 2-infected and transformed cells. J Mol Biol. 1976;106(3):749-74. Epub 1976/09/25. PubMed PMID: 185395.
10. Wall R, Philipson L, Darnell JE. Processing of adenovirus specific nuclear RNA during virus replication. Virology. 1972;50(1):27-34. Epub 1972/10/01. PubMed PMID: 5081848.
11. Darnell J. RNA: Life's Indispensable Molecule. 1st ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2011. 416p.
12. Thomas M, White RL, Davis RW. Hybridization of RNA to double-stranded DNA: formation of R-loops. Proceedings of the National Academy of Sciences of the United

- States of America. 1976;73(7):2294-8. Epub 1976/07/01. PubMed PMID: 781674; PMCID: PMC430535.
13. Casey J, Davidson N. Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nucleic Acids Res.* 1977;4(5):1539-52. Epub 1977/01/01. PubMed PMID: 19730; PMCID: PMC343772.
 14. Berget SM, Flint SJ, and Sharp PA. Characterization of viral, single-stranded DNA sequences in adenovirus infected cells. New York: Academic Press; 1976. 81-96 (Baltimore D, Huang AS, Fox CF, editors. *Animal Virology*).
 15. Lewis JB, Anderson CW, Atkins JF, Gesteland RF. The origin and destiny of adenovirus proteins. *Cold Spring Harbor symposia on quantitative biology.* 1975;39 Pt 1:581-90. Epub 1975/01/01. PubMed PMID: 1057484.
 16. Gelinas RE, Roberts RJ. One predominant 5'-undecanucleotide in adenovirus 2 late messenger RNAs. *Cell.* 1977;11(3):533-44. Epub 1977/07/01. PubMed PMID: 884734.
 17. Berget SM, Moore C, and Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA.* 1977; 74, 3171-3175.
 18. Berk AJ. Discovery of RNA splicing and genes in pieces. *Proceedings of the National Academy of Sciences of the United States of America.* 2016;113(4):801-5. Epub 2016/01/21. doi: 10.1073/pnas.1525084113. PubMed PMID: 26787897; PMCID: PMC4743779.
 19. Berget SM, Sharp PA. A spliced sequence at the 5'-terminus of adenovirus late mRNA. *Brookhaven symposia in biology.* 1977(29):332-44. Epub 1977/05/12. PubMed PMID: 754866.
 20. Sambrook J. Adenovirus amazes at Cold Spring Harbor. *Nature.* 1977;268(5616):101-4. Epub 1977/07/14. PubMed PMID: 593301.
 21. Gilbert W. Why genes in pieces? *Nature.* 1978;271(5645):501. Epub 1978/02/09. PubMed PMID: 622185.
 22. Crick F. Split genes and RNA splicing. *Science (New York, NY).* 1979;204(4390):264-71. Epub 1979/04/20. PubMed PMID: 373120.
 23. Berk AJ, Sharp PA. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell.* 1977;12(3):721-32. Epub 1977/11/01. PubMed PMID: 922889.
 24. Berk AJ and Sharp PA. Spliced early mRNAs of simian virus 40. *Proc. Nat. Acad. Sci. USA.* 1978;75:1274-1278.
 25. Berget SM, Berk AJ, Harrison T, Sharp PA. Spliced segments at the 5' termini of adenovirus-2 late mRNA: a role for heterogeneous nuclear RNA in mammalian cells. *Cold Spring Harbor symposia on quantitative biology.* 1978;42 Pt 1:523-9. Epub 1978/01/01. PubMed PMID: 277361.
 26. Donoghue DJ, Rothenberg E, Hopkins N, Baltimore D, and Sharp PA. Heteroduplex analysis of the nonhomology region between Moloney MuLV and the dual host range derivative HIX virus. *Cell.* 1978; 14: 959-970.
 27. Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. Are snRNPs involved in splicing? *Nature.* 1980;283(5743):220-4. Epub 1980/01/10. PubMed PMID: 7350545.

28. Manley JL, Fire A, Cano A, Sharp PA, Gefter ML. DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. *Proceedings of the National Academy of Sciences of the United States of America*. 1980;77(7):3855-9. Epub 1980/07/01. PubMed PMID: 6933441; PMCID: PMC349725.
29. Padgett RA, Hardy SF, Sharp PA. Splicing of adenovirus RNA in a cell-free transcription system. *Proceedings of the National Academy of Sciences of the United States of America*. 1983;80(17):5230-4. Epub 1983/09/01. PubMed PMID: 6577417; PMCID: PMC384226.
30. Konarska MM, Padgett RA, Sharp PA. Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell*. 1984;38(3):731-6. Epub 1984/10/01. PubMed PMID: 6567484.
31. Hernandez N, Keller W. Splicing of in vitro synthesized messenger RNA precursors in HeLa cell extracts. *Cell*. 1983;35(1):89-99. Epub 1983/11/01. PubMed PMID: 6194902.
32. Ruskin B, Krainer AR, Maniatis T, Green MR. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell*. 1984;38(1):317-31. Epub 1984/08/01. PubMed PMID: 6088074.
33. Grabowski PJ, Padgett RA, Sharp PA. Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. *Cell*. 1984;37(2):415-27. Epub 1984/06/01. PubMed PMID: 6722880.
34. Padgett RA, Konarska MM, Grabowski PJ, Hardy SF, Sharp PA. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science (New York, NY)*. 1984;225(4665):898-903. Epub 1984/08/31. PubMed PMID: 6206566
35. Konarska MM, Grabowski PJ, Padgett RA, Sharp PA. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature*. 1985;313(6003):552-7. Epub 1985/02/14. PubMed PMID: 2578627.
36. Grabowski PJ, Seiler SR, Sharp PA. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell*. 1985;42(1):345-53. Epub 1985/08/01. PubMed PMID: 3160482.
37. Brody E, Abelson J. The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science (New York, NY)*. 1985;228(4702):963-7. Epub 1985/05/24. PubMed PMID: 3890181.
38. Konarska MM, Sharp PA. Association of U2, U4, U5, and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA. *Proceedings of the National Academy of Sciences of the United States of America*. 1988;85(15):5459-62. Epub 1988/08/01. PubMed PMID: 2969592; PMCID: PMC281776.
39. Sharp PA, Konarska MM, Grabowski PJ, Lamond AI, Marciniak R, and Selier S. Splicing of messenger RNA precursors. *Cold Spring Harbor Symp. Quant. Biol.* 1987: 71(52).
40. Moore MJ, Query CC, and Sharp Pa. Splicing of precursors to messenger RNAs by the spliceosome. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1993. P 303-357. (Gesteland R and Atkins J, editors. *The RNA World*).

41. Pikielny CW, Rymond BC, Rosbash M. Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes. *Nature*. 1986;324(6095):341-5. Epub 1986/11/03. doi: 10.1038/324341a0. PubMed PMID: 3537805.
42. Moore MJ, Sharp PA. Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science (New York, NY)*. 1992;256(5059):992-7. Epub 1992/05/15. PubMed PMID: 1589782.
43. Boutz, P.L., Bhutkar, A., and Sharp, P.A. Detained introns are a novel, widespread class of post- transcriptionally spliced introns *Genes Dev*. 2015 Jan 1;**29**(1):63-80.doi: 10.1101/gad.247361.114. PubMed PMID: 25561496; PubMed Central PMCID: PMC4281565.
44. Ruby SW, Abelson J. Pre-mRNA splicing in yeast. *Trends Genet*. 1991 Mar; 7(3): 79-85. Review. Pubmed PMID: 2031287