Regulation of alternative splicing in cell nuclei

Most eukaryotic pre-mRNAs contain intervening sequences (introns) that must be removed in order to place the coding sequences (exons) in a protein-reading frame. The precise removal of introns from pre-mRNAs by the pre-mRNA splicing machine is, therefore, an essential step in the regulation of gene expression. The human genome project revealed that alternative splicing and pre-mRNA processing are major contributors to the functional diversity of products encoded in the human genome. It has been estimated that 40-75% of the ~25,000 human genes undergo alternative splicing, thereby increasing the coding potential of the human genome by more than one order of magnitude. It is therefore not surprising that changes in splicing-site selection and alternative splicing are frequently associated with human pathologies and cancer.

The mechanism of the chemical transformations involved in pre-mRNA splicing, has been extensively worked out based on studies in vitro, mainly of pre-mRNAs composed of two exons and an intervening intron. However, the regulation of splicing and alternative splicing of the multi-intronic pre-mRNAs transcribed in vivo is still not well understood, particularly from the standpoint of the structure of the supraspliceosome – the nuclear macromolecular machine that catalyzes the splicing reactions. To better understand the complex regulation of alternative splicing we employed, in collaboration with the group of Prof. Ruth Sperling at the Hebrew University, a combination of functional, biochemical and structural analyses of the supraspliceosome complex isolated from living mammalian cells.

The supraspliceosome structure

The supraspliceosome is one of the largest protein-RNA complexes of which structural analysis has thus far been attempted. To study this huge splicing machine we took a top-down approach. Namely, we isolated the splicing machine from nuclei of living cells and initiated the analyses of its composition, structure and function.

To develop a protocol for the isolation of supraspliceosomes from mammalian cell nuclei under physiological conditions, we optimized conditions whereby over 85% of "H-labeled RNA polymerase II (pol II) transcripts could be released to the nuclear supernatant after mild sonication of the nuclei and removing the chromatin. Subsequent fractionation of such nuclear supernatants in sucrose gradients revealed that the general population of labeled pol II transcripts, including several specific pre-mRNA transcripts, sedimented at the 20S region of the gradient (Sperling et al., 1985; Spann et al., 1989; Sperling and Sperling, 1998). Visualization by EM of aliquots from fractions across the gradient revealed a peak of large tetrameric structures, having overall dimensions of 50 x 50 x 35 nm (Spann et al., 1989). Immunoprecipitation experiments, using antibodies against splicing factors, showed that all five spliceosomal U snRNPs are integral components of the supraspliceosome (Sperling and Sperling, 1990; Miriami et al., 1995; Yitzhaki et al., 1996; Sperling and Sperling, 1998). Similar experiments showed that all known protein splicing factors, including the non-snRNP protein splicing factor U2AF and all members of the SR protein family, are also integral components of supraspliceosomes (Sperling and Sperling, 1990; Miriami et al., 1995; Yitzhaki et al., 1996; Sperling and Sperling, 1998). A remarkable feature of supraspliceosomes is that they package pre-mRNA transcripts of different sizes and of different number of introns into complexes of a unique size and hydrodynamic properties, indicating their universal nature.

Structural studies of the supraspliceosome

The complexity and large size of the supraspliceosome makes EM the method of choice for its structural analysis. Three-dimensional (3-D) image reconstruction of individual supraspliceosomes by automated electron tomography of negatively stained (Medalia et al., 1997; Sperling et al., 1997), and of frozen hydrated complexes (Medalia et al., 2002) showed the supraspliceosome as forming a closed structure, composed mainly of four similar subcomplexes. Mass measurements by scanning transmission EM (STEM) showed that the supraspliceosome has a mass of 21 MDa, whereas each of its main...
subcomplexes has a mass of 4.8 MDa (Müller et al., 1998). Using a positive staining protocol, which allowed us the visualization of nucleic acids, we could show strands and loops of RNA emanating from positively stained supraspliceosomes (Müller et al., 1998), indicating that the RNA is loosely bound and therefore accessible for probing. Two-dimensional image restoration of ice-embedded particles revealed new high-resolution features, such as holes within the subcomplexes and fibers, presumably the pre-mRNA covered with proteins, through which the subcomplexes are interconnected (Medalia et al., 2002). These observations support our working hypothesis that each supraspliceosome packs one pre-mRNA and can therefore be responsible for its post-transcriptional processing. In this model (Figure 1), the supraspliceosome provides a platform to juxtapose exons about to be spliced, and each of the four major subcomplexes represents a spliceosome that can splice the intron wound around it. The supraspliceosome complex provides simultaneous multiple interactions for the assembly of the splicing machine with an intron, likely resulting in better efficiency and specificity of recognition of large introns. Furthermore, it should provide better specificity and efficiency of recognition of multi-intronic pre-mRNAs, which constitute the majority of the pre-mRNA population (Sperling et al., 1997; Müller et al., 1998). Thus, the supraspliceosome configuration is compatible with the requirements for alternative splicing and with the fact that the splicing of multi-intronic pre-mRNAs does not occur in a sequential manner.

Structural studies of native spliceosomes

The four substructures of the supraspliceosome are interconnected in a flexible way and may thus adopt different angular settings, which impose a significant restriction on reaching high resolution in EM image analyses. We have therefore developed a methodology to prepare and isolate the monomeric spliceosomal subcomplexes (termed native spliceosomes) from supraspliceosomes, and carried out image reconstruction of frozen hydrated particles by the single particle technique. For the isolating of native spliceosomes we specifically cleaved the general population of pre-mRNAs within supraspliceosomes, while keeping the snRNAs within these substructures intact. The resulting monomeric native spliceosomes were purified by centrifugation in a glycerol gradient, where they sedimented at the 60-70S region. The biochemical composition of the 60-70S native spliceosomes is very similar to that of 200S supraspliceosomes and they exhibit RNA splicing in vitro – hence the term native spliceosomes (Azubel et al., 2004; Azubel et al., 2006). The isolation of native spliceosomes and their relative stability enabled us to perform 3-D cryo-EM structural analysis by the single particle technique (Azubel et al., 2004). The 3-D structure was reconstructed from 9297 raw single-particle images at a resolution of 20 Å, which is the highest resolution obtained thus far for any splicing-active complex.

Figure 2 summarizes these structural studies. The structure (Figure 2, left and middle images) depicts the native spliceosome as an elongated globular particle made up of two distinct subunits. This finding is consistent with our previous STEM mass measurements, which revealed two major, equally populated, distinct groups of small particles with masses of 1.5 MDa and 3.1 MDa, which together add up to close to the 4.8 MDa mass of the native spliceosome (Müller et al., 1998). The two subunits are interconnected to each other leaving a tunnel in between, which is large enough to allow the pre-mRNA to pass through. The other side of the native spliceosome exposes a cavity that could provide a place to transiently store the pre-mRNA. Because RNA molecules are quite susceptible to degradation, and the number of bases that are involved in the splicing reaction represents only a small fraction of the entire pre-mRNA, the cavity might protect the part of the pre-mRNA that is not directly involved in the splicing reaction from nonspecific degradation.

The two subunits vary also with respect to the distribution of high densities within the native spliceosome. Because RNA is denser than protein, the localization of regions of high density can provide some information about internal organization of RNA and protein components. We have shown that the large subunit is a suitable candidate to accommodate the penta-snRNP, as the high density regions were found on the side of the native spliceosome exposing a cavity that could provide a place to transiently store the pre-mRNA. The smaller subunit is less dense, which might be consistent with the lack of proteins that are specifically localized in this subunit. Because RNA is denser than protein, the localization of regions of high density can provide some information about internal organization of RNA and protein components. We have shown that the small subunit is a suitable candidate to accommodate the penta-snRNP, as the high density regions were found on the side of the native spliceosome exposing a cavity that could provide a place to transiently store the pre-mRNA. The smaller subunit is less dense, which might be consistent with the lack of proteins that are specifically localized in this subunit.
in which the small subunits, which are in the center of the supraspliceosome, are related to one another by a right angle forming a four-fold pattern (Cohen-Krausz et al., 2007). These studies culminated in an integrated working model (Figure 3) that can account for alternative splicing in general and for alternative 5’ splice site selection in particular (see below).

**Alternative 5’ splice site selection**

The remarkable accuracy and specificity of the splicing reaction stems from an array of cis and trans-acting factors. The cis-acting elements include the 5’ and 3’ splice sites (SSs), a branch point, a poly-pyrimidine track, and splicing enhancer and silencer sequences. The trans-acting factors include the five spliceosomal uridine-rich small nuclear ribonucleoprotein particles (snRNPs); several non-snRNP proteins, such as the SR-protein family; and hnRNP proteins.

A bioinformatic survey of EST data bases revealed that alternative 5’SS selection accounts for ~8% of the alternative splicing events that are conserved between the human and mouse genomes. This estimate is by far lower than expected because a survey carried out by Miriami et al. (2002) on 446 annotated genes revealed that sequences that conform to the 5’SS consensus are highly abundant in introns, yet, there is no evidence that such latent 5’SSs are selected for splicing. In this study, a total of 10490 latent 5’SSs were found within 1601 introns. It was also found that more than 95% of these intronic latent 5’SSs have an in frame stop codon between the authentic and the latent 5’SSs. Thus, utilization of such a latent 5’SS could potentially introduce a premature termination codon (PTC) in the mature mRNA. Such PTC-carrying mRNAs could be harmful to the cell by producing a truncated protein, which could function in a deleterious way (e.g. loss-of-function; dominant-negative). Following this observation we hypothesized that selection of latent 5’SSs is suppressed such that the introduction of PTCs into mRNAs is avoided. Hence, the phenomenon was termed stop codon mediated Suppression Of Splicing (SOS).

In order to validate this hypothesis, Li et al. (2002) carried out an exhaustive mutational analysis on two model genes: CAD (abbreviated for the multifunctional enzyme carbamoyl-phosphate synthetase, aspartate trans-carbamylase, and dihydro-orotase) and IDUA (abbreviated for α-L-iduronidase). In these analyses latent splicing was elicited from mini-gene constructs only when all in frame PTCs were eliminated either by direct mutations or by frame shifting. These experiments have specifically implicated the PTCs as the causal factor for the suppression of splicing from the latent 5’SS. Because a stop codon is inherently defined by an open reading frame (ORF), the SOS machinery needs a start point to register and establish the reading frame for a proper identification of a PTC. To this end we have whether the translation start codons (AUG sequences), and specifically the first AUG, affect the SOS phenomenon. Transfections with CAD constructs wherein start codons were mutated, elicited latent splicing, thus showing that AUG sequences are required to sustain the SOS mechanism (Kamhi et al., 2006). Importantly, although the SOS mechanism is dependent on the establishment of an ORF, this mechanism is fully active when protein translation, or even the pioneer round of translation, is inhibited (Kamhi et al., 2006). These and other published data (Li et al., 2002; Wachtel et al., 2004), portray the SOS mechanism as an RNA surveillance mechanism that is distinct from the nonsense mediated mRNA decay (NMD) pathway.

New evidences from our group now show that the initiator tRNA (ini-tRNA) may act, in a manner that is independent on its role in protein translation, as an...
SOS trans-acting-factor that marks the AUG to establish a reading frame. We tested this possibility using gene constructs harboring a mutation in their first ATG. As expected, SOS for pre-mRNAs expressed from these constructs was abrogated and they exhibited latent splicing. However, SOS could be rescued when each of the ATG mutants was cotransfected with a mutated ini-tRNA construct in which the anti-codon was mutated to complement the respective ATG mutation. This was demonstrated on two minigene constructs, CAD and IDUA, each carrying either ATG to ACG mutation or a ATG to AAA mutation. Importantly, cotransfection of the ATG mutant with an ini-tRNA mutant harboring a non-complementary anti-codon mutation, did not affect the level of latent RNA. It should be noted that the effect of the ini-tRNA on SOS is not dependent of translation, because the rescue of the AUG phenotype by the complementary ini-tRNA mutant was maintained in cells where translation initiation had been inhibited. Moreover, we were able to show that the mutated forms of the ini-tRNA are not charged with amino acid, suggesting that only the RNA moiety, of the tRNA is required for SOS. Consistent with this finding, we also show that ini-tRNA, but neither the elongator methionyl-tRNA nor lysyl-tRNA, fractionated with supraspliceosomes and native spliceosomes, suggesting a possible involvement of ini-tRNA in splicing regulation.

Our study present evidence for the recognition of ORFs in the cell nucleus prior to RNA splicing, and points at a possible connection between the maintenance of ORF and splice site selection. A working model that accounts for this connection is shown in Figure 4.

**Selected publications**


