SPLIT GENES AND RNA SPLICING

Nobel Lecture, December 8, 1993

by

PHILLIP A. SHARP

Department of Biology and the Center for Cancer Research, Massachusetts
Institute of Technology, Cambridge, MA 02139 – 4307, USA

INTRODUCTION

By the late 70s the physical structure of a gene was firmly established from work in bacteria. The sequences of the gene, the RNA and the protein were colinearly organized and expressed. Since the science of genetics suggested that genes in eukaryotic organisms behaved similarly to those of prokaryotic organisms, it was naturally assumed that this bacterial gene structure was universal. It followed that if the gene structure was the same, then the mechanisms of regulation were probably very similar, and thus what was true of a bacterium would be true of an elephant.

However, many descriptive biochemical aspects of the genetic material and its expression in cells with nuclei suggested that the simple molecular biology of gene expression in bacteria might not be universal. First, obviously, RNAs transcribed from nuclear genes are physically separated from the translational machinery in the cytoplasm. Thus, the nuclear compartment could be the site of selective RNA processing and transport. Second, the DNA content of eukaryotic germ cells varied significantly between organisms without an apparent variation in the number of genes. Some organisms appeared to have ten times as much DNA as was required to encode all of the proteins. Third, the previously described phenomenon of heterogeneous nuclear RNA (hnRNA) suggested that long RNAs were transcribed from diverse nuclear sequences (1). These hnRNAs had a short half-life relative to cytoplasmic mRNAs and thus could potentially be precursors to mRNAs. Furthermore, both the long hnRNA and the shorter mRNA appeared to have modified 5′ and 3′ termini in common, a ^{7}mGpppX cap (2, 3, 4) and polyadenylation tracts (5, 6, 7), respectively. The meaning of these observations concerning hnRNAs remained controversial as it was not possible to establish a precursor-product relationship between the nuclear RNA population and the cytoplasmic mRNA.

Whether or not the structure of genes in eukaryotic cells was the same as that in bacteria was not really questioned at that time. The important issue was to establish the exact biochemical pathway between a gene in the nucleus and its mRNA in the cytoplasm. This hypothetical pathway was pictured as beginning with initiation of transcription by RNA polymerase II(B) which would proceed'through completion of transcription beyond the terminus of the gene. The nuclear precursor RNA transcribed from the gene was potentially processed and the mRNA selectively transported to the cytoplasm. Regulation of gene expression, the basis of almost all interesting biology, including cancer, cellular responses to infection, and development, would primarily result from changes in the rates or efficiencies of the various steps in the pathway. Thus, understanding the pathway of eukaryotic gene expression promised new insights into the mechanisms of regulation and the biochemical events controlling both diverse biological and biomedical phenomena.

BACKGROUND

The late stage of adenovirus infection was chosen as the best system for studying the pathway of nuclear gene → nuclear precursor RNA → cytoplasmic mRNA → protein. We had previously established several restriction endonuclease cleavage maps of the viral genome (Fig. 1; 8) and had used fragments of the genome to map the positions of cytoplasmic mRNAs generated during the early and late stages of infection (9). We had also determined the abundance (in copies per cell) of both nuclear and cytoplasmic RNAs (10), and this suggested that RNAs from both compartments could be obtained in adequate amounts to compare their structures directly using the electron microscope and the then recently developed RNA·DNA hybridization methods. Furthermore, these earlier studies established that sets of viral RNA sequences were restricted to the nucleus, suggesting a selection in processing and/or transport of only certain RNA sequences to the cytoplasm (11). Finally, several studies had established that long nuclear

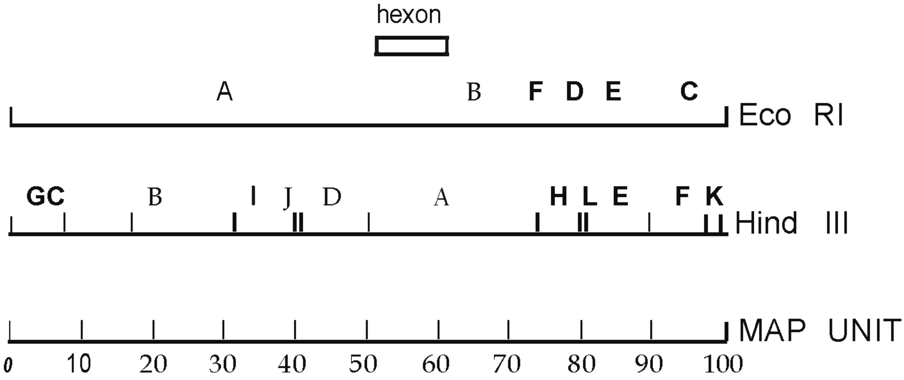

Fig. I: Map of cleavage sites for restriction *endonucleases Eco*RI *and HindIII.* The approximately 35,000 bp of adcnovirus 2 DNA was assigned as 100 map units. The positions of cleavage sites are denoted by vertical lines and fragments are lettered on the basis of length. The r and l viral strands are transcribed to the right and left, respectively. The sequences constituting the body of mRNA specifying the abundant hexon (II) protein are encompassed by the bar about the boundary of *Eco*RI fragments A and B.

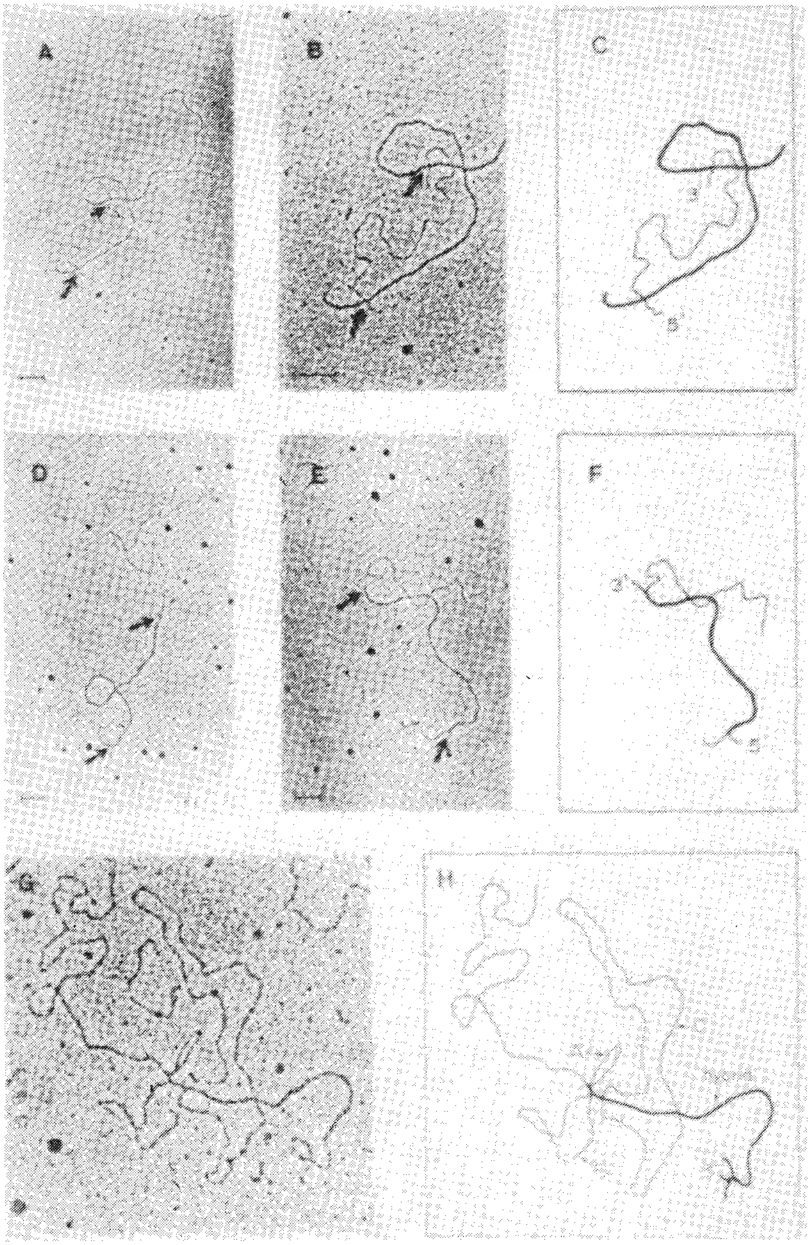

Fig. 2: Electron micrographs of hybrids of hexon mRNA andfragments of Ad2DNA (13). Examples of the latterR-toop hybrids observed after incubation of hexon mRNA and duplex *Hind*III A fragment DNA arc shown in A and B and is diagrammed schematically in C. Similarly, two examples of hybrids of hexon mRNA and the single-stranded *Hind*III A fragment are shown in D and E. A schematic. of the hybrid structure shown in E is given in F. The single-stranded RNA at the end of the hybrid region is represented by a wave-like line. In A, B, D, and E the positions of the KNA tails at the 5′ and 3′ ends of the hybrids are denoted by arrows. An example of a hybrid between single-stranded *Eco*RI A DNA and hexon RNA is shown in G and diagrammed in H. The hybrid region is indicated by a heavy line; loops A, B, and C (single-stranded unhybridized DNA) are joined by hybrid regions resulting from annealing of upstream DNA sequences to the 5′ tail of hexon mRNA. Bars on micrographs represent 0.1 μM.

RNA was transcribed from the adenovirus genome during the late stages of infection (12) and the stable cytoplasmic RNAs were shorter than the predominant nuclear RNAs. Thus, the production of RNAs during the late stage of adenovirus infection presented a paradigm for the heterogeneous nuclear RNA phenomenon associated with cellular genes.

DISCOVERY OF RNA SPLICING

Comparison of the structure of a cytoplasmic mRNA to that of a nuclear precursor RNA (13) required the purification of a specific homogeneous mRNA. The most abundant mRNA, which encoded the adenovirus hexon protein, was separated from other viral mRNAs by gel electrophoresis and used in electron microscope mapping studies (14). Ray White, as a fellow at Stanford, was the first to recognize that RNA·DNA hybrids were more stable than DNA·DNA duplexes in high concentrations of formamide (15). This phenomenon was characterized physically by Davidson's lab (16) and was the basis of a convenient R-looping technique whereby RNA forms a hybrid with a DNA strand, displacing the other strand of DNA into an easily observed loop. The adenovirus mRNA for hexon protein was mapped by the R-loop method to the *Hind*III A fragment of adenovirus 2 (13).

Inspection of the R-loops between the hexon-mRNA and the *Hind*III B DNA fragment revealed the presence of RNA tails at both the 5′ and 3′ ends of the hybrid (Fig. 2A, B and C). The single strand RNA tail at the 3′ end was expected, as the RNA was known to be polyadenylated post-transcriptionally. The single strand RNA tail at the 5′ end of the hybrid was not expected; however, this 5′ tail of RNA could have been displaced from the RNA·DNA hybrids by formation of duplex DNA by a process called branch migration. In fact, similar 5′ tails had been observed previously in R-loop mapping of adenovirus mRNA and had been ascribed to such branch migration (17, 18). Arguing against branch migration however was the finding that the lengths of the 5′ tails were relatively uniform, 170 nucleotides. To eliminate this potential of competing DNA sequences displacing RNA sequences, we used a denatured single-strand of the *Hind*III fragment (Fig. 2D, E, and F) to form a RNA·DNA hybrid. Surprisingly, the 5′ RNA tail still did not form a hybrid with the adjacent viral sequences, suggesting that these RNA sequences were derived from other DNA sequences.

Could the DNA sequences transcribed to form the 5′ tail sequence, the leader RNA sequence, be located upstream of the body of the hexon mRNA, perhaps as part of the long nuclear RNA? To test this possibility, a strand of the ***EcoRI*** A DNA fragment was hybridized with the hexon mRNA. This fragment contained all of the linear viral sequences which could have been transcribed by the polymerase before encountering the body of the hexon mRNA. Surprisingly, and wonderfully, the leader sequences hybridized to three short tracts of DNA sequences, creating three different size loops A, B and C of intervening single-strand DNA (Fig. 2G and H). The length of these loops mapped the positions of the leader sequences, **L1** , L2, L3, to approximately 16.9, 19.8, and 26.9 map units on the genome, respectively. The distances between the bases of the three loops permitted an estimate of the lengths of the two internal leaders to be 80 and 110 nucleotides, respectively. The 5′ proximal leader was estimated to be quite short, but longer than fifteen nucleotides.

RNA splicing was the mechanism we proposed for generation of the final hexon mRNA (13). It was known that the nucleus of virus-infected cells contained long RNAs transcribed from the viral sequences between the 5′ most leader **Ll** , located at *17* map units, and the body of the hexon mRNA, located between 51.7 and 61.3 map units (19, 20). Thus, the long nuclear RNA probably contained sequences for all three leaders, Ll , L2 and L3, as well as for the body of the mRNA. These sequences were conjectured to be joined by excision of the intervening sequences and ligation of the flanking RNAs, a process dubbed RNA splicing (Fig. 3). In fact, the nuclear RNAs were quite abundant and were easily visualized by electron microscopy of the RNA·DNA hybrids (21). Analysis of the structure of these RNA·DNA hybrids revealed the presence of potential intermediates, where only Ll had been spliced to L2 and where only Ll, L2 and L3 had been joined by splicing:

The RNA splicing hypothesis reconciled many paradoxes. Both the longer nuclear RNAs and the shorter cytoplasmic mRNAs could share 5′ cap termini, ^{7}mGpppX, and 3′ polyA tracts (22), because internal sequences were removed from the nuclear precursor. More specifically for adeno-

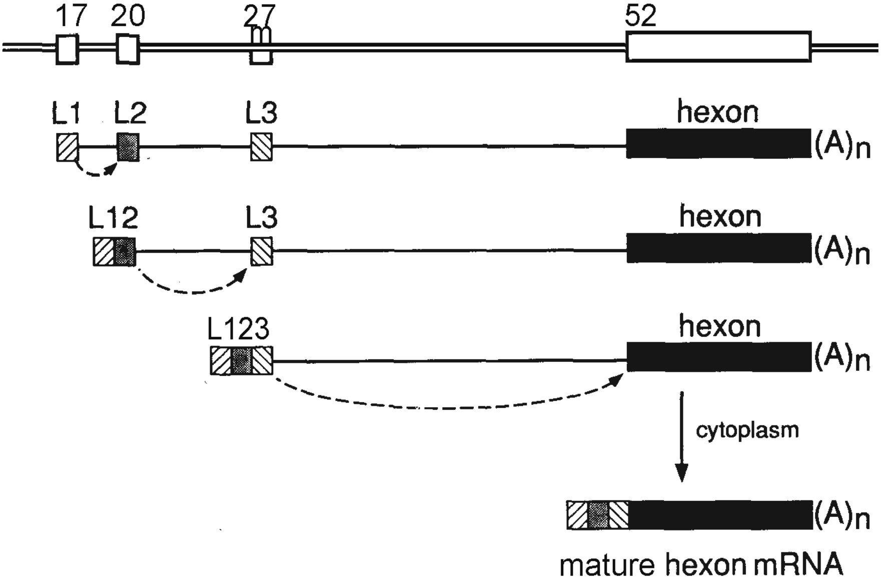

Fig. 3: Proposed RNA splicing mechanism for synthesis of mRNA for hexon protein. **A long nuclear precursor RNA is transcribed from 16.9 map units through the polyA addition site at the end of the body of hexon mRNA. The region of the Ad2 genome from which the precursor RNA is transcribed is shown at the top of the figure. The four RNA segments in the cytoplasmic mRNA are processed from this precursor by excision of intervening sequences** *(denoted by dashed arrows).*

virus, previous evidence had suggested that many different viral mRNAs appeared to share a common 5′ terminal sequence (23). This sequence, a long Tl oligonucleotide, contained the cap and an apparently unique sequence eleven residues in length. If many of the late viral mRNAs were processed by splicing from the same type of precursor RNA, then they could share a common sequence at their 5′ termini. More importantly, 'the RNA splicing hypothesis provided an explanation for the hnRNA phenomenology associated with cellular genes. Heterogeneous nuclear RNA transcribed from diverse cellular genes could be processed by RNA splicing into shorter cytoplasmic mRNAs. Thus, most cellular genes probably contained sequences which were removed by RNA splicing, i.e. they were split genes.

SPLIT GENES: INTERVENING SEQUENCES OR INTRONS IN CELLULAR GENES

Shortly after the discovery of RNA splicing and split genes in adenovirus, a number of cellular genes were also shown to have introns or intervening sequences. For example, the globin genes contained two intervening sequences (24, 25), the ovalbumin gene was split into eight sets of sequences (26), and the immunoglobulin genes contained both short and long introns (27). In fact, the average cellular gene contains approximately eight introns and the primary transcription unit is typically four times larger than the final mRNA. Shortly thereafter, it was recognized that there was a limited set of conserved sequences at each intron boundary (28). Interestingly, these consensus sequences were common of vertebrate, plant and yeast cells (29) suggesting the splicing process was evolutionarily general. Introns in the latter organisms are generally shorter and have more highly conserved sequences at their boundaries.

Phylogenetic comparison of the sequences of homologous genes from a variety of organisms revealed that intron sequences had drifted much more rapidly than exon sequences. This suggested that intron sequences are generally not functional, at least in the context of requiring long tracts of specific sequences. Furthermore, the length of introns in homologous genes varied significantly during evolution, suggesting little constraint. Finally, it was clear that specific introns could be lost during evolution. The mechanism responsible for the exact deletion of introns is probably related to gene conversion using a cDNA copy of the mRNA or a partially spliced intermediate RNA. This process has been documented for the removal of introns from yeast genes (30) and raises the question of why introns persisted during evolution.

MUTATIONS IN CELLULAR GENES

Many human diseases are caused by mutations that interfere with RNA splicing. Approximately one quarter of all mutations in the human globin genes underlying thalassemia are mutations in sequences specifying correct

RNA splicing. Thus, conserving information for accurate splicing is a constraint on genetic systems. The limited nature of the mutations altering splicing in the globin family is interesting (31). All of the characterized mutations either alter the conserved 5′ or 3′ splice site sequences at the boundaries of the intron, or are changes within introns that generate new consensus splice sites (Fig. 4). The former is simple – mutation of the highly conserved GU or AG sequences at the boundaries of the intron commonly inactivates splicing at that site-this frequently leads to the activation of a nearby cryptic splice site. The latter is more complicated. In this case, a mutational change within an intron- CT → GT activates this site as a 5′ splice site and, as a consequence, an upstream AG is activated as a 3′ splice site for further splicing. This results in the generation of a short exon from the previous intron sequences (Fig. 4). Thus, mutations can alter both the position of splice sites and the number of exons to generate novel proteins.

Mistakes are probably not uncommon in splicing of RNA from complex genes. Exons can occasionally be skipped and, in some cases, circular RNAs can be generated (32, 33). Since specific functions have not been assigned to many of these RNAs, their generation may reflect noise in the system. Interestingly, it has recently been proposed that an mRNA surveillance system exists in cells which destroys RNAs as they enter the cytoplasm if they contain an open reading frame interrupted by a translational termination signal (34). This system would probably degrade most mRNAs with splicing errors as they are transported to the cytoplasm.

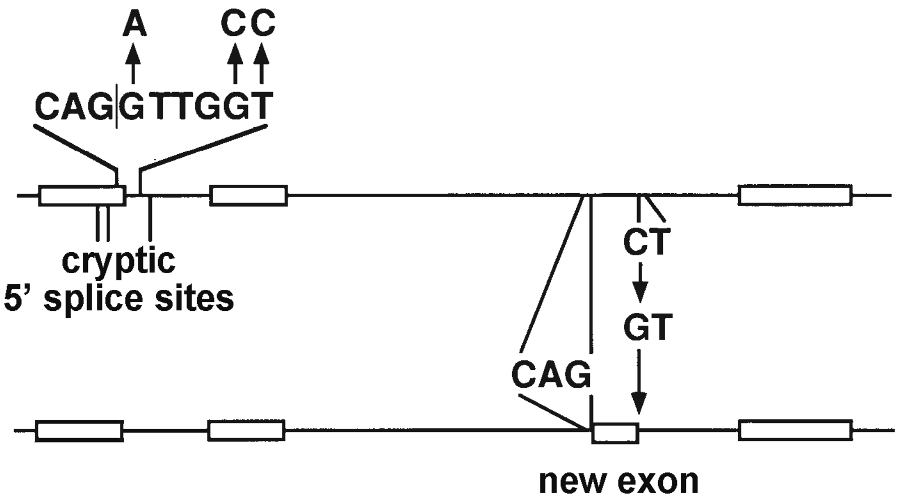

Fig. 4: β-Globin mutations. β-Globin genes of thalassemia patients possessing mutations which affect RNA splicing (31). The single base changes in four independent patients are indicated by arrows from the diagram of the two intron structure of the β-globin gene. Three of the mutations alter conserved sequences of the 5′ splice site and a fraction of the mRNA from these mutant genes are processed at the cryptic 5′ splice sites. The fourth mutation is within the sequences of the second intron and creates a 5′ splice site at this position. This results in the induction of the indicated sequences as a new exon. Exon sequences are represented by rectangles, intron sequences by lines.

WHY SPLIT GENES AND RNA SPLICING?

It is diffkult to confidently account for why introns have been conserved during evolution within the genes of most, if not all, eukaryotes. Clearly the intron-exon structure of genes has been very important in the generation of new genes during evolution. For example, the fibronectin gene of man is composed of three types of exons, each of which encodes a specific type of protein folding domain (Fig. 5; 35). These same three protein-folding domains and corresponding exon structures are found in other genes, some of which encode cell surface receptors and blood coagulation proteins. Thus, the fibronectin gene was created by tandem and dispersed duplication of exon units using breakage and joining within the intron sequences. Since these exon units are assembled precisely by RNA splicing and encode functional protein domains, the final protein is stable and has multiple functions. This theme of duplication and utilization of an exon unit has been a common mechanism for the generation of new genes encoding many cell surface receptors and other types of proteins in vertebrates. Thus, the presence of exon units which correspond to functional protein domains has been critical for the evolution of complex organisms.

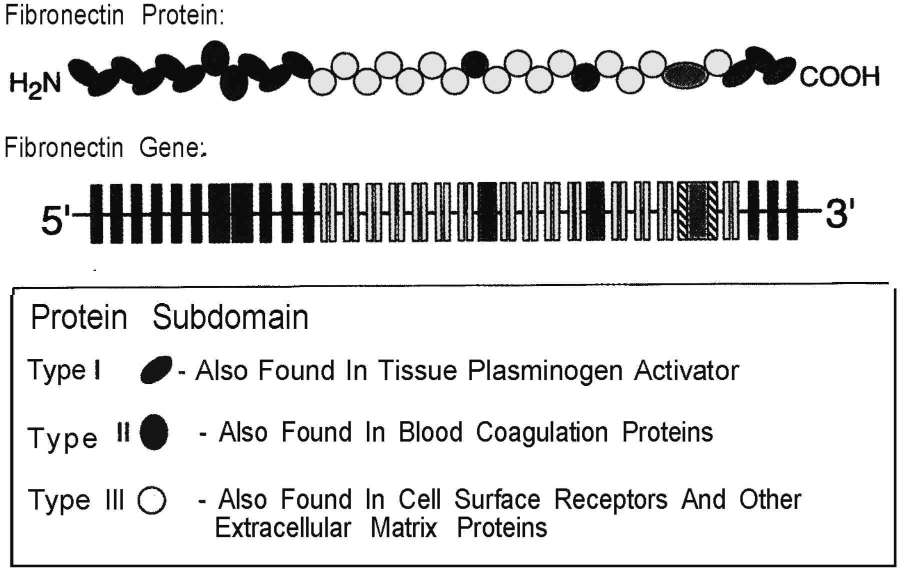
Fibronectin Protein:
H_2N
COOH
Fibronectin Gene:
5'
3'
Protein Subdomain
Type I - Also Found In Tissue Plasminogen Activator
Type II - Also Found In Blood Coagulation Proteins
Type III ○ - Also Found In Cell Surface Receptors And Other Extracellular Matrix Proteins

Fig. 5: Fibronectin gene evolved by exon duplication. The **large extracellular matrix protein, fibronectin, is primarily composed of three types of protein domains, I, II, and III denoted as tilted ellipsoids, vertical ellipsoids and circles, respectively. As indicated at the bottom, homologous domains are found in other cellular proteins. Each of these domains are encoded by a defined exon pattern, one or two exons, denoted as verticle rectangles in the middle. These exon units are duplicated in the ftbronectin gene and also in the other genes containing these domains. In two cases, larger rectangles, the intron separating the typical two exon structures of the type III subdomain, have been deleted. These patterns suggest that descendants of a common progenitor exon configuration was used to generate parts of these genes (36).**

The ability to alternatively select different combinations of exons at the stage of RNA splicing to generate proteins with different functions is clearly critical for the viability of many vertebrate organisms. Approximately, one of every twenty genes is expressed by alternative pathways of RNA splicing in different cell types or growth states. For example, nuclear precursor RNAs from the fibronectin gene are alternatively spliced into mRNAs that encode 20 different proteins. These proteins have slightly different functions reflecting their variant structures. Furthermore, a set of exons is not included in mRNAs processed in liver cells but is included in mRNAs in other cell types (Fig. 6). The fibronectin secreted by the liver more readily circulates in the blood stream because it has fewer protein domains which are recognized by receptors on the cell surface than the other forms of fibronectin (36). The fibronectin secreted by other cells is more typically found in the intracellular matrix of solid tissue. A similar variation in cellular adhesion is observed with alternative splicing patterns in another part of the fibronectin mRNA. Thus, through alternative splicing, the same set of gene sequences can be utilized for different functions.

Alternative splicing of precursor RNAs in different cell types must reflect differences in the factors regulating the splicing process. Only in few cases have these factors been identified, primarily by the analysis of mutants which are defective in the regulatory step. The development of sexual dimorphism in the fruit fly *Drosophila* is regulated at the level of alternative RNA splicing by expression of a cascade of genes. Early in development, the

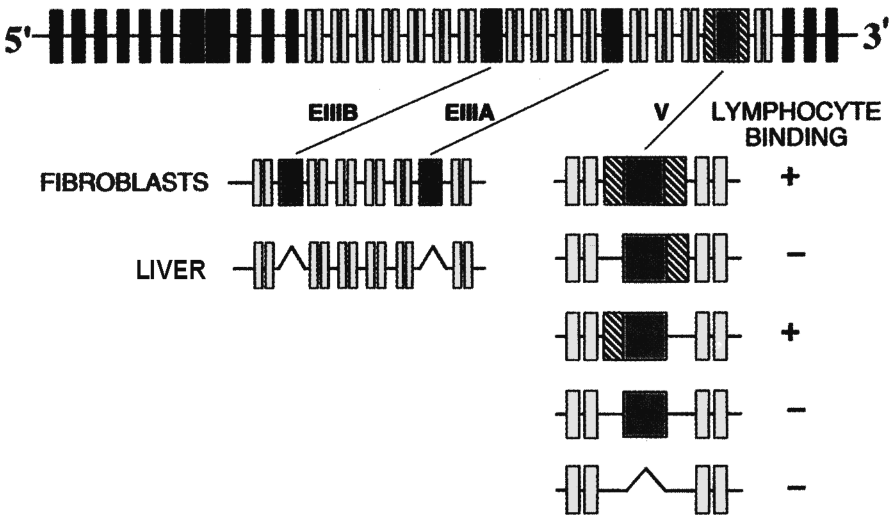

Fig. 6: Alternative splicing of RNA encodingfibronectin. The exons denoted as EIIIB and EIIIA are spliced into the mature RNA synthesized in many cell types including fibroblasts. However, in the liver, these two exons are not included, are skipped, in synthesis of the mRNA. Furthermore, five variations for the splicing patterns of the V region are shown. All of these variations on splicing are found in most cell types. The fibronectin proteins encoded by the first and third mRNA bind differentially to receptors on lymphocytes. In total, at least twenty different proteins are synthesized from the single fibronectin gene.

female or male specific pattern of splicing is set by a reading of the ratio of sex chromosomes to autosomal chromosomes. After setting this switch, the developmental process is controlled by an autoregulatory process which maintains the male or female splicing pattern in many if not all cells (37). Interestingly, most of the proteins encoded by these sex regulating genes are members of a family whose common features are one or more RNA recognition subdomains and a pronounced tract of the repeating amino acid sequence Arg-Ser (38). It is likely that these factors directly interact through the Arg-Ser tract with other proteins critical for formation of the splicing machinery.

SPLIT GENES IN THE PROGENOTE?

The split gene structure must be very old, clearly predating the divergence of organisms that gave rise to plants, yeast and man (29). First, as will be described later, a complex splicing machinery, which involves greater than 50 – 100 proteins and five small RNAs, exists in the nucleus of all eukaryotes. This machinery is highly conserved and could not have arisen spontaneously in the various lineages. Second, introns are conserved in positions within homologous genes found in the lineages of plants and animals. This evidence establishes that split genes and RNA splicing were present in common progenitor organisms a billion years ago.

Several scientists have speculated that genes originally evolved as exons and that the progenote organism from which current prokaryotic and eukaryotic organisms evolved may have had a split gene structure (39 – 42). These primordial exons are pictured as encoding sequences for stable protein folding domains. Assembly of a number of exon sequences by RNA splicing would be expected to produce a protein composed of stable folding domains which have a high probability of being functional either structurally or catalytically. If genes originally evolved in this fashion, the positions of introns in relationship to protein secondary structure might not be random. Evidence to support this hypothesis has been sought in the exon-intron structure of evolutionarily old proteins critical for energy metabolism. For example, the enzyme pyruvate kinase (PK) is conserved in structure and functions in bacteria, yeast and chicken (Fig. 7; 43). The gene is not interrupted with introns in bacteria and yeast, but contains ten introns in chicken, nine within the coding sequences. The positions of these nine introns are not random when compared to the protein structure. First, in the N-terminal part of the protein, the first three introns are positioned between repeating structural motifs of α helix-β sheet. Second, in the mononucleotide binding fold. PK shares a common intron position with one other old metabolic enzyme, alcohol dehydrogenase. In PK, the eighth intron occupies a position similar to that of an intron in the dehydrogenase. Since the generation of these two enzymes from a common protein subdomain clearly predated the evolutionary divergence of prokaryotic and eukaryotic organisms, these results suggest that the progenitor organisms for

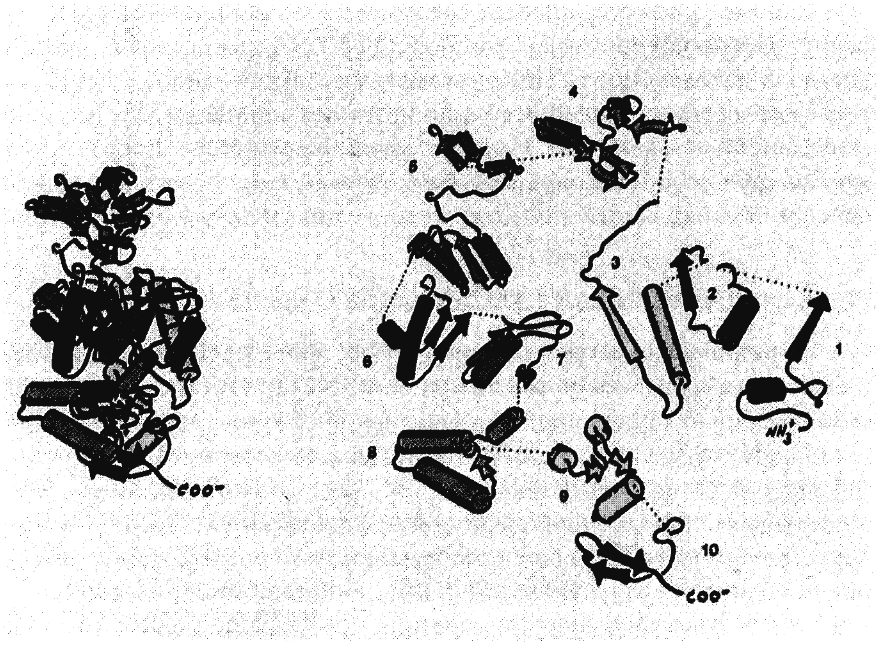

Fig. 7: The structure of pyruvate kinase and intron positions. Pyruvate kinase is shown schematically as a protein with secondary and tertiary structure on the left. On the right, this structure is expanded at the positions of introns in the tertiary structure of the protein. Note that the first three introns all fall between a-helix (barrels) and β-sheet (arrows) repeats. Intron 8 is positioned in approximately equivalent positions in the mononucleotide binding fold of PK and maize alcohol dehydrogenase. *Reprinted* with permission from *ref* 44.

both prokaryotic and eukaryotic organisms may have had a split gene structure more typical of a current eukaryotic cell (for discussion see 44).

IS THE GENE AN EXON?

The gene was first genetically defined as a unit of inheritance which is associated with a locus on a chromosome. The chemical definition of a gene has become much more difficult however, as the complexity of genetic information and its modifications are discovered. The existence of alternative splicing of exons, where information at an exon unit can be optionally expressed in some cells and not in others, suggests that an exon might correspond to a gene. That is, an exon corresponds to the minimal amount of information which is expressed as a discrete unit. This concept becomes particularly relevant when the trans-splicing process is considered (45). In this case, exons transcribed from different loci, and in many cases different chromosomes, are joined by RNA splicing. Trans-splicing of exons and introns has been established in the parasite trypanosomes (46), the flat worm C. elegans (47), and suggested for some human genes (48). Clearly, in the case of trans-splicing, both the unit of inheritance and the locus on a chromosome can correspond to a single exon.

It is unlikely, however, that the current working definition of a gene as a linear collection of exons which are joined by RNA splicing will be radically altered in the near future. This is probably wise, as the existence of multiple processes collectively called RNA editing further complicates the biochemical definition of a gene (49). However, given the possibility that the earliest unit of genetic information may have evolved as an exon, the general concept of exons as gene units may be more valid than any other proposal.

SPLICING OF NUCLEAR PRECURSORS TO mRNAs

The development of a reaction composed of soluble cellular components which accurately processed precursors to mRNAs (pre-mRNAs) was critical for advancing our understanding of splicing (50). When combined with the use of highly radioactive pre-mRNA substrates, a biochemical analysis of the splicing process became feasible (51, 52). Not surprisingly, kinetic RNA intermediates in the splicing reaction were soon identified (53, 54). Surprisingly, these intermediates had a lariat structure where the 5′ most nucleotide of the intron was joined in a 2′-5′ phosphodiester bond to an adenosine within the intron (55). Since the adenosine is covalently bonded through both 3′-5′and 2′-5′ phosphodiester linkages, this forms an RNA branch. The existence of such branches had just been described from nuclease digestion studies of total nuclear RNA from human cells (56). Formation of the branch appears simultaneously with cleavage at the 5′ splice site and generates the lariat RNA which typically migrates more slowly than the pre-mRNA during electrophoresis through a tight porosity polyacrylamide gel.

The splicing of pre-mRNA proceeds in two steps (Fig. 8, left). As mentioned above, the first step consists of cleavage at the 5′ splice site with the concomitant formation of the branch. At this stage, the 5′ exon RNA has a 3′ hydroxyl group and the lariat intermediate RNA contains the intron and 3′ exon. The second step consists of cleavage of the RNA at the 3′ splice site with concomitant joining of the two exons. The intron is released as a lariat RNA and is reasonably stable in reactions in vitro. This contrasts with the situation in vivo where intron RNAs are almost always rapidly degraded.

The fact that the intermediate state, consisting of two RNAs, was efficiently converted to the final products strongly suggested that these RNAs remain bound in a complex. The complex was identified by its rate of sedimentation in a glycerol gradient, 60S, and was designated to be a spliceosome or splicing body (57, 58). As anticipated from earlier work suggesting the importance of small nuclear ribonucleoprotein particles in splicing, the spliceosome contained the small nuclear RNAs (snRNAs) U2, U4, U5 and U6 and, under certain conditions, U1 (59). Thus, the spliceosome, much like a ribosome, contains a substrate RNA and a number of stable cellular RNA-protein components.

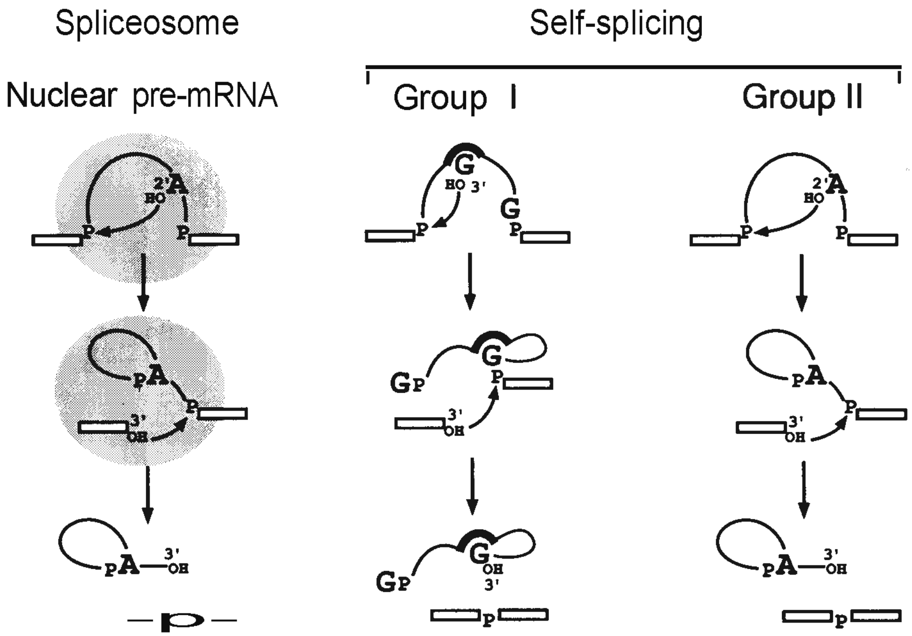

Fig. 8: Comparison of self-splicing and nuclear pre-mRNA splicing mechanism. **The first column outlines the mRNA precursor splicing mechanism. The shaded circle represents a multicomponent complex, the spliceosome, which promotes the splicing reaction. The second column outlines the splicing mechanism of self-splicing introns of the group I type. This process is catalyzed by RNA structures within the intron (dark semicircle), which contains a guanosine binding site, and utilizes a guanosine (C) factor in the first step. The third column outlines the splicing mechanism of self-splicing introns of the group II type. This process is also catalyzed by RNA structures within the intron but utilizes, instead of a cofactor, an adenosine residue (A) within the intron to form a lariat RNA. All three mechanisms proceed by two steps, reaction at the 5′ splice site and then reaction at the 3′ splice site. The fate of the phosphates at the 5′ and 3′ splice sites is indicated.**

GROUP I, GROUP II AND THE SPLICEOSOME

Comparison of the two steps in splicing by the spliceosome to the RNA-catalyzed self-splicing reactions of group I and II introns shows some striking similarities (Fig. 8). In all three cases, the first step is cleavage at the 5′ splice site. In group I introns, this cleavage requires a guanosine which specifically occupies a binding site in the catalytic intron sequences (60). The 3′ hydroxyl group on this guanosine is activated and through a transesterification reaction displaces the 3′ hydroxyl of the 5′ exon. Group II self-splicing introns cleave at the 5′ splice site by activating the 2′ OH at the branch site, producing a lariat RNA (61, 62) much like that produced by the spliceosome. The second step for all three processes involves a reaction at the 3′ splice site to join the exons and displace the intron. The simplest mechanism for the second step of splicing of the group I and II introns is a single transesterification reaction. This has been proven to be the case for reactions by group I introns.

The similarities between the spliceosomal process and the self-splicing introns suggested that these reactions might be evolutionarily related. This is particularly the case for the group II and spliceosome reactions. Thus, the snRNAs in the spliceosome might be pictured as a group II intron in pieces, where these RNAs form the catalytic sites for both reactions. The proteins in the spliceosomes could be necessary for recognition of the precursor RNA, arranging the snRNAs in catalytic structures and rearrangement of the components to facilitate completion of the process. Early in evolution all introns might have been self-splicing. Then, during the passage of time, tram-acting components may have developed which executed the splicing of introns by recognition of sequences at the splice sites, thus permitting atrophy of the cis-sequence within each intron (63). *Trans*-acting segments of group II introns which complement the splicing of introns with partial catalytic structure have been documented (see reference 63).

Years after the discovery of the lariat intermediate in the *cis*-splicing process, studies of the *trans*-splicing reaction revealed a branched intermediate state. Thus, the chemistry of the truns-splicing reaction is very similar to that of the cis-splicing process (64). It is therefore highly likely that the *cis-* and truns-splicing processes are variations of a single fundamental mechanism. Trypanosome parasites which exclusively synthesize mRNA by truns-splicing of a short leader RNA do not apparently express Ul snRNA or U5 snRNA, but do express U2, U4 and U6 snRNAs (65).

The sequence complementarity between the 5′ end of Ul snRNA and the consensus sequences at the 5′ splice site led to the hypothesis that this interaction was critical for splicing (66, 67). This hypothesis has been confirmed by a number of studies including the inhibition of splicing in vitro by addition of anti-sera specific for Ul snRNP (68). After the discovery of lariat RNA, a consensus sequence was recognized in the region flanking the branch site. This consensus sequence is complementary to a conserved internal sequence in U2 snRNA, and mutational analysis has shown that this interaction is also critical in splicing (69). Ul and U2 snRNAs recognize consensus sequences at the 5′ splice site and branch site as early steps in the splicing reaction (Fig. 9A). At the same stage of splicing, U4 and U6 snRNAs are bound to one another through an extended region of complementarity (70). SnRNA U6 is unique relative to the other snRNAs in that: it is not bound by core peptides recognized by the Sm lupus antisera, it is transcribed by a different polymerase, and it has a different type of modification at its 5′ terminus. U6 snRNA is also the most conserved member of the snRNA family and may have a unique catalytic role. The U4/6 snRNP forms a specific complex with U5 snRNP and this tri-snRNP complex is thought to bind to the other components in formation of the spliceosome (71).

After formation of the spliceosome, there are major rearrangements of the snRNAs (Fig. 9B). Both genetic and biochemical experiments show that U6 and U2 snRNAs are paired through an extensive tract of complementarity in the spliceosome (72, 73). Formation of the U2 and U6 structure

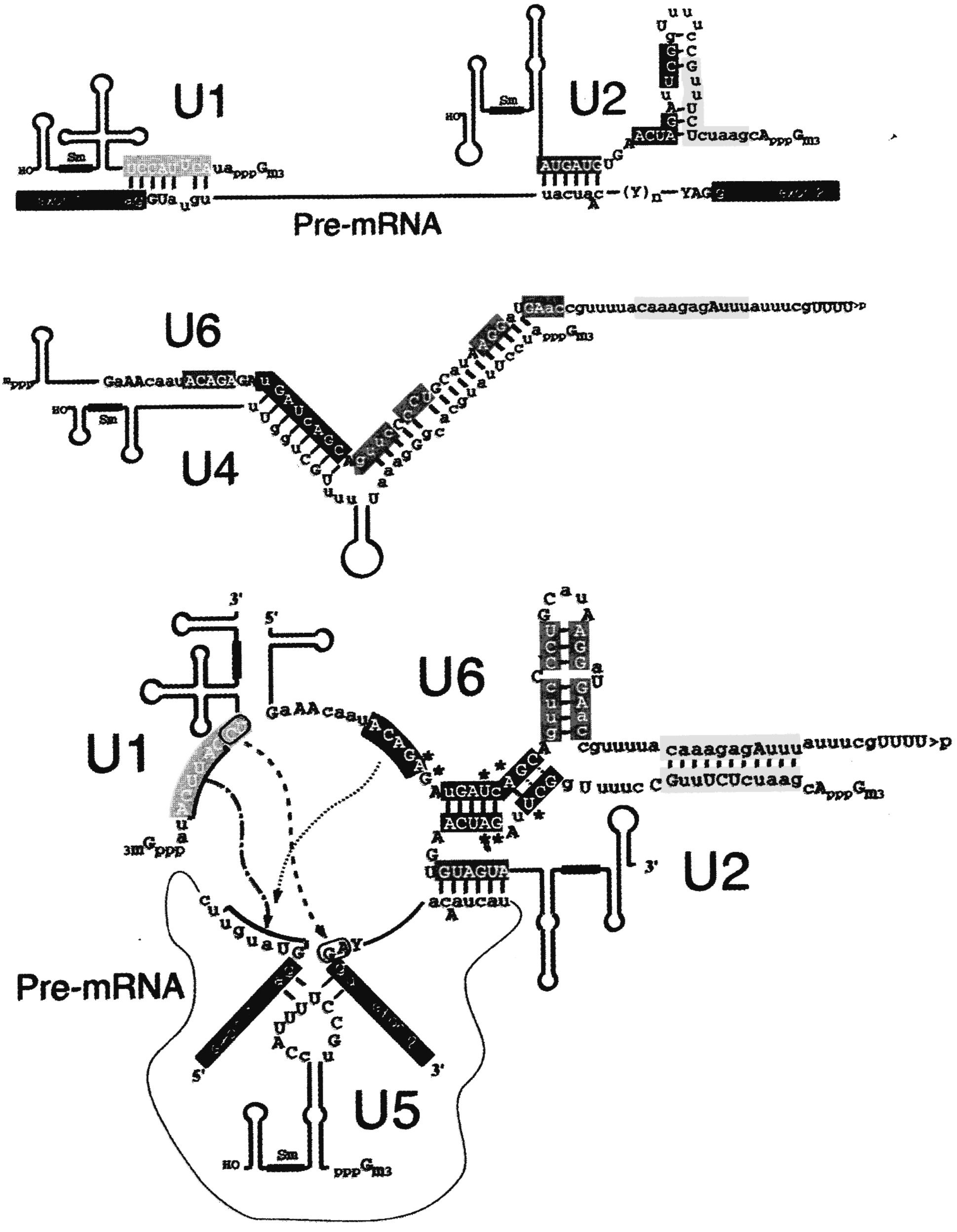

Fig. 9: RNA interactions between spliceosomal snRNAs and pre-mRNA substrates. (A) *(Top)* Base-pairing interactions between Ul and U2 snRNAs and pre-mRNA are indicated on left and right of intron, respectively. (Bottom) Extensive base-pairing between U4 and U6 snRNAs. (B) Interactions between Ul, U2, U5, and U6 snRNAs and pre-mRNA. In both A and B, pre-mRNA consensus sequences and snRNA sequences are those of S. cerevisiae; uppercase nucleotides are highly conserved between S. cerevisiae and the known sequences of other organisms (excluding trypanosomes, which do not have a GUAGUA sequence in U2 snRNA). Different shaded areas highlight sequences in U2 and U6 snRNAs that change base-pairing partners during the spliceosome cycle. Internal snRNA secondary structures that do not change between A and B are shown as stylized stems and loops. Asterisks indicate snRNA positions at which mutations specifically block the second step of splicing.

requires dissociation of U4 snRNA from U6 snRNA. In fact, U4 snRNA can be released from the active spliceosome after this transition. The complementarity between Ul snRNA and the 5′ splice site sequence is probably also dissociated before the first step of splicing. The 5′ splice site almost certainly pairs with another region of U6 snRNA (74, 75). U5 snRNA is thought to be important in recognition of the exon sequences immediately flanking the splice sites. Mutations in these flanking sequences frequently have significant effects on splicing efficiency and these effects can be reduced by complementary changes in U5 snRNAs (76). It is possible that there are further rearrangements in the secondary structures formed by the snRNAs between the first and second steps in splicing. There are mutational changes in U2 and U6 (positions denoted by *Fig. 9) which only inactivate the second step (72, 73). Inspection of the network of complementarity formed between snRNAs and the pre-mRNA in the spliceosome reveals a concentration of snRNA structure, perhaps tertiary, near the branch site and 5′ splice site sequences. This arrangement of snRNAs might form the catalytic site for the first step. Perhaps further rearrangements would form another catalytic site for the second step.

FORMATION OF THE SPLICEOSOME

A number of stable complexes containing snRNAs and pre-mRNA have been partially characterized and placed in a spliceosome cycle (Fig. 10; 77). The commitment complex, CC, forms on the pre-mRNA by recognition of the 5′ splice site sequence by Ul snRNP and sequences encompassing the branch site and 3′ splice sites (78, 79, 80). For mammalian systems, a pyrimidine tract near the 3′ splice site and branch site is particularly important. The subsequent binding of U2 snRNP results in formation of the very stable A complex. The stable binding of U2 snRNP to the pre-mRNA depends critically on a protein U2AF (81), which recognizes the polypyrimidine tract through prototypical RNA-binding domains and signals interactions with the other splicing components through a tract of Arg-Ser amino acid repeats. Interestingly, Ul snRNP which is also required for stable U2 snRNP binding, has a bound protein with a tract of Arg-Ser repeats, the **Ul** – 70 kd polypeptide. Whether U2AF and **Ul** – 70 kd directly communicate across the intron is not clear.

The Bl spliceosome forms when the U4/U6/U5 tri-snRNA complex binds the A complex. It is probably in Bl complex that U4 snRNA dissociates from U6 snRNA and perhaps Ul snRNA from the 5′ splice site. These events define the B2 complex which is the precursor of the Cl complex. The first step of splicing has occurred in the Cl complex. In generation of the C2 complex, it is likely that there are further rearrangements that are necessary to catalyze the second step in splicing. The C2 complex dissociates and the newly joined exons are released from the snRNP·intron complex I. In the cell, the spliced product migrates to the cytoplasm. The lariat intron is retained with the snRNPs and this I complex turns over. The snRNPs are

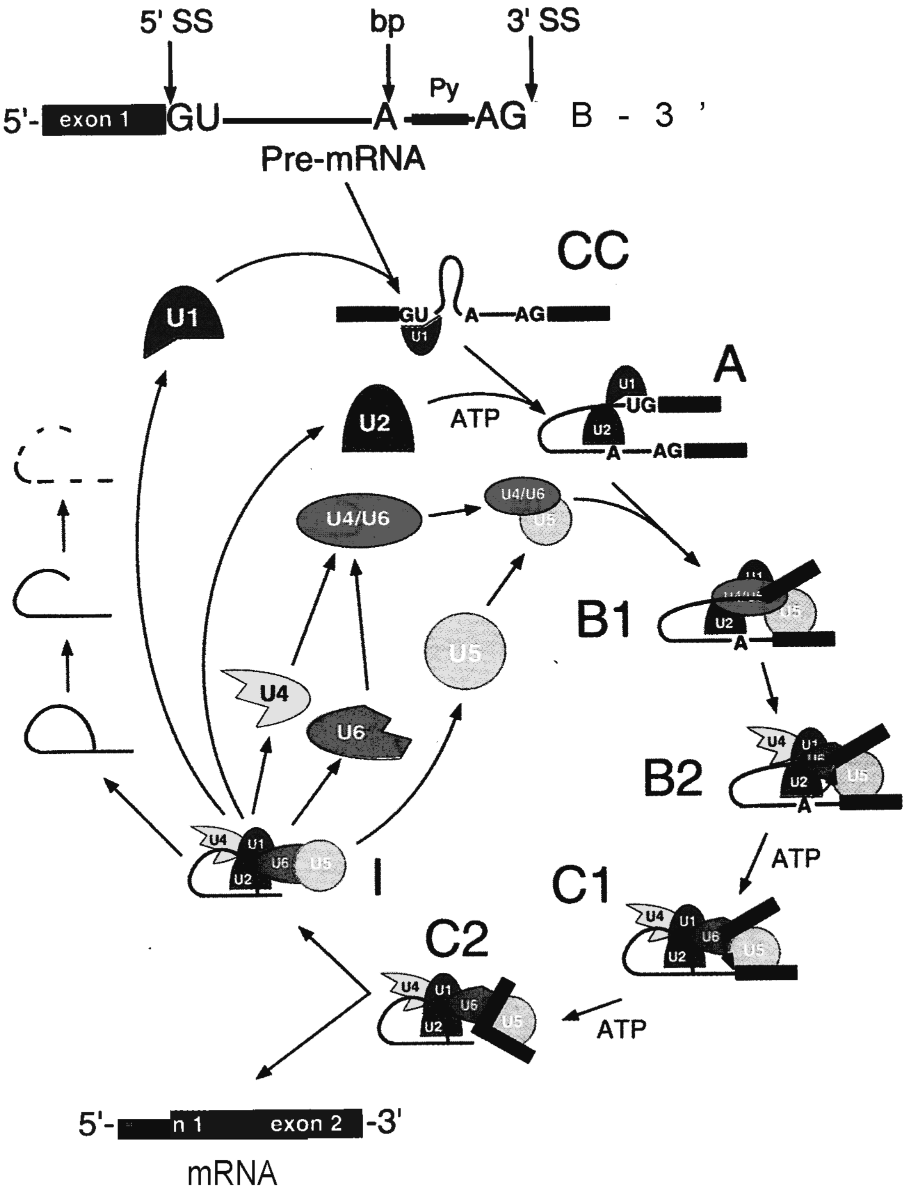

Fig. 10: Schematic representation of the spliceosome cycle in terms of the role of small nuclear ribonucleoprotein (snRNP) particles in pre-mRNA splicing. Pre-mRNA (top line), containing two exons separated by an intron, enters splicing complexes with snRNPs and exits as mRNA (bottom line) and excised lariat intron (left border). Other, non-snRNP factors are required for spliceosome formation, but have been omitted for simplicity. CC, A, Bl, B2, Cl, C2, and I represent complexes within the splicing pathway that have been distinguished biochemically and/or genetically. 5′ SS, 3′ SS, bs, and Py indicate 5′ and 3′ splice sites, branch site, and polypyrimidine tract, respectively. The individual snRNPs indicated are Ul, U2, U4, U5, and U6.

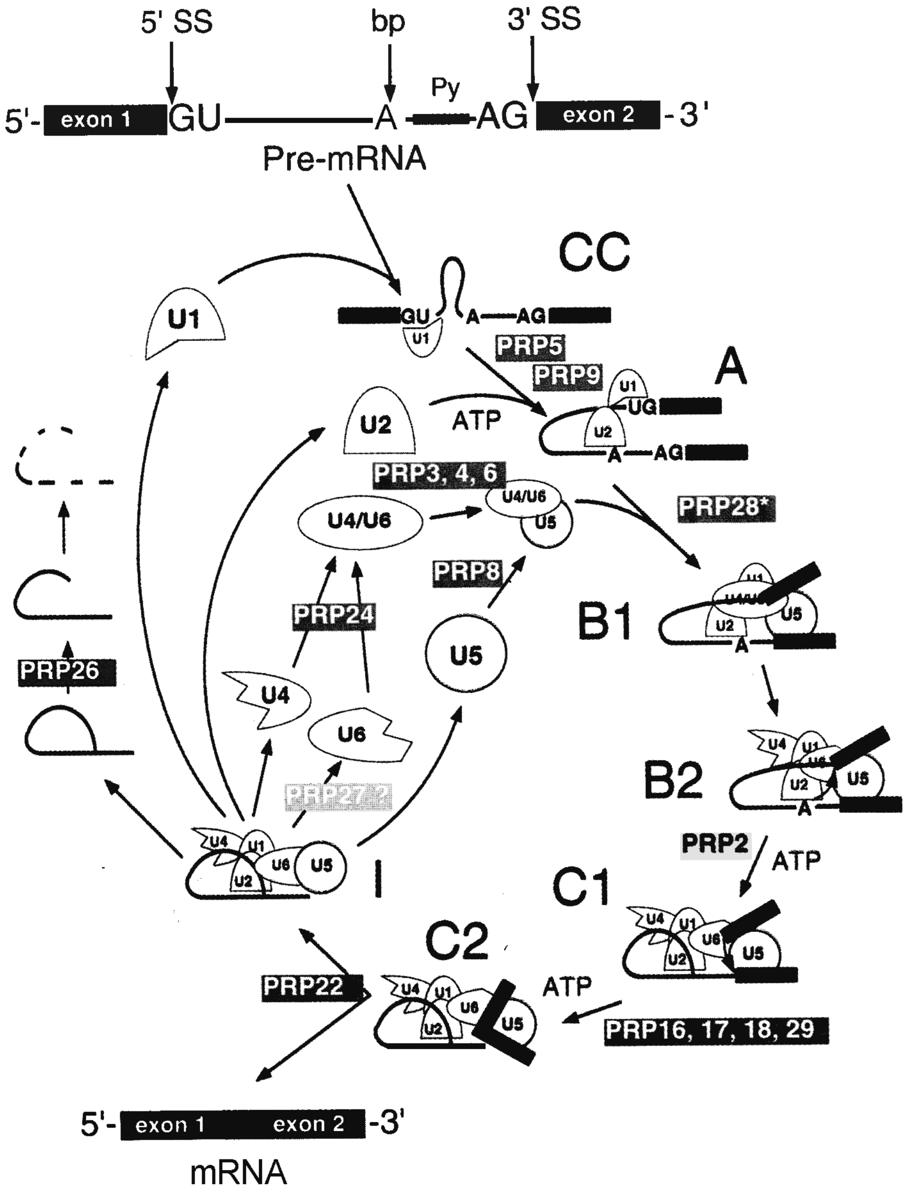

Fig. 11: Transitions in the spliceosome cycle which require a PRPprotein. The particular PRP mutant is listed beside the arrow indicating the transition in the cycle in vitro which requires the mutant protein.

recycled for further splicing in vivo, while the intron RNA is degraded. Since all snRNAs are very stable, a particular snRNA likely participates in many splicing cycles.

Assembly and rearrangements of complexes as elaborate as those in the spliceosome cycle would be expected to require a number of proteins. Sophisticated genetic analysis has identified a number of yeast gene products that are important for splicing either in vivo or in vitro (82, 83). Temperature sensitive mutations in PRP (precursor RNA processing) genes cause the accumulation of unspliced pre-mRNA in the nucleus and/or generate extracts defective for splicing in vitro (Fig. 11). Surprisingly, it is estimated that at least a hundred different genes encode products important for RNA splicing, about 2% of the total yeast genome. Thus, the splicing apparatus is a significant component of the nucleus. The amino acid sequences predicted for some of the PRP genes suggests that they have functions such as RNA binding, RNA helicase and protein-protein interaction properties. Matching these hypothetical functions to processes in spliceosomes will be challenging.

Reassuringly, the steps in the spliceosome cycle where particular PRP proteins are required (Fig. 11) are consistent with the cycle as defined by kinetic and biochemical methods. Most transitions between specific forms of the spliceosome require one or more specific proteins. Furthermore, a number of PRP mutants are defective in splicing because of their inability to reassemble snRNPs for further splicing. Thus, both genetic and biochemical results prove that the spliceosome cycle is the process responsible for excision of introns from split genes.

THE CHEMISTRY OF THE SPLICEOSOME

Both steps in the spliceosome process involve reactions between hydroxyl groups and a phosphodiester bond. The products and substrates in both steps contain the same number of covalent bonds and thus each step could be accomplished with a single transesterification reaction. This has been shown to be the case for group I self-splicing introns by analysis of the stereochemistry of the two reactions (Fig. 12; 84, 85, 86). Phosphorothioates in diester bonds of RNA are chiral, having either a Rp or Sp stereochemistry depending upon the position of the sulfur atom. The specific stereochemistry can be determined by its sensitivity to particular nucleases. If a chiral phosphorothioate group participates in a single transesterification reaction, its stereochemistry is inverted, for example from Rp to Sp. This fact permits counting of the number of transesterification reactions in a process and, thus, the detection of any potential transient intermediate.

As implied above, Rp and Sp phosphorothioates are not equally active in many catalytic sites. This is commonly interpreted as reflecting the unique role of one of the oxygen groups on the phosphate in interaction with a metal ion, a role which could not be equally well-fulfilled by a sulfur group.

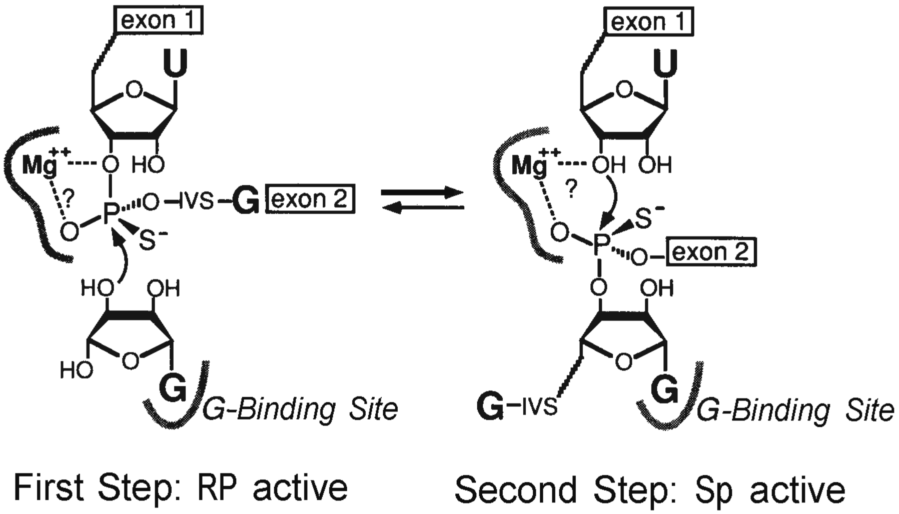

Fig. 12: The stereochemistry of the first and second step reactions of the group I self-splicing site. A sulfur atom is indicated at the positions where it does not inhibit the reactions, Rp and Sp positions for the first and second steps, respectively (84, 85, 86). The interactions between the oxygen group on the phosphate and the Mg+ + groups are hypothetical, proposed to explain why a phosphorothioate with the sulfur atom in this position is not reactive.

These principles of phosphorothioate chemistry have been analyzed for the group I self-splicing intron and now have been studied in the spliceosome.

The two steps of splicing by the group I intron are probably executed as a forward and reverse reaction at a single catalytic site (Fig. 12). In the first step, the guanosine cofactor is bound to the G binding site and its 3′ OH group is activated for the transesterilication reaction. In this reaction, the Rp chiral phosphorothioate is active while the Sp is not. This suggests that the oxygen group in the equivalent Sp position is uniquely recognized in this reaction, perhaps by a metal ion. Consistent with a single transesterification mechanism during the first step, the Rp chiral center is converted to a Sp phosphorothioate product (84). In the second step, the guanosine group at the 3′ splice site occupies the G binding site and the 3′ OH of the 5′ exon is activated for the transesterification reaction. As would necessarily be the case by the principle of microscopic-reversibility, the Sp chiral phosphorothioate is active in this reaction and the Rp is not (86). The Sp substrate phosphodiester bond generates a Rp product. Thus, stereochemical analysis of the group I process strongly supports the hypothesis of a single catalytic center for both steps.

The stereochemistry of the two steps in the spliceosome process was investigated by synthesis of substrate RNAs containing either a Rp or Sp phosphorothioate at either the 5′ or 3′ splice site (87). The particular stereoisomer was incorporated by a combination of chemical synthesis, oligo-primed transcription and ligation of RNAs by use of a bridging oligodeoxynucleotide and T4 DNA ligase (88). Splicing of the phosphorothioate-containing RNA substrates was tested in nuclear extracts after

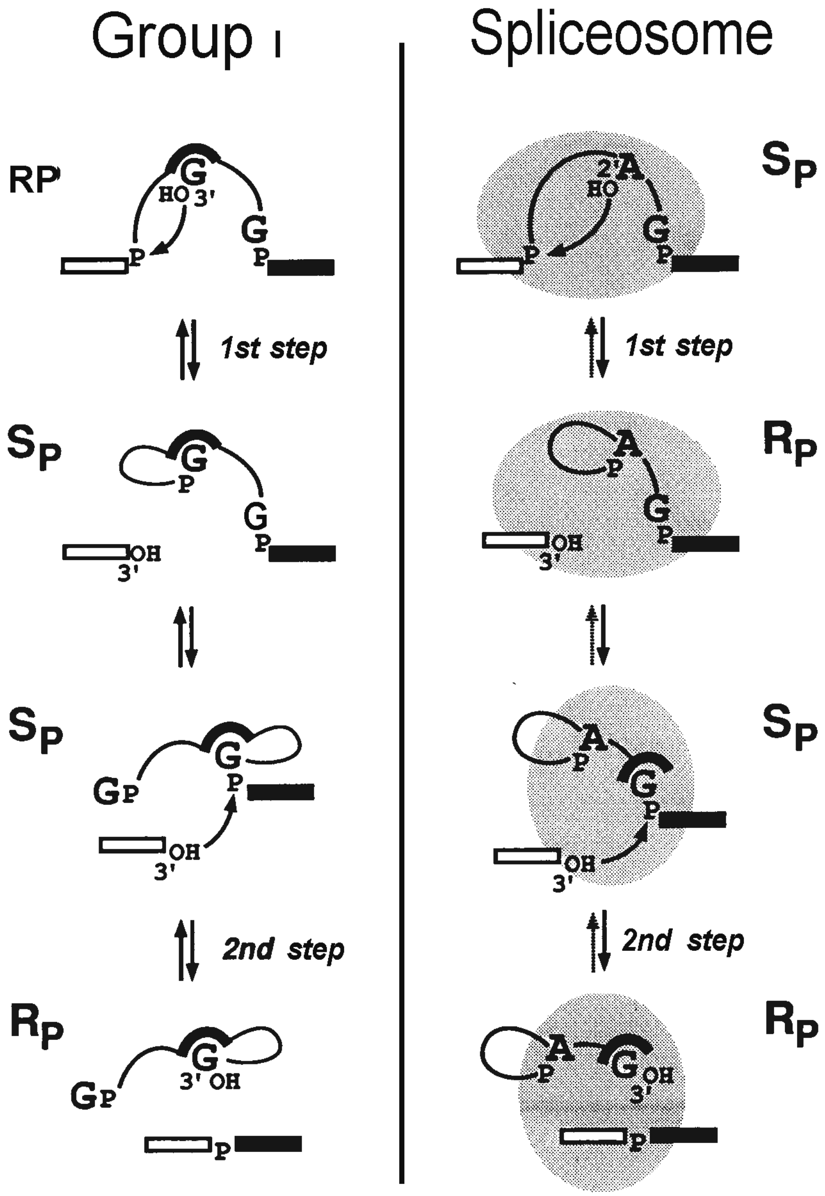

Fig. 13: Chemical mechanisms and stereochemical configurations of the two steps of pre-mRNA splicing (right) and group I intron self-splicing (left) First-step differences between spliceosomal and group I splicing include opposite phosphorothioate diastereomer preferences, different reaction nucleophiles (Y-OH versus 3′-OH), differential placement (5′ versus 3′) of the conserved guanosine (G) with respect to the phosphate at the 5′ splice site. Second-step similarities include preference for the Sp phosphorothioate diastereomer, use of a 3′-OH as the nucleophile and a conserved guanosine (C) attached through a 3′ oxygen to the phosphate at the 3′ splice site. The phosphorothioate diastereomer (Rp or Sp), which is not significantly inhibitory, is indicated for the substrate for each step. The diastereomer that results is indicated for the product of each step. Small dashed arrows indicate the potential (but not yet observed) reversibility of each step of nuclear pre-mRNA splicing. The dark hemicircle, indicated in the spliceosome for the second step, designates a hypothetical site similar to that of group I introns.

different periods of incubation. Analysis of the substrate and product RNAs showed that (a) both steps in splicing involve a single transesterification reaction with inversion of the chiral stereochemistry of the active phosphorothioate, and (b) the Sp diastereomer was the active phosphorothioate in both steps, while the Rp diastereomer was not (8'7).

A comparison of the stereochemistry and critical components in the two steps in splicing for the group I self-splicing process and the spliceosome process is informative (Fig. 13). As discussed before, the two steps in the group I case are forward and reverse reactions within a single catalytic center. This is almost certainly not the situation for the spliceosome process. In both steps of the latter process, the Sp phosphorothioate was active; thus, the second step is not a reverse of the first step. These results are most consistent with the proposal that the spliceosome generates two different catalytic centers for the two steps. The constituents in these centers may partially overlap but the centers must be distinct. This is also consistent with the different chemical nature of reaction components in the two steps. A 2′ hydroxyl on an adenosine is activated in the first step, while a 3′ hydroxyl group on the 5′ exon is activated during the second step. Furthermore, the conserved guanosine residue is linked to the activated phosphate by a 5′ bond in the first step and a 3′ bond in the second. Thus, the spliceosome process, and probably the group II self-splicing process, must involve two distinct catalytic sites.

The catalytic site in the spliceosome responsible for the second step is probably similar to that of the group I self-splicing introns. These two sites share several common characteristics. Both catalyze the transesterification reaction using (a) a Sp, and not Rp, phosphorothioate at the active site, (b) the activated phosphate is linked to the 3′ position to a conserved guanosine residue, and (c) both activate a 3′ hydroxyl group on the 5′ exon. These similarities are probably determined by some common chemical structure among the two catalytic sites. Perhaps this common micro-structure reflects a shared RNA tertiary structure for the catalytic centers and a shared evolutionary origin.

NUCLEAR STRUCTURE AND RNA SPLICING

The processing of introns occurs from both nascent RNA which is being extended by polymerase and from post-transcriptional precursor RNA (89). Thus, the splicing apparatus must be located proximal to the gene, as well as in the regions between the gene and the nuclear pores. After completion of splicing, the mRNA is transported to the cytoplasm through one of the approximately 4,000 nuclear pores of a typical mammalian cell (90). The mechanism by which nuclear RNAs are transported from sites of transcription and splicing to pores remains a mystery.

The subnuclear locations of a number of proteins and RNAs important for RNA splicing have been studied by light and electron microscopy. The most striking feature is the concentration of snRNPs, U2, U4/6 and U5 in

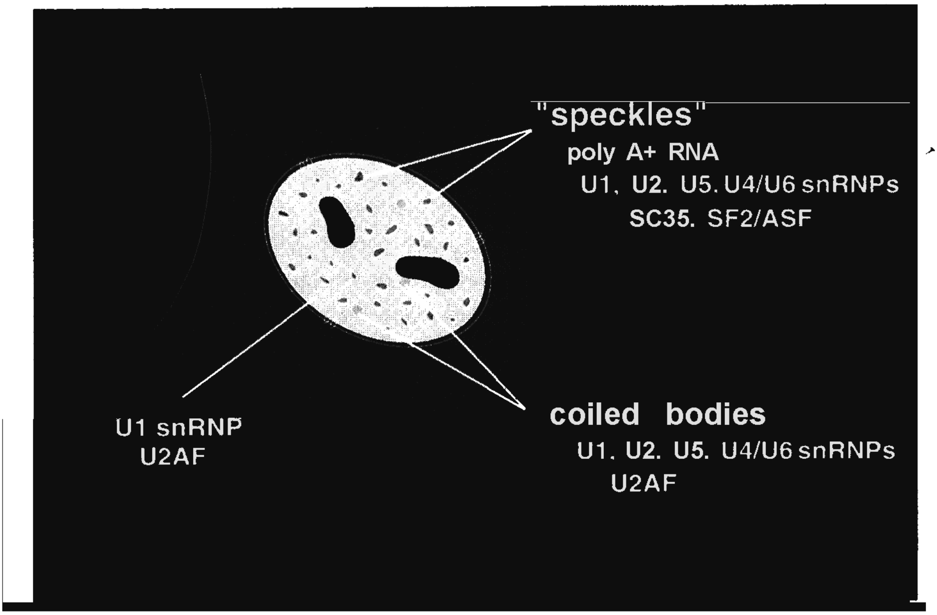

Fig. 14: The nucleus. Subnuclear localization of splicing factors. Specific antisera and in situ hybridization methods have been used to localize components of the spliceosome and splicing factors in the nucleus of mammalian cells. The outline of the cell is shown without structure in the cytoplasm. Three nuclear pores are diagrammed in the nuclear membrane. The large dark structures are nucleoli,.site of ribosomal RNA synthesis. Some of the components of the speckles, regions containing pre-mRNA, are listed. The coiled bodies are different and do not appear to contain pre-mRNAs (93). The general shadowing represents the nuclear distribution of Ul snRNP and U2AF.

20 – 50 speckled structures and also in 1 – 5 foci called coiled bodies (Fig. 14; 91, 92). This contrasts with the subnuclear location of Ul snRNP, which, although also concentrated in speckles and foci, is more uniformly dispersed throughout the nucleus. These speckled structures correspond to regions described by electron microscopists as interchromatin granule clusters and the perichromatin fibril network. High resolution in situ hybridization methods and pulse labeling studies locate newly synthesized RNA in the speckled regions and also in curvilinear tracks which extend from the gene toward the nuclear periphery (93, 94). The further the nuclear precursor RNA is located along the tracks which lead from its gene of origin, the lower the relative concentration of intron sequences as compared to exon sequences. This suggests that the precursor RNA moves through the speckled structures as it is being processed and transported (94). Active spliceosomes are probably concentrated in these regions.

Several proteins important for RNA splicing are also concentrated in speckled structures with the snRNPs. These include the Arg/Ser proteins SF2/ASF (95, 96) and SC35 (97). Surprisingly, these proteins, as well as the snRNPs in active spliceosomes, are probably attached to an operationally defined nuclear matrix. This matrix is the structure that remains in the nucleus after almost all of the chromatin proteins and DNA are extracted by sequential treatments with DNase 1 and high salt (98). The integrity of the fibrillar matrix which remains, is sensitive to RNase suggesting a major role for RNA in its structure. Spliceosomes have been shown to be associated with the nuclear matrix as it has been possible to process pulse-labeled endogenous precursor RNA extracted as part of the matrix (98).

Active spliceosomes may be associated with the nuclear matrix through interactions with proteins in the Arg/Ser family (99, 100). A panel of monoclonal antibodies has been generated to components in the nuclear matrix of human cells. Three of these monoclonal antibodies specifically stain the matrix and extensively co-localize with the snRNPs in the speckled structures. Surprisingly, these three antibodies specifically immunoprecipitate spliceosome complexes which contain exon sequences (101). Thus, the antibodies immunoprecipitate a fraction of the precursor RNA, the lariat intermediate RNA and the associated 5′ exon, and the spliced exon product RNAs. The antibodies do not immunoprecipitate the intron containing spliceosome complex I even though it contains most of the snRNPs and associated protein components. These remarkable results suggest that the matrix components bind to exon RNA sequences that are specifically assembled into splicing complexes.

The proteins recognized by the monoclonal antibodies to the nuclear matrix are most likely members of the Arg/Ser family of proteins or are associated with the Arg/Ser proteins. Precursor RNAs are not immunoprecipitated from reactions containing nuclear extracts depleted of Arg/Ser proteins. Further addition of preparations of purified Arg/Ser proteins will block specific immunofluoresence staining of subnuclear structures. These results are consistent with a model where the Arg/Ser family of proteins are associated with or form the nuclear matrix (101). Newly synthesized RNA sequences are recognized by the Arg/Ser proteins such as SC35, ASF, the U1 70 kd and UPAF and become associated with the matrix, perhaps, through interactions of the Arg/Ser tracts. On this structure, complete spliceosomes would form encompassing the matrix proximal splice sites and execute splicing. The exon product RNAs would remain matrix bound and move by some unknown mechanism to the nuclear pore.

CONCLUSION

The discovery of split genes and RNA splicing has been critical for studies of the biology of eukaryotic organisms. Gene regulation is central to all biological phenomena and RNA splicing is important in the regulation of genes, particularly when precursor RNAs are processed by alternative pathways to generate mRNAs encoding different proteins. The mechanism of splicing by the spliceosome is probably related to the self-splicing process of group I and II introns. The spliceosome process is old in an evolutionary sense, perhaps as old as the ribosomal process responsible for translation. Thus, the eukaryotic cell can be conjectured as consisting of two compartments, the nucleus where the spliceosome processes RNA precursors by RNA catalysis and the cytoplasm where the ribosome translates mRNAs by RNA catalysis. The distinct subnuclear locations of spliceosome-related components suggest a compartmentalized organization for the nucleus. Further studies of RNA splicing and transport hold promise of revealing the nature of the organizational specificity of the nucleus.

REFERENCES

1. Darnell, J. E., Jr. (1975) *Harvey Lectures, 69,* 1-47.
2. Furuichi, Y., Morgan, M., Muthukrishnan, S. and Shatkin, A. J. (1975) *Proc. Natl.* Acad. *Sci. USA,* 72, 362 - 366.
3. Wei, C. M. and Moss, B. (1974) *Proc. Natl. Acad. Sci. USA,* 71, 3014-3018.
4. Rottman, F., Shatkin, A. and Perry, R. P. (1974) Cell, 3, 197-199.
5. Edmonds, M., Vaughan, M. H., Jr., and Nakazoto, H. (1971) *Proc. Natl. Acad. Sci. USA,* 68, 1336- 1340.
6. Lee, S. Y., Mendecki, J. and Brawerman, G. (1971) *Proc. Natl. Acad. Sci. USA,* 68, 1331- 1335.
7. Darnell, J. E., Jr., Wall, R. and Tushinski, R. J. (1971) *Proc. Natl. Acad. Sci. USA,* 68, 1321- 1325.
8. Pettersson, U., Mulder, C., Delius, H. and Sharp, P. A. (1973) *Proc. Natl. Acad. Sci. USA,* 70, 200 - 204.
9. Sharp, P. A., Gallimore, P. H. and Flint, S. J. (1974) *Cold Spring Harbor Symp. Quant.* Biol., 34, 457 - 474.
10. Flint, S. J., Gallimore, P. H., and Sharp, P. A. (1975)J. *Mol. Biol., 96, 47 – 68.*
11. Flint, S. J. and Sharp, P. A. (1976)J. *Mol.* Biol., 106, 749-771.
12. Bachenheimer, S. and Darnell, J. E. (1975) *Proc. Natl. Acad. Sci.* USA, 72, 4445 - 4449.
13. Berget, S. M., Moore, C. and Sharp, P. A. (1977) *Proc. Natl. Acad. Sci.* USA, 74, 3171-3175.
14. Lewis, J., Atkins, J. F., Anderson, C., Baum, P. R. and Gesteland, R. F. (1975) *Proc. Natl. Acad. Sci. USA, 72, 4445-4449.*
15. Thomas, M., White, R. L. and Davis, R. W. (1976) *Proc. Natl. Acad. Sci. USA, 73,2294 – 2298.*
16. Casey, J. and Davidson, N. (1977) *Nucleic Acid Res., 4,* 1539-1552.
17. Westphal, H., Meyer, J. and Maizel, J. (1976) *Proc. Natl. Acad. Sci. USA, 73,* 2069-2071.
18. Chow, L. T., Roberts, J. M., Lewis, J. B. and Broker, T. R. (1977) *Cell,* 11, 819-836.
19. Goldberg, S., Weber, J. and Darnell, J. E., Jr. (1977) *Cell,* **10,** 617-622.
20. Weber, J., Jelinek, W. and Darnell, J. E., Jr. (1977) *Cell,* **10,** 611-616.
21. Berget, S. M. and Sharp, P. A. (1979)J. *Mol. Biol.,* **129,** *547-565.*
22. Perry, R. P. and Kelley, D. E. (1976) *Cell, 8, 433 -442.*
23. Gelinas, R. E. and Roberts, R. J. (1977) *Cell,* **11, 533-544.**
24. Jeffreys, A. J. and Flavell, R. A. (1977) *Cell, 12,* 1097- 1108.
25. Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. V. and Leder, P. (1978) *Proc. Natl. Acad. Sci. USA, 75, 725 – 729.*
26. Breathnach, R., Manel, J-L. and Chambon, P. (1977): *Nature, 270,* 314-319.
27. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, 0. and Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA, 75, 1485 –* 1489.
28. Breathnach, R. and Chambon, P. (1981) *Annu. Rev. Biochem., 50, 349-384.*
29. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. and Sharp, P. A. (1986) *Annu. Rev. Biochem., 55* 1119-1150.

30. Fink, G. R. (1987) Cell, 49, 5 – 6.
31. Treisman, R, Orkin, S. and Maniatis, T. (1983) ***Nature, 302,*** 591-596.
32. Nigro, J. M., Cho, K. R., Fearson, E. R., Kern, S. E., Ruppert, J. M., Oliver, J. D., Kinzler, D. W. and Vogelstein, B. (1991) Cell, 64, 607-613.
33. Cocquerelle, C., Daubersies, P., Majerus, M.-A., Kerckaert, J.-P. and Bailleul, B. (1992) EMBO J., **11,** 1095-1098.
34. Pulok, R. and Anderson, P. (1993) Genes *Dev.,7,* 1885- 1897.
35. Patel, R. S., Odermatt, E., Schwarzbauer, J. E. and Hynes, R. 0. (1986) EMBO ***J., 6, 2565-2572.***
36. Hynes, R. 0. (1989) ***Fibronectin,*** Springer-Verlag, New York.
37. Baker, B. S. (1989) ***Nature,*** 340, 521-524.
38. Tian, M. and Maniatis, T. (1992) ***Science,*** 256, 237-240.
39. Gilbert, W. (1978) ***Nature,*** 271, 501.
40. Darnell, J. E., Jr. (1978) ***Science,*** 202, 1257.
41. Doolittle, W. F. (1978) ***Nature, 272, 581.***
42. Blake, C. C. F. (1978) ***Nature, 273, 267.***
43. Stone, E. M. and Schwartz, R. J., editors (1990) ***Intervening Sequences in Evolution and Development*** , Oxford University Press, New York.
44. Lonberg, N. and Gilbert, W. (1985) ***Cell,*** 40, 81-90.
45. Agabian, N. (1990) Cell, 61, 1157-1160.
46. Borst, P. (1986) ***Ann. Rev. Riochem.,*** 55, 701-732.
47. Krause, M. and Hirsh, D. (1987) Cell, 49, 753-761.
48. Vellard, M., Sureau, A., Soret, J., Martinerie, C. and Perbal, B. (1992) ***Proc. Natl. Acad. Sci. USA, 89, 251*** 1-2515.
49. Benne, R., van den Burg, J., Branhenhoff,J. P. J., Sloof, P., van Boom, J. H. and Tramp, M. C. (1986) ***Cell,*** 46, 819-826.
50. Padgett, R. A., Hardy, S. F. and Sharp, P. A. (1983) ***Proc. Natl. Acad. Sci. USA, 80,523O – 5234.***
51. Green, M. R., Maniatis, T. and Melton, D. A. (1983) Cell, 32, 681-694.
52. Hernandez, N. and Keller, W. (1983) ***Cell, 35, 89-99.***
53. Grabowski, P. J., Padgett, R. A. and Sharp, P. A. (1984) ***Cell,*** 37, 415-427.
54. Krainer, A. R., Maniatis, T., Ruskin, B. and Green, M. R. (1984) ***Cell, 36, 993 – 1005.***
55. Konarska, M. M., Grabowski, P. J., Padgett, R. A. and Sharp, P. A. (1985) Nature, 313, 552-557.
56. Wallace, J. C. and Edmonds, M. (1983) ***Proc. Natl. Acad. Sci. USA, 80, 950-954.***
57. Grabowski, P. J., Seiler, S. R. and Sharp, P. A. (1985) ***Cell, 42, 345 -353.***
58. Brody, E. and Abelson, J. (1985) ***Science, 228, 963-967.***
59. Steitz, J. A., Black, D. L., Gerke, V., Parker, K. A., Kramer, A., Frendeway, D.,and Keller, W. (1988) In ***Structure and function of major and minor small nuclear ribonucleoprotein particle s*** (ed. M. L. Birnstiel), pp. 115 – ***154,*** Springer-Verlag, New York.
60. Cech, T. R. (1985) Cell, 43, 713-716.
61. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A. and Cheng, H.-L. (1986) ***Cell, 44, 213-223.***
62. van der Veen, R., Arnberg, A. C., van der Horst, G., Bonen, L., Tabak, H. F. and Grivell, L. A. (1986) ***Cell, 44, 225-234.***
63. Sharp, P. A. (1991) ***Science, 254, 663.***
64. Hannon, G. J., Maroney, P. A., Denker, J. A. and Nilsen, T. W. (1990) Cell, 61, 1247-1255.
65. Watkins, K. R., Dungan, J. M. and Agabian, N. (1993) ***Cell,*** in press.
66. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. (1980) ***Nature, 283, 220- 224.***
67. Rogers, J. and Wall, R. (1980) ***Proc. Natl. Acad. Sci. USA, 77. 1877-*** 1879.

68. Padgett, K. A., Mount, S. M., Steitz, J. A. and Sharp, P. A. (1983) Cell, 35, 101 – 107.
69. Parker, R., Siliciano, P. G. and Guthrie, C. (1987) Cell, 49, 229-239.
70. Guthrie, C. and Patterson, B. (1988) ***Annu. Rev. Genet.,*** 22, 387-419.
71. Konarska, M. M. and Sharp, P. A. (1987) Cell, 49, 763-774.
72. Madhani, H. D. and Guthrie, C. (1992) Cell, 71, 803-817.
73. McPheeters, D. S. and Abelson, J. (1992) Cell, 71, 819-831.
74. Sawa, H. and Abelson, J. (1992) ***Proc. Natl. Acad. Sci. USA, 89,*** 11269- 11273.
75. Wassarman, D. A. and Steitz, J. A. (1992) ***Science,*** 257, 1918- 1925.
76. Newman, A. and Norman, C. (1992) ***Cell,*** 68, 743-754.
77. Moore, M. J., Query, C. C. and Sharp, P. A. (1993) In The ***RNA World,*** Cold Spring Harbor Laboratory Press, pp. 303 - 357.
78. Seraphin, B. and Rosbash, M. (1991) *EMBO* J., 10, 1209-1216.
79. Michaud, S. and Reed, R. (1991). Genes ***Dev.,*** 5, 2534-2546.
80. Jamison, S. F. and Garcia-Blanco, M. A. (1992) ***Proc. Natl. Acad. Sci. USA, 89, 5482 – 5486.***
81. Ruskin, B., Zamore, P. D. and Green, M. R. (1988) ***Cell,*** 52, 207-219.
82. Ruby, S. W. and Abelson, J. (1988) ***Trends Genet., 7, 79-85.***
83. Guthrie, C. (1991) Science, 253, 157-163.
84. McSwiggen, J. A. and Cech, T. R. (1989) ***Science, 244, 679-683.***
85. Rajagopal, J., Doudna, J. A. and Szostak, J. W. (1989) ***Science, 244, 692-694.***
86. Suh, E.-R. and Waring, R. B. (1992) ***Nucleic Acids Res., 20, 6303 -6309.***
87. Moore, M. J. and Sharp, P. A. (1993) ***Nature, 365, 364- 368.***
88. Moore, M. J. and Sharp, P. A. (1992) ***Science, 256, 992-997.***
89. Beyer, A. L. and Osheim, Y. N. (1988) Genes *Dev.,* 2, 754- 765.
90. Mehlin, H., Daneholt, B. and Skoglund, U. (1992) ***Cell, 69,*** 605-613.
91. Spector, D. L. (1993) ***Annu. Rev. Cell Biol., 9, 265-315.***
92. Nyman, U., Hallman, I-I., Hadlaczky, G., Pettersson, I., Sharp, G. and Ringertz N. R. (1986) *J.* ***Cell.*** *Biol.,* 102, 137- 144.
93. Carter, K. C., Bowman, D., Carrington, W., Fogarty, K., McNeil, J. A., Fay, F. S. and Lawrence,J. B. (1993) Science, 259, 1330 - 1334.
94. Xing, Y., Johnson, C. V., Dobner, P. R. and Lawrence, J. B. (1993) Science 259, 1326-1330.
95. Ge, H. and Manley, J. L. (1990) ***Cell, 62, 25-34.***
96. Krainer, A. R., Conway, G. C. and Kozak, D. (1990b) ***Cell, 62, 35 -42.***
97. Fu, X.-D. and Maniatis, T. (1990) ***Nature, 343, 437-441.***
98. Zeitlin, S., Wilson, R. C. and Efstratiadis, A. (1989) J. ***Cell. Biol.,*** 108, 765-777.
99. Roth, M. B., Murphy, C. and Gall, J. G. (1990)J. ***Cell. Biol.,*** 111, 2217-2223.
100. Roth, M. B., Zahler, A. M. and Stolk, J. A. (1991)J. ***Cell. Biol., 115, 587-596.***
101. Blencowe, B. J., Nickerson, J. A., Issner, R., Penman, S. and Sharp, P. A. (1993) in preparation.