CRACKING THE CODE OF 3'SS SELECTION IN *S.cerevisiae*

Markus Meyer

DOCTORAL THESIS UPF

Barcelona, January 2010

Thesis director

Dr. Josep Vilardell (Gene Regulation/CRG)

UNIVERSITAT POMPEU FABRA

Yikes, I m practically shaking at the thought of writing the acknowledgments. The thing is, most of this thesis will be read only by the courageous few who have agreed to sit on my tribunal. Because frankly, who really cares about 3' splice site selection in *Saccharomyces cerevisiae*? But this part is different. It will be scrutinized and sounded for a mention. At least this is what I imagine, because that's about how far I got with my friends' theses.

Many things have happened since I got to Barcelona, and it is a great feeling to finally having reached a goal that I had set myself long ago. But the best part is that I am coming out of this adventure with incredible memories and many new friends. I won't remember all the pesky mutants I have picked and the frustration of doing countless primer extensions or IPs (I hate IPs). The real value of the last few years lies in the people I've met and in the places I've spent time at.

We've had heaps of fun at the CRG. This is thanks to the many brilliant people who have crowded this place over the years. Some of you have become close friends and have made a real difference. Together we've blown off steam and gone wild at beer sessions. We've gone off to the countryside for calçotadas and swished down the Pyrenean slopes. Sometimes we just laughed for no particular reason in a corner. All the frustrations accumulated because of misbehaving yeast were forgotten in those moments of fun, and such respites have enabled me to go on and finish this thesis.

Outside of the CRG, I have discovered a vibrant and vivid city that is also jammed with fantastic people. I've lived with some of them. We've also gone to concerts, celebrated the New Year, la Mercè and Sant Joan. Together we strode down the streets of Barcelona in search of the perfect place for a few drinks, and found many. They've tolerated my eccentricities, and the only thing I have imported from Switzerland: cheese fondue. On one occasion we even ate one on a boiling day in the middle of July. It is heartening to know that all these memories, and many more that I cannot mention here for fear of undermining the seriousness of this work, will accompany me in my future peregrinations.

As most of you have noticed, I've been popping over to Berlin quite a lot in the last year or so. And I didn't make the trips just for Glühwein and Currywurst. I have met a lot of sensational people there, but one of them in particular made even airport security and cattle-like transport tolerable.

Meeting so many people has not made me forget my family and friends who were around before I embarked on this journey. They have always been supportive, even though they are usually quite nervous to find out what I will come up with next. I thank them for their patience. I've been terrible at keeping in contact, and I apologize for this.

I thank every one of you for accompanying me during this stage of my life and for having made it all so memorable. My gaze is now set on a new and exciting adventure. I cannot imagine anything better than starting afresh in a new fantastic place with someone truly extraordinary. Incredible as it sounds, this is what I am throwing myself into, and I couldn't be happier.

INDEX

SUMMARY

English

The informational content of 3’ splice sites is low and the mechanisms whereby they are selected are not clear. Here we enunciate a set of rules that govern their selection. For many introns, secondary structures are a key factor, because they occlude alternative 3’ss from the spliceosome and reduce the effective distance between the BS and the 3’ss to a maximum of 45 nucleotides. Further alternative 3’ss are disregarded by the spliceosome because they lie at 9 nucleotides or less from the branch site, or because they are weak splice sites. With these rules, we are able to explain the splicing pattern of the vast majority of introns in *Saccharomyces cerevisiae*.

When in excess, L30 blocks the splicing of its own transcript by interfering with a critical rearrangement that is required for the proper recognition of the intron 3' end, and thus for splicing to proceed. We show that the protein Cbp80 has a role in promoting this rearrangement and therefore antagonizes splicing regulation by L30.

Castellano

Tanto la información que define el sitio de splicing 3’ como los mecanismos de selección del mismo son poco conocidos. En este trabajo, proponemos una serie de reglas que gobiernan esta selección. Las estructuras secundarias son claves en el caso de muchos intrones, porque son capaces de ocultar sitios de splicing alternativos 3’ al spliceosoma, y además reducen la distancia efectiva entre el punto de ramificación y el sitio de splicing 3’ a un máximo de 45 nucleotidos. Otros sitios de splicing alternativo 3’ no son considerados por el spliceosoma como tales porque se encuentran a 9 nucleotidos o menos del punto de ramificación, o porque son sitios de splicing débiles. Con estas reglas somos capaces de explicar el splicing de la mayoría de intrones de *Saccharomyces cerevisiae.*

El exceso de proteína L30 bloquea el splicing de su propio tránscrito porque interfiere con la reorganización necesaria para el correcto reconocimiento del 3’ final del intrón, y por tanto de su splicing. Demostramos que la proteína Cbp80 está implicada en promover esta reorganización y que por tanto antagoniza la regulación del splicing por L30.

KEY WORDS

Splicing

Spliceosome

Pre-mRNA

Intron

mRNA

5’ splice site

3’ splice site

Branch site

Selection

Secondary structure

Splicing regulation

L30

RPL30

Cbp80

U1 snRNP

U2 snRNP

Rearrangement

ABBREVIATIONS

DNA – Deoxyribonucleic acid

RNA – Ribonucleic acid

mRNA – messenger RNA

pre-mRNA – premature mRNA

CBC – Cap binding complex

m7Gppp – 7-methyl guanosine

U snRNA – Uridine-rich small nuclear RNA

snRNP – Small nuclear ribonucleoprotein particle

5'ss – 5' splice site

3'ss – 3' splice site

BS – Branch site

Ppy – Polypyrimidine tract

Nt – Nucleotide

CC1 – Commitment complex 1

CC2 – Commitment complex 2

PS – Pre-spliceosome

SP – Spliceosome

IC – Inhibited complex

INTRODUCTION

1. Splicing in the context of gene expression

In eukaryotic cells, genetic information is stored in the nucleus in the form of DNA. The expression of this information is achieved through a series of processes that have mostly been studied as a succession of discrete stages through which the information is processed in a linear manner. After chromatin remodeling, RNA polymerase II transcribes the DNA into a messenger molecule termed pre-messenger RNA (pre-mRNA). The 5' end of this transcript acquires a m7GpppN cap (Gu and Lima, 2005) that protects it from nucleases. It also serves as a platform for binding of the cap binding complex (CBC) made up of the proteins Cbp80 and Cbp20. The pre-mRNA is then spliced, polyadenylated, in certain cases further edited, and exported to the cytoplasm where this now mature messenger RNA (mRNA) is finally translated into a protein by the ribosome. However, it is becoming increasingly apparent that transduction of genomic information is not simply a collection of sequential steps as previously thought, but that the different processes are intertwined and can occur at least partially concomitantly (**Fig. 1** and Hagiwara and Nojima, 2007; Orphanides and Reinberg, 2002; Proudfoot et al., 2002). For instance, the chromatin architecture was known to influence transcription, but it now appears to also affect the outcome of splicing (Schwartz et al., 2009; Tilgner et al., 2009). Similarly, the process of capping happens cotranscriptionally, and many splicing factors are also recruited to the nascent transcript that is still attached to the polymerase (Tardiff et al., 2006). The interdependence between the various stages of gene expression is also implied by the function of certain proteins at more than one stage. This is the case for Npl3 for example in yeast that is involved in chromatin remodeling, splicing and RNA export (Kress et al., 2008). More generally, SR proteins in higher eukaryotes have functions in all steps of gene expression (Zhong et al., 2009). Each step offers many possibilities for regulation and the coupling of the various events provides an elaborate regulatory network that allows to tightly control the output of genomic

information transduction. This interconnection also produces a highly responsive system that is able to integrate external cues to which gene expression has to adapt in order to ensure survival of the cell

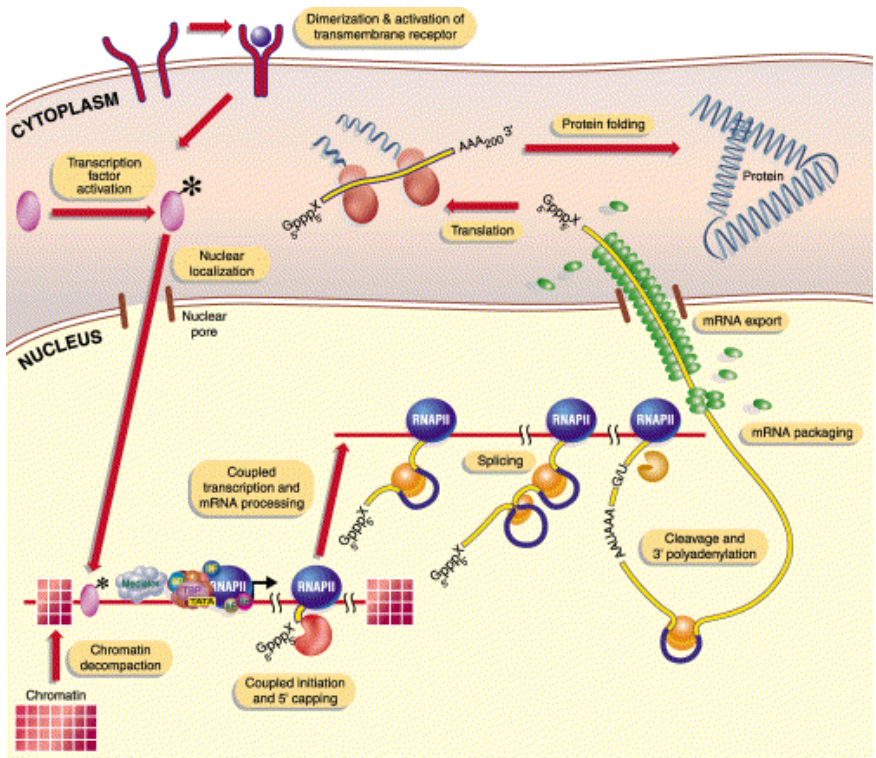


Figure 1. The contemporary view of gene expression *(from . (Orphanides and Reinberg, 2002))***.** Rather than a succession of independent steps, the process of gene expression is now viewed as a subdivision of a continuous process.

2. Splicing

Genomic information is in many cases not a linear thread the can be directly translated into a protein. Instead, the DNA template contains non-coding regions that have to be removed from the primary transcript in order to generate a functional message that will produce a protein upon translation. The non coding regions are called introns, and are interspersed among the coding regions, or exons. The presence of such intervening sequences is not

marginal and most metazoan genes are interrupted by one or several of these introns. In *S.cerevisiae,* 5% of the about 6000 annotated genes contain one intron, and a minority of those contain two (Spingola et al., 1999). However, these genes are highly transcribed and about 10'000 of the 38'000 mRNA molecules produced each hour in every cell are derived from intron-containing genes (Holstege et al., 1998).

Splicing is the process through which intervening introns are removed from the pre-mRNA and exons are joined together. Introns are characterized by several *cis* elements that define their boundaries and allow the cellular machinery by which they are excised to recognize them. These elements are rather degenerate in metazoans but are well defined in *S.cerevisiae*. In the latter, the 5'exon-intron junction, or 5' splice site (5'ss) is characterized by the consensus sequence GUAUGU (the first nucleotide of the intron is underlined, see **Fig. 2**). At the other brink of the intron lies the 3' splice site (3'ss) that is quite enigmatically defined only by the very minimal splicing cue HAG ([U/C/A]AG, the last nucleotide of the intron is underlined). Between the two extremities, often near the 3'ss, lies the branchpoint sequence (BS) UACUA**A**C (the branchpoint adenosine is highlighted) and in metazoans, the BS is followed by an extensive ploypyrimidine tract (Ppy tract), but this feature is less conserved in *S.cerevisiae*.

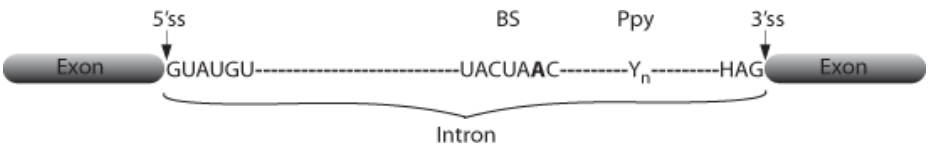


Figure 2. Schematic representation of conserved intronic sequences. Introns are delimited by the consensus sequences of the 5' splice site (5'ss) and 3' splice site (3'ss). Additional sequences important for splicing are also shown; the branch site (BS, the branch site adenosine is shown in bold) and the pyrimidine-rich tract (Ppy).

2.a The splicing reaction

The chemistry of splicing consists of a set of two subsequent transesterification reactions. The removal of an intron is initiated by the nucleophilic attack of the 2' hydroxyl of the branch point adenosine on the phosphate of the 5'ss (**Fig. 3**). This first transesterification results in an intron branched by a 2'-5' phosphodiester bond (Konarska et al., 1985). The second part of the reaction entails a nucleophilic attack of the now free 3'hydroxyl of the 5'ss on the 5' phosphate of the 3'ss, which leads to joined exons and an excised intron in the form of a lariat (Konarska et al., 1985; Moore and Sharp, 1993).

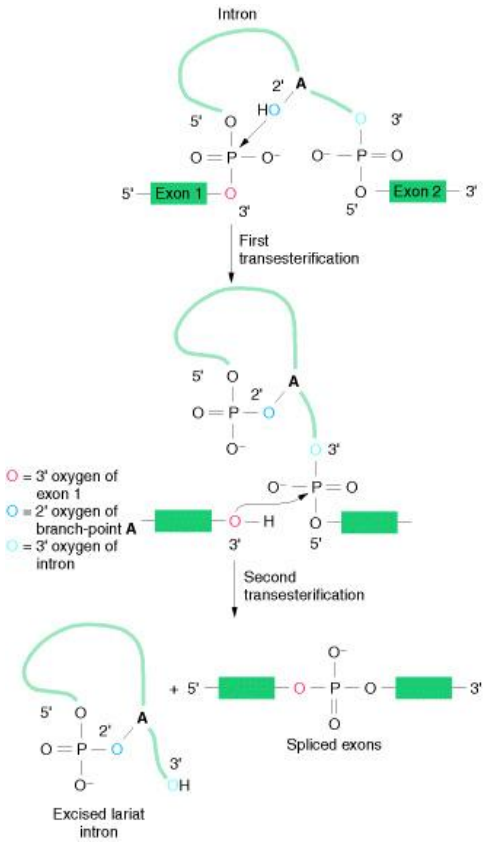


Figure 3. The catalytic steps of pre-mRNA splicing *(from Griffiths et al., 1999)***.** In the first transesterification reaction, the ester bond between the 5' phosphorus of the intron and the 3' oxygen of exon 1 is exchanged for an ester bond with the 2' oxygen of the branch site adenosine (**A**). In the second trasesterification reaction, the ester bond between the 5' phosphorus of exon 2 and the 3' oxygen of the intron is exchanged for an ester bond with the 3' oxygen of exon 1. The result of these transesterification reactions are an intron in the form of a lariat and joined exons.

3. The spliceosome

Several types of introns have been identified and classified into different groups based on their sequence features. Group II introns for example, are present in certain plants, lower eukaryotes and bacteria and are able to excise themselves autocatalytically from RNA molecules through the action of the ribozyme that is encoded within their sequence (Fedorova and Zingler, 2007; Toor et al., 2008). They are arguably the precursors of modern introns that have lost the ability of self-splicing and therefore rely on *trans* acting factors to be remove from the pre-mRNA. The cellular machinery that carries out this function is the spliceosome. It is largely conserved from yeast to human, both in basic components and design principal.

3.a The complexity of the spliceosome

The spliceosome is a complex multi-megadalton molecular machine (Nilsen, 2003). It is composed of 5 basic building blocks termed U snRNPs (Uridine-rich small nuclear ribonucleoprotein particles) and many auxiliary non-snRNP protein factors. Each of the U snRNPs (U1, U2, U4, U5 and U6) is composed of a short snRNA (short nuclear RNA) and several proteins, of which some are common core components of the U snRNPs and some are specific. The total number of proteins involved in splicing in humans has been estimated to between 170 and 300 proteins depending on the purification technique and the spliceosomal complexes studied (Jurica and Moore, 2003; Rappsilber et al., 2002; Wahl et al., 2009). This number is somewhat lower for the *S.cerevisiae* spliceosome and oscillates around the number of 100 proteins (Fabrizio et al., 2009). Interestingly, over 85% of the reduced set of proteins that compose the yeast spliceosome have a clear evolutionarily conserved counterpart in the human spliceosome (**Fig. 4**). Conversely, most of the proteins that copurify with the spliceosome in human but not in yeast do not have a conserved equivalent in the latter (Behzadnia et al., 2007; Bessonov et al., 2008). This

suggests that the yeast splicing machinery is the evolutionarily conserved core design of the more complicated human spliceosome.

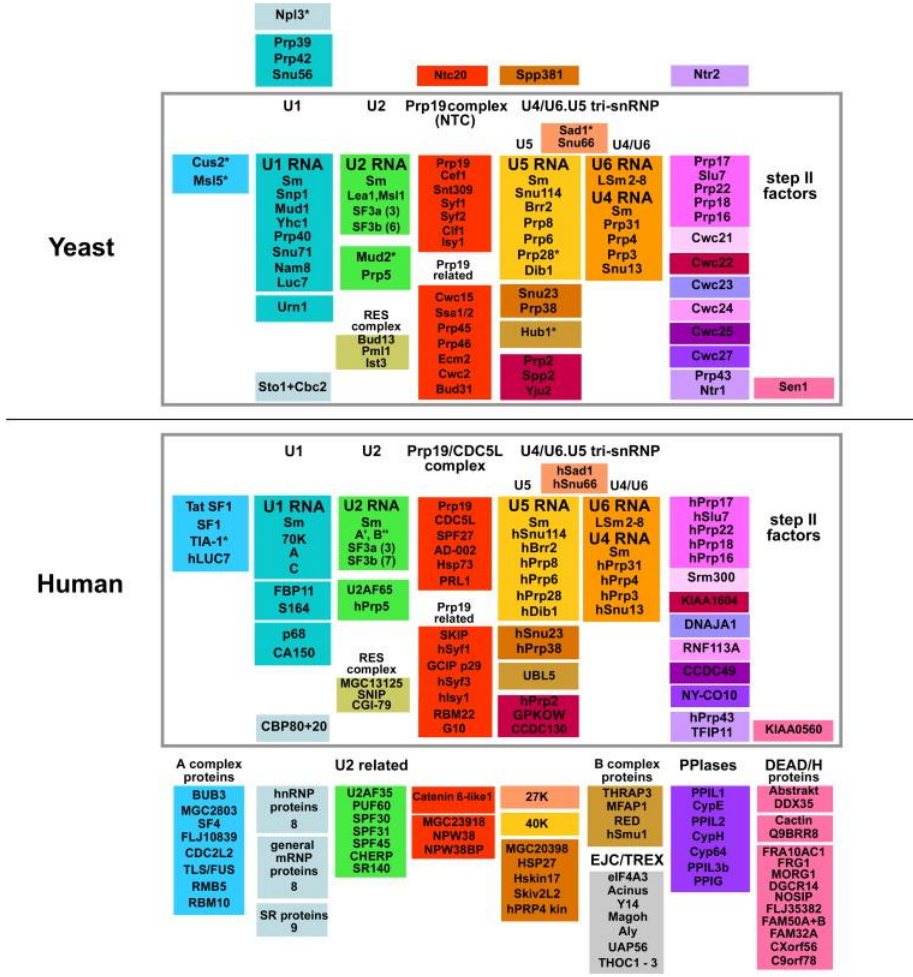


Figure 4. Proteomics of yeast and human spliceosomes *(from Fabrizio et al., 2009).* Proteins associated with the yeast spliceosome are shown in the upper rectangle. Proteins located above do not have a human counterpart. In the lower rectangle are indicated the proteins associated with purified spliceosomes. Under this same rectangle are indicated proteins that are purified with the human spliceosome but do not have a clear yeast counterpart.

The staggering complexity of the spliceosome can seem superfluous when considering the fairly simple biochemical reaction it catalyzes, especially since certain introns (group I and II) are able to perform this reaction without the help of external factors. Furthermore, the fact that some RNA structures within the spliceosome bear striking resemblances to catalytically crucial structures in the self-splicing Group II introns makes this discrepancy even more astonishing (Collins and Guthrie, 2000; Nilsen, 1998). Indeed, it suggests that the splicing mechanisms of self-splicing introns and of those assisted by the spliceosome are similar (Valadkhan and Manley, 2001). To council these observation, it has been speculated that the high degree of complexity of the spliceosome is due to several factors. Firstly, the spliceosome has to be able to recognize and process a great variety of introns that tremendously differ in size and sequence, which requires a high degree of flexibility. Secondly, the spliceosome has to ensure a prominent degree of accuracy since a shift of only one nucleotide in the splice site choice introduces a frame shift, thereby dramatically changing the protein produced from the mRNA. Therefore, every splicing signal is inspected multiple times by various factors. In addition, the accuracy is also checked by multiple proofreading steps along the splicing path (see below). Third, splicing is not a constitutive reaction that happens at the same rate under all conditions. It is a regulated process that is influenced by external events. In *S.cerevisiae*, for example, it was shown that amino acid starvation or changes in carbon source had a direct effect on splicing of certain groups of introns (Pleiss et al., 2007). This implies a certain degree of spliceosomal flexibility that can be regulated in response to external cues. This is even more true in the case of metazoans were alternative splicing that has the potential to be regulated, occurs on most transcripts (Wang et al., 2008a).

3.b Enzyme or ribozyme?

The spliceosome is composed both of proteins and RNAs. It is therefore legitimate to wonder which of the two constituents carries out the actual catalysis. Indeed, in most RNPs, although both components are essential for function, one primarily serves as a scaffolding or guiding entity, while the other facilitates the biochemical reaction. The extensive mechanistic and structural similarities of the spliceosome to self-splicing Group II introns (Valadkhan, 2007) long lead to believe that the spliceosome is a ribozyme helped for accuracy, speed and flexibility by proteins. However, the crystal structure of a 250 amino acid domain of Prp8, a U5 snRNP protein located at the catalytic center of the spliceosome (Pena et al., 2008; Ritchie et al., 2008; Yang et al., 2008), cast a doubt upon this quasi certainty. The structure revealed an RNase H domain. This raises the possibility that the spliceosome, rather than being a ribozyme, is actually a RNP enzyme. This uncertainty will remain until the structures of the entire Prp8 protein and the active site of the spliceosome have been resolved.

4. The splicing cycle

One of the fundamental questions in the field of splicing was for some time whether the spliceosome is a pre-assembled complex such as the ribosome, or if it assembles *de novo* onto each pre-mRNA substrate. The first model was supported by the purification under low salt conditions of a holospliceosome, a penta-snRNP complex that contains all five U snRNPs (Stevens et al., 2002). This piece of evidence suggested that the spliceosome engages its substrate as a multi-snRNP complex. However this model has many detractors, because of findings *in vitro*, and especially since the advent of techniques that enable to follow spliceosome assembly onto the nascent transcript *in vivo* (Gornemann et al., 2005; Lacadie and Rosbash, 2005; Tardiff et al., 2006). The emerging picture is one of a spliceosome that assembles onto the substrate in an ordered piece-by-piece assembly. The U1

snRNP is the first to engage the pre-mRNA and is necessary for the recruitment of all subsequent U snRNPs. This first step is then followed by the arrival of U2 snRNP and finally by the engagement of the U4/U6·U5 tri-snRNP.

4.a Spliceosome assembly

The spliceosome is built in a stepwise manner onto the nascent transcript as inferred by *in vivo* studies. However, most of what is known about spliceosome assembly has been determined *in vitro*. Indeed, the fact that the assembly is carried out in a stepwise manner, implies the formation of intermediate complexes, and they can be assembled, then resolved and visualized for example by native gel techniques (Pikielny et al., 1986) (**Fig. 5**). Such approaches have proven invaluable tools to understand the stepwise construction of the spliceosome.

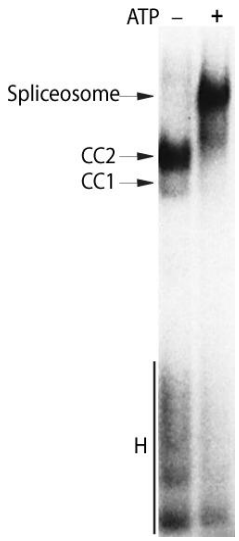


Figure 5. Yeast spliceosome assembly visualized by native gel *(adapted from Caspary and Seraphin, 1998)***.** *In vitro* assembly of the spliceosome visualized by native gel analysis shows that apart from a heterogeneous complex (H), the commitment complex 1 (CC1) and commitment complex 2 (CC2) are formed (first lane). Addition of ATP (second lane) triggers the formation of the spliceosome at the expense of the earlier complexes. The pre-spliceosome is not observed by this technique.

Native gel techniques, combined with proteomic analysis of purified complexes, *in vitro* splicing reactions and others, have lead to a comprehensive view of the splicing cycle. Components of the splicing machinery are recruited to the pre-mRNA and form a series of complexes that

will be detailed here. The yeast nomenclature is used, because an important part of our knowledge of spliceosome formation has been acquired from studies in *S.cerevisiae*.

4.b Commitment complexes

At the onset of spliceosome assembly, the pre-mRNA is engaged by the U1 snRNP particle through a base-pairing interaction between the 5' end of the U1 snRNA and the 5'ss (Seraphin et al., 1988; Siliciano and Guthrie, 1988; Zhuang and Weiner, 1986). The CBC has been shown to facilitate this recruitment through a protein-protein interaction with U1 snRNP (Lewis et al., 1996; Zhang and Rosbash, 1999) and is particularly important for introns with poor complementarity to U1 snRNA (Colot et al., 1996; Fortes et al., 1999a; Fortes et al., 1999b; Lewis et al., 1996). This initial contact commits the substrate to the splicing pathway and the U1 snRNP-pre-mRNA complex is therefore designated as the commitment complex 1 (CC1, **Fig. 6**). The formation of this first complex is followed by the assembly of the commitment complex 2 (CC2). In this step, the 3'end of the intron is identified by BBP (Berglund et al., 1997) and Mud2 (Abovich et al., 1994) possibly as a preformed heterodimer (Wang et al., 2008b). BBP is the yeast homolog of human SF1 and binds the flanking nucleotides of the branch point adenosine through its KH domain (Arning et al., 1996; Berglund et al., 1998; Kramer, 1992; Liu et al., 2001; Peled-Zehavi et al., 2001; Rain et al., 1998). A direct interaction has been shown between BBP and the U1 snRNP protein Prp40 (Abovich and Rosbash, 1997; Kao and Siliciano, 1996). This suggests that the extremities of the intron are bridged by a cross-intron bridging interaction to stabilize the CC2. The function of Mud2 is not clear, but its homology to the large subunit of the human U2AF heterodimer ($U2AF^{65}$) hints at a function. This protein interacts with the pyrimidine-rich sequence that often follows the BS in metazoan (Zamore and Green, 1989; Zamore et al., 1992), and is required for the subsequent U2 snRNP binding (Ruskin et al., 1988). Interestingly, the small subunit of the U2AF ($U2AF^{35}$) that directly

contacts the 3'ss in humans (Merendino et al., 1999; Wu et al., 1999) has no clear homolog in *S.cerevisiae*, and the 3'ss is altogether dispensable for the first catalytic step of splicing (Rymond and Rosbash, 1985).

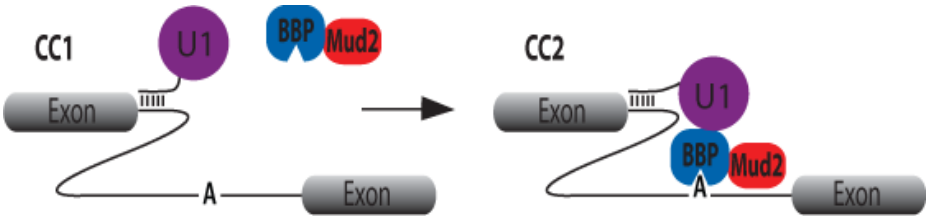


Figure 6. Formation of the commitment complexes. The pre-mRNA is first engaged by the U1 snRNP to form the commitment complex 1 (CC1). Subsequent identification of the 3' end of the intron by BBP and Mud2 leads to the formation of commitment complex 2 (CC2) that is stabilized by a cross-intron bridging interaction.

4.c The pre-spliceosome

Formation of the pre-spliceosome entails the displacement of BBP and Mud2, as well as the binding of U2 snRNP to the BS (Parker et al., 1987; **Fig. 7**). The transition from CC2 to pre-spliceosome is mediated by DExD/H-box proteins, which explains why, contrarily to the previous stages, this step is ATP-dependent.

Mud2 and BBP have to be removed in order for U2 snRNA to base-pair with the branch site sequence. The removal of Mud2, and possibly also BBP, is thought to be achieved through the ATP-dependent action of the DExD/H-box protein Sub2. Indeed, deletion of *MUD2* suppresses the lethality of the *SUB2* deletion (Kistler and Guthrie, 2001), suggesting that Sub2 is dispensable when BBP is not stabilized on the BS by Mud2.

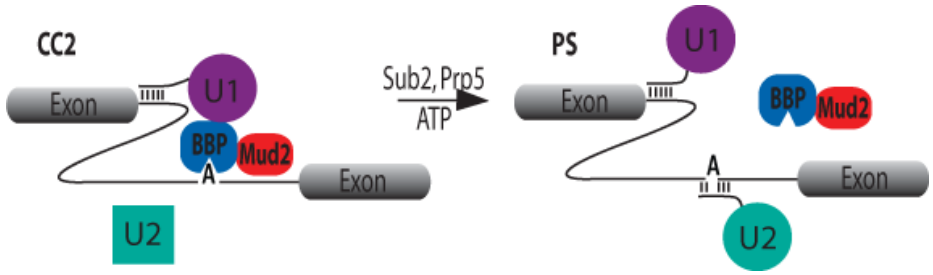


Figure 7. Pre-spliceosome formation. The transition from the CC2 to the pre-spliceosome (PS) results from the displacement of BBP and Mud2 and the base pairing of the U2 snRNA to the BS. This step is driven by the RNA helicases Sub2 and Prp5, which make it ATP-dependent.

The last requirement for U2 snRNP to dock on the branchpoint sequence is conformational. Indeed, the U2 snRNA can adopt two distinct conformations of which one is active, and one is not. The U2-stem IIc form is unsuitable for spliceosome assembly to proceed and has to be converted to the active U2-stem IIa conformation in which the branchpoint-binding sequence is rendered accessible (Hilliker et al., 2007; Perriman and Ares, 2007). The protein that drives this U2 snRNA activation is another ATPase, the DExD/H-box protein Prp5 (O'Day et al., 1996; Perriman and Ares, 2000; Perriman et al., 2003). The current model for this rearrangement is that the non-essential U2 snRNP protein Cus2 stabilizes the non-productive conformation. When Cus2 is removed by Prp5, the U2 snRNA changes conformation and spliceosome assembly proceeds. This model is supported by the fact that pre-spliceosome assembly becomes ATP-independent in the absence of Cus2 *in vitro*.

The adenosine residue of the BS itself is not base paired, but bulges out of the U2 snRNA-BS helix, which will allow its utilization as a nucleophile in the first catalytic reaction (Query et al., 1994).

4.d The full spliceosome

The last step of spliceosome assembly is the most complicated and the least understood. It involves the joining and displacement of many factors, along with numerous rearrangements that make it very labile. The assembly of this last complex starts with the recruitment of the tri-snRNP U4/U6·U5 to the pre-spliceosome (Konarska and Sharp, 1987), and will lead to the formation of the complex capable of carrying out the sequential transesterification reactions (**Fig. 8**). The CBC has been implicated in promoting the association of the tri-snRNP to the developing spliceosome (Gornemann et al., 2005; O'Mullane and Eperon, 1998).

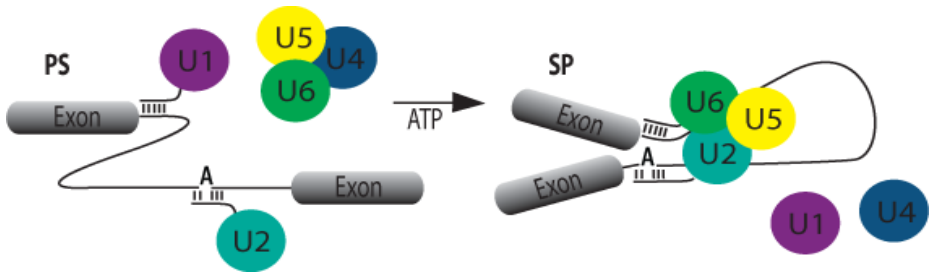


Figure 8. Spliceosome formation. The tri-snRNP U4/U6·U5 is added to the pre-spliceosome and the active spliceosome (SP) is formed upon several ATP driven rearrangements and release of U1 and U4 snRNPs.

The initial recruitment of the tri-snRNP triggers a cascade of events that result in the juxtaposition of the 5' and 3'ss and formation of the active catalytic core of the spliceosome. During these rearrangements, the snRNA-snRNA and pre-mRNA-snRNA interaction network is completely remodeled (**Fig. 9**). In the tri-snRNP, the U6 snRNA is base-paired to the U4 snRNA. This duplex is unwound (Lamond et al., 1988) and U6 snRNA base-pairs to the 5'ss in place of U1 snRNA (Wassarman and Steitz, 1992) as well as to the U2 snRNA, close to the BS duplex (Datta and Weiner, 1991; Hausner et al., 1990; Madhani and Guthrie, 1992; Wu and Manley, 1991). These rearrangements are conducive to the release of both U1 and U4 snRNPs. U5

snRNP remains and interacts with both exons, thereby serving as a platform on which the two exons are positioned for the second step of splicing (Newman and Norman, 1992; Sontheimer and Steitz, 1993; Wassarman and Steitz, 1992; Wyatt et al., 1992).

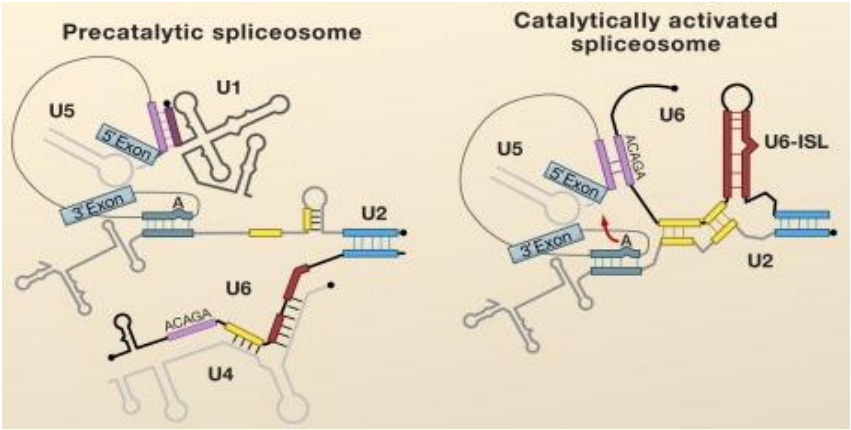


Figure 9. Spliceosomal RNA network *(from Wahl et al., 2009)***.** The RNA network that arises from tri-snRNP engaging the pre-spliceosome is shown on the left side. This network undergoes numerous rearrangements and results, after release of U1 and U4 snRNPs, in the active spliceosome as shown on the right. The critical base-pairing interactions are highlighted.

4.e Powering spliceosome assembly

The entire cascade is mediated by RNA helicases which offers an opportunity for a proofreading mechanism (see below). After pre-spliceosome formation that is driven by Sub2 and Prp5 as mentioned above, the formation of the spliceosome and subsequent steps are powered by at least six helicases: Prp28, Brr2, Prp2, Prp16, Prp22, and Prp43 (de la Cruz et al., 1999; Silverman et al., 2003; Staley and Guthrie, 1998, and see **Fig. 10** for the entire splicing cycle and the RNA helicases that drive it).

Prp28 facilitates the displacement of U1 snRNP by destabilizing its interaction with the 5'ss and thus clearing the way for U6 snRNA binding

(Chen et al., 2001; Staley and Guthrie, 1999). It may facilitate the U1 to U6 snRNA transition further by participating in rearrangements in the U6 snRNA conformation prior to docking (Staley and Guthrie, 1999; Strauss and Guthrie, 1991).

The release of the U4 snRNP is likely mediated by unwinding of the U4/U6 snRNA base pairing by U5 snRNP protein Brr2 (Raghunathan and Guthrie, 1998). This helicase is furher involved after the splicing reaction has been completed. It possibly drives spliceosome disassembly by destabilizing the U2/U6 snRNA duplex (Xu et al., 1996). Both of the Brr2 activities are either directly or indirectly regulated by the only GTPase of the spliceosome, Snu114 (Bartels et al., 2002; Brenner and Guthrie, 2005; Small et al., 2006).

Prp2 is thought to cause a structural reorganization of the spliceosome that makes it active for the first catalytic reaction (Kim and Lin, 1996; Roy et al., 1995). Similarly, Prp16 may have the same function to drive the second trasesterification reaction, possibly by promoting reformation of stem IIa in the U2 snRNA (Hilliker et al., 2007; Perriman and Ares, 2007; Schwer and Guthrie, 1991, 1992).

4.f At the end of the splicing cycle

After the two transesterification reactions, splicing is completed, and the spliceosome is dismantled. Prp22 and Prp43 are involved in facilitating the release of respectively the spliced mRNA and the lariat (Arenas and Abelson, 1997; Company et al., 1991; Schwer and Gross, 1998). The proteins Ntr1 and Ntr2 have been shown to assist Prp43 in its role by targeting it to its site of function (Boon et al., 2006; Tsai et al., 2005). The splicing cycle is closed by the recycling of the different components of the spliceosome to perform further rounds of splicing. This is made possible in part by the RNA helicase Prp24 that helps to restore the original conformation of the tri-snRNP by promoting U4/U6 base pairing (Raghunathan and Guthrie, 1998).

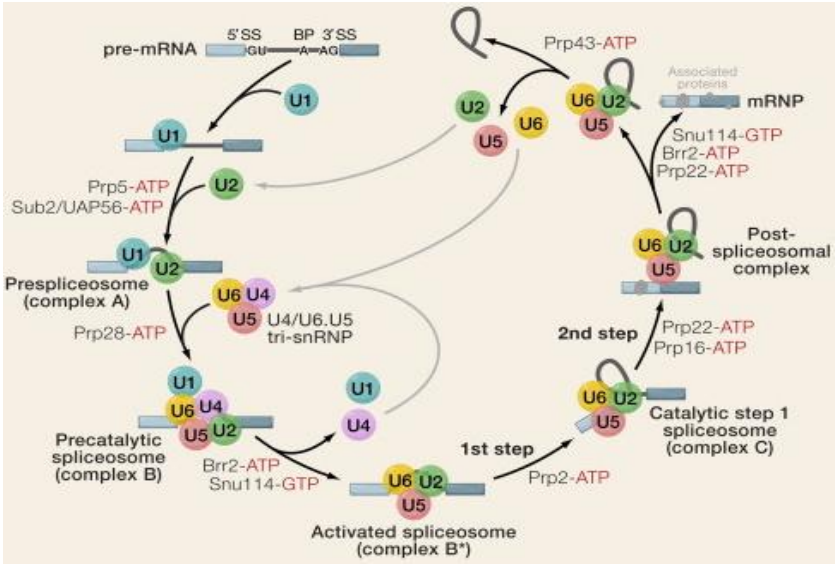


Figure 10. Splicing cycle and RNA helicases *(from Wahl et al., 2009).* All the complexes of spliceosome assembly that can be resolved biochemically are shown here. The RNA helicases that hydrolyse ATP, or GTP in the case of Snu114, to power the splicing cycle are indicated at the step(s) where they have been shown to operate.

5. Selection of the 5' and 3'ss

The informational content of the 5'ss is rather high due to its relatively long consensus sequence. A certain degree of degeneracy is even permitted, with no apparent defect in splicing efficiency. The selection of the 5'ss is therefore rather straightforward and is achieved through base-pairing to the U1 snRNA during CC1 formation (Lerner et al., 1980; Ruby and Abelson, 1988; Seraphin and Rosbash, 1989; Zhuang and Weiner, 1986) with the help of the protein U1C (Pomeranz Krummel et al., 2009). As seen above, the 5'ss is then recognized again by base-pairing to U6 snRNA upon the engagement of the tri-snRNP.

The selection of the 3'ss is more enigmatic because of the scarcity of information contained in its very short sequence HAG. The 3'ss is even dispensable for the first catalytic step of splicing and is only absolutely

required for the second step (Rymond and Rosbash, 1985). A 5' to 3' scanning mechanisms from the BS or the polypyrimidine tract onward that chooses the first AG dinucleotide downstream of the branch site has been proposed in human (Smith et al., 1993; Smith et al., 1989). However, this mechanism has fallen out of favor, largely because of the observation that the first AG is in many cases not the splice site that is preferred by the spliceosome. This is the case for example in the *S.cerevisiae* intron of *VMA10* where the 3'ss is preceded by two HAGs (**Fig. 11**). The question therefore remains as to how the 3'ss is identified. Several elements have been put forward in order to explain the 3'ss choice. The splicing machinery can be positively influenced to favor a second AG after the BS by the introduction of a poly-U stretch in its 5' vicinity (Patterson and Guthrie, 1991). The 5'ss sequence (Goguel and Rosbash, 1993) and the sequences surrounding the intron (Crotti and Horowitz, 2009) have also been shown to have a certain influence on 3'ss selection. Finally, the distance from the BS to the AG is a limiting factor for splicing (Cellini et al., 1986; Luukkonen and Seraphin, 1997). Several proteins have also been shown to have a function in the second step of splicing, and may be involved in 3'ss selection. The best example of such a protein is Slu7. It is required for the second step of splicing both *in vivo* and *in vitro* (Ansari and Schwer, 1995; Frank and Guthrie, 1992; Frank et al., 1992; Jones et al., 1995), and at this stage, cross links to the 3'ss (Umen and Guthrie, 1995). A mutant allele, *slu7-1*, causes a selective defect in the usage of a 3'ss that is located at over 12 nucleotides downstream of the BS (Frank and Guthrie, 1992).

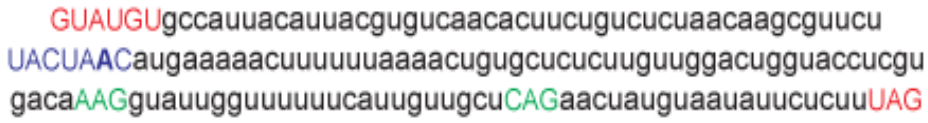


Figure 11. *VMA10* intron. The 162 nucleotides of the intron are shown here. The extremities (5' and 3'ss) are depicted in red, the BS in blue (the BS adenosine is in bold), and the alternative HAGs situated between the BS and the 3'ss are in green.

However, in spite of many efforts, the mechanism of 3'ss selection has not yet been solved. Indeed, the different effects described here are not general rules that allow a global understanding of 3'ss selection, especially since most of them have been observed for only one intron. Matters are made worse by further complications. In plants for example, it has been observed that a 3'ss sequestered in a hairpin could only act as a splicing acceptor in the presence of a downstream 3'ss (Liu et al., 1995).

5.a Splicing fidelity

Mis-splicing events lead to altered genetic messages that are potentially deleterious for the cell and can cause severe diseases (Cooper et al., 2009). The spliceosome goes to great length to avoid such messages. Reports of mutations in splicing factors that result in splicing of suboptimal substrates that would normally be discarded, point towards a mechanism of kinetic proofreading (Burgess and Guthrie, 1993; Konarska et al., 2006; Liu et al., 2007; Mayas et al., 2006; Query and Konarska, 2006; Umen and Guthrie, 1996; Villa and Guthrie, 2005; Xu and Query, 2007). Indeed, mutations with such an effect are often found in the ATPase domains of the DExD/H-box ATPases that drive spliceosome assembly and the splicing reaction itself. According to this model, splicing progresses only if completion of a particular step precedes ATP hydrolysis that drives the next step. In case of a suboptimal substrate that slows down spliceosome assembly or catalysis, the correct chain of events is disrupted. Consequently, the suboptimal substrate is discarded because the time imparted to a particular stage, which is marked by the hydrolysis of the ATP that powers the next step, has been over-reached (**Fig. 12**). Mutations in the ATPase domains of the RNA helicases therefore allow for splicing of a suboptimal substrate by lowering the rate of ATP hydrolysis and thus allowing for more time for a particular reaction.

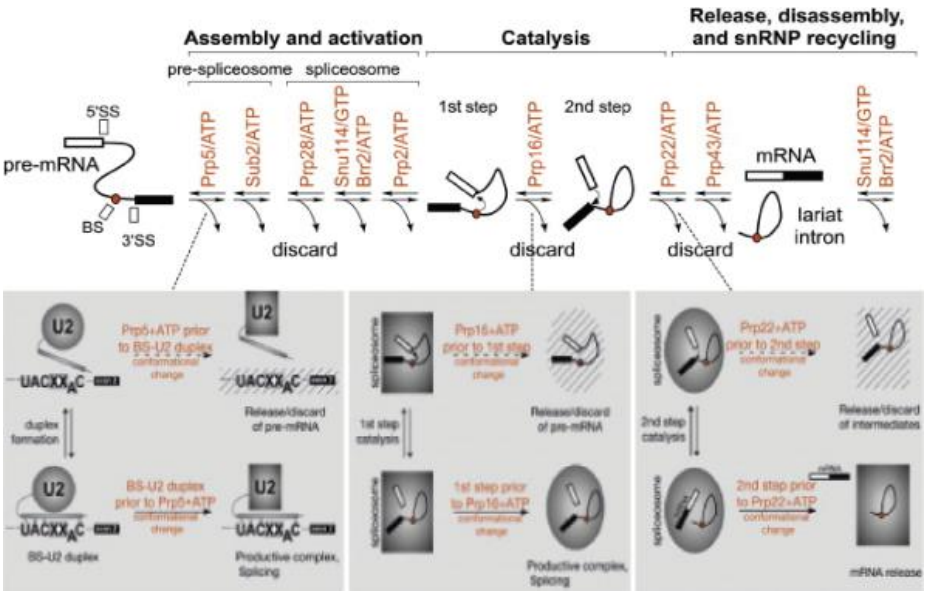


Figure 12. Kinetic proofreading along the splicing cycle *(adapted from Smith et al., 2008)***.** Every step at which the splicing fidelity can be checked by the kinetic proofreading mechanism is indicated. The best characterized examples are shown in the lower panels.

A compelling case is the one of Prp5 that has been shown to proofread the BS-U2 snRNA interaction (Xu and Query, 2007). A conformational change of the U2 snRNP mediated by Prp5 ATP hydrolysis has to occur after BS-U2 snRNA duplex formation for splicing to go forward. If this is not the case, the substrate is discarded. In case of a wt Prp5, the substrate has to be optimal to proceed along the splicing path, but BS mutations can be superseded by mutations in the ATPase domain of Prp5. The level of suppression by *prp5* alleles correlates inversely with their ATPase activity. Similarly, the ATPase activity of Prp5 from organisms in which the BS is less highly conserved than in *S.cerevisiae* is lower.

Although the best studied instances of kinetic proofreading are Prp5, Prp16 and Prp22, it is very probable that each ATP-driven step offers an opportunity to check splicing fidelity.

6. Alternative splicing

Alternative splicing enables an organism to dramatically increase its coding capacity by generating several transcript, and ultimately proteins, from a single genomic locus (reviewed in (Chen and Manley, 2009)). It also serves as a crucial expression regulation tool by introduction of alternative regulatory elements and by coupling to the nonsense mediated decay pathway that degrades premature stop codon-containing transcripts (Lareau et al., 2007). Alternative splicing events include exon skipping, alternative 5' or 3'ss usage and intron retention. The combination of these events produces a wide array of transcripts with a variety of functions and expression patterns.

Although a few cases of alternative splicing have been reported in yeast (Davis et al., 2000; Juneau et al., 2009), their occurrence is rather marginal (Yassour et al., 2009). This scarcity is in stark contrast to the abundance of alternative transcripts found in metazoans. In human cells for example, no less than 92 to 94% of transcripts undergo at least one alternative splicing event (Wang et al., 2008a). Such events often arise from suboptimal splicing signals or non-ideal intron and exon lengths and require a high degree of flexibility on the part of the spliceosome. Regulation of alternative splicing is carried out mostly by *trans* acting factors that recognize auxiliary sequences in the pre-mRNA that serve as splicing enhancers or silencers. One of the main families of proteins responsible for regulation are SR proteins, which have no clear homologues in *S.cerevisiae*. Modulation of alternative splicing can be achieved in several ways such as by hindrance or promotion of communication between spliceosomal components, or by occlusion of splicing cues.

7. Splicing regulation in *S.cerevisiae*

The first part of this section is presented in the form of a publication in which we summarized what is known about the regulation of splicing in *S.cerevisiae*. It is followed by a section that specifically describes splicing regulation of the *RPL30* transcript by the L30 protein.

7.a Publication 1

The quest for a message: budding yeast, a model organism to study the control of pre-mRNA splicing

Markus Meyer and Josep Vilardell

Meyer M, Vilardell J. The quest for a message: budding yeast, a model organism to study the control of pre-mRNA splicing. Brief Funct Genomic Proteomic. 2009; 8(1): 60-7.

7.b Splicing regulation by L30

L30 is an essential ribosomal protein that is encoded by the *RPL30* gene. The *RPL30* transcript adopts a secondary structure known as a kink-turn (White et al., 2004) that is similar to the binding site of the L30 protein in the ribosome (**Fig. 13**, Halic et al., 2005). This structure contains the first nucleotides of the 5'ss and is bound and stabilized by L30 when the protein is produced in excess and cannot be incorporated into a ribosome. This binding results in blockage of full spliceosome assembly (Eng and Warner, 1991; Vilardell and Warner, 1994) and translation (Dabeva and Warner, 1993). The ability of L30 to regulate the level of *RPL30* mRNA is essential to sustain optimal biological fitness (Li et al., 1996), and suggests a high toxicity of L30 over-expression.

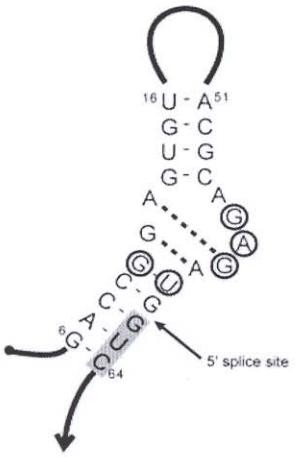


Figure 13. The kink-turn of *RPL30* pre-mRNA *(from Macias et al., 2008).* Numbers are indicated relatively to the start of transcription and the cap is depicted by a solid circle. The structure includes the first nucleotides of the 5'ss (grey box) and the nucleotides that have been implicated in the interaction with L30 by x ray crystallography are circled (Chao and Williamson, 2004).

Since the 5'ss is partially occluded by the secondary structure, it was initially hypothesized that splicing regulation was achieved through blockage of U1 snRNA docking onto the 5'ss (Eng and Warner, 1991). However, the fact that U1 snRNP is present in the inhibited complex (IC) in conjunction to L30, called for a revision of this hypothesis (Vilardell et al., 2000; Vilardell and Warner, 1994). This finding was confirmed by ChIP analysis in which a stabilization of U1 snRNP onto the nascent transcript can be observed when L30 is over-expressed (**Fig. 14**). The same data however indicates that the regulation is likely to occur between the steps of CC2 formation and pre-spliceosome assembly, because contrarily to U1 snRNP, U2 snRNP is prevented from engaging with the nascent transcript.

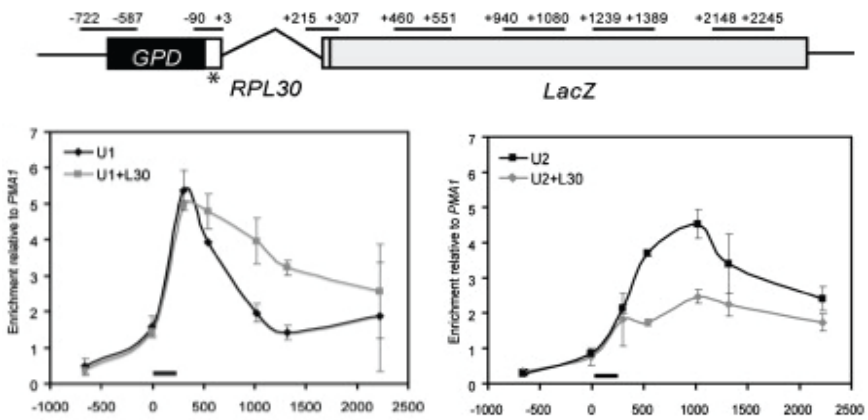


Figure 14. Co-transcriptional recruitment of U1 and U2 snRNPs onto the *RPL30*-LacZ transcript *(adapted from Macias et al., 2008)***.** The ChIP profile of U1 and U2 snRNPs are shown both in wt and under L30 over-expression conditions. The recruitment of U2 snRNP to the nascent transcript is impaired by over-expression of L30, whereas U1 snRNP is stabilized onto the transcript in this same condition.

L30 does not affect the binding of BBP or Mud2 to the intron, neither does regulation depend on the activity of the Sub2 and Prp5 or the presence of Cus2 (Macias et al., 2008 and unpublished results from Sara Macías and Josep Vilardell). The lack of activity of all these factors in regulation of splicing by

L30 is puzzling. It suggests that U2 snRNP is prevented from docking to the BS by the blockage of an internal rearrangement of the nascent spliceosome required for U2 snRNP to latch onto the intron. The nature of this rearrangement is not clear for the moment, and the means by which L30 hinders it remains to be uncovered (**Fig. 15**).

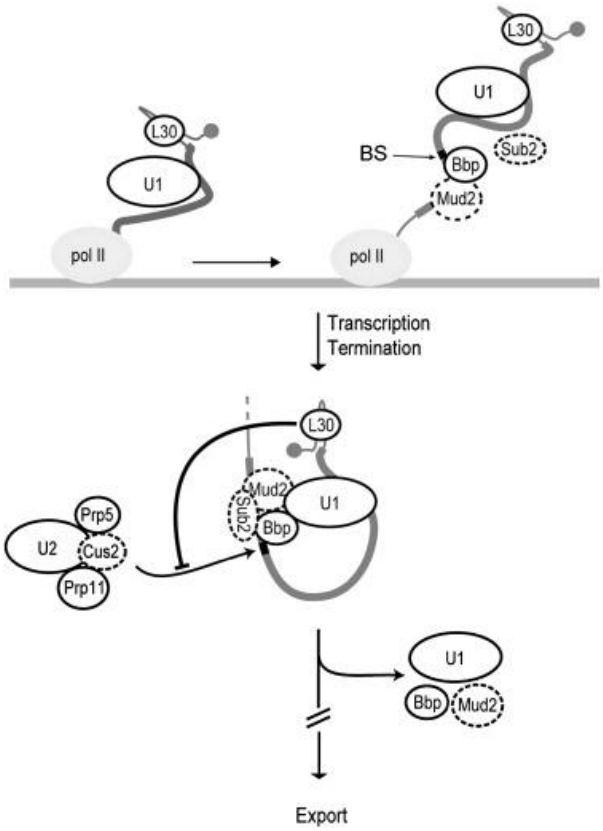


Figure 15. Model of splicing inhibition of the *RPL30* transcript by L30 *(from (Macias et al., 2008))***.**

OBJECTIVES

1. Uncover the mechanisms that govern 3’ss selection in the introns of *S.cerevisiae*.

2. Identify the molecular basis of the genetic interaction between *CBP80* and the regulation by L30 of intronic 3'end region recognition.

RESULTS

In this section, the scientific results are presented in the form of research papers that are in the process of being published.

1. Publication 2

Deciphering 3' splice site selection in the *Saccharomyces cerevisiae* genome

Markus Meyer, Mireya Plass, Eduardo Eyras and Josep Vilardell

2. Publication 3

RPL30 regulation of splicing reveals distinct roles for Cbp80 in U1 and U2 snRNP co-transcriptional recruitment

Mireia Bragulat, Markus Meyer, Sara Macías[1], Maria Camats, and Josep Vilardell

Running title:

Deciphering 3' splice site selection in the *Saccharomyces cerevisiae* genome

Markus Meyer[1], Mireya Plass[2], Eduardo Eyras[2,3], and Josep Vilardell[1,4,*]

[1]Centre de Regulació Genòmica (CRG), Barcelona, Spain; [2]Computational Genomics Group, Universitat Pompeu Fabra (UPF), Barcelona, Spain; [3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain; [4]Institut de Biologia Molecular de Barcelona (IBMB), Barcelona, Spain.

Key words: 3'ss selection, secondary structure, splice site strength, branch site, *S.cerevisiae*, intron

*Corresponding autor:
Josep Vilardell
Present address: IBMB
Baldiri Reixach 10-12
08028 Barcelona
Spain
Tel: +34.93.4034668
Fax: +34.93.4034979
e-mail: josep.vilardell@ibmb.csic.es

ABSTRACT

Most eukaryotic ORFs are interrupted by introns, and their precise removal by the spliceosome is critical for gene expression. Given the low informational content of intronic 3' ends (3'ss), it has not been possible to explain the basis for their selection. Using the model system *S. cerevisiae*, we have addressed this problem using a combination of molecular and bioinformatics approaches. Our results allow to explain selection of all 3'ss with a single exception. We show that 3'ss selection hinges on both the substrate flexibility of the spliceosome, as well as the folding properties of introns. Remarkably, increasing intron size to achieve proper 3'ss does not affect cellular fitness, suggesting new evolutionary routes to regulation.

INTRODUCTION

Splicing is the evolutionarily highly conserved process whereby non-coding intronic regions are removed from pre-mRNA during mRNA synthesis. Accurate identification of intron boundaries (splice sites [ss]) is crucial. Errors result in incoherent and possibly deleterious transcripts, and is conductive to diseases (see[1]). Splicing is performed by the spliceosome, a complex ribonucleoprotein machine[2]. The 5'ss is identified through base-pairing to U1 snRNP[3-5]. Picking out the right 3'ss, only encoded by the dinucleotide AG, is a greater challenge because of its high frequency of occurrence. In metazoans the selected AG is bound by the dimer U2AF, which is partially conserved in yeast[6]. A mechanism of scanning from the branch site (BS) or polypyrimidine tract onwards to the first AG that would then be used as a 3'ss has been proposed[7]. However, in many cases, this first AG is not the splice site. This apparent discrepancy can in part be explained by *cis* elements. Indeed 3'ss can be made stronger by a U-rich tract[8], the 5'ss sequence can guide 3'ss choice[9], the exonic sequence just after the AG can help identify the true splice site[10], and distance from the BS also constitutes a critical element[11,12]. Moreover, additional protein factors play a role as well (for review see[13,14]). These data notwithstanding, accurate prediction of the 3'ss remains elusive; yet it is required to truly decode most eukaryotic pre-mRNAs.

The relative simplicity of the *Saccharomyces cerevisiae*, with only about 5% of genes that contain an intron[15], and a reduced collection of splicing modulators[16], provides a unique opportunity for a global analysis of 3'ss selection. Furthermore, the BS is generally conserved and unique in each transcript. The yeast transcriptome is well characterized, and while some cases of alternative splicing have been reported[17], alternative isoforms are largely underrepresented[18], indicating a predominantly straight forward 3'ss selection. This is notable, because possible alternative splice sites are not a rare occurrence. In fact, well over a third of introns contain up to eight possible 3'ss (sequence HAG, H : [UCA]) between the BS and the annotated

3’ss, and nothing apparently distinguishes them from their annotated counterparts. Here we undertook a bioinformatics and molecular approach to understand how the 3’ss is selected by the spliceosome and to decode a set of fundamental rules that allow us to explain the selection of the 3’ss in all but one introns of *S.cerevisiae*. Pre-mRNA folding and spliceosome flexibility are required for proper 3'ss selection, suggesting important evolutionary implications and new ways to regulate splicing.

RESULTS

The distance between the BS and the 3’ss (from the nucleotide after the BS adenosine, to the last nucleotide of the 3’ss) varies greatly, from 10 to 155 nucleotides. This is surprising in the light of the fact that earlier reports indicate that a large distance drastically hinders splicing efficiency[11,19]. Interestingly, our analyses indicated that the number of cryptic 3'ss between the BS and the annotated 3’ss is significantly lower (p = 1.71 x 10^{-17}) than what could be expected in absence of a bias (see **Methods** and **Sup. Data**). This implies a selective pressure against HAGs upstream of the annotated 3’ss. Even so, spliceosomal mechanisms to ensure proper 3'ss selection are inferred by the limited occurrence of splicing variants in yeast (unpublished data, and[17,18]). Likewise, as most yeast transcripts are efficiently spliced[20], the inhibitory effect of relatively large BS-3'ss distances must be overcome. Remarkably, these processes are likely to be related, as insinuated by our finding in the *VMA10* transcript. This RNA contains an intron with a long BS-3'ss region (**Sup. Fig. 1**). When a cryptic 3'ss, 48 nt upstream from the real 3'ss, is inactivated by a mutation, splicing efficiency to the annotated one is decreased (**Sup. Fig. 1**). This "long-distance" effect suggested the possibility of a structure between the BS and the annotated 3’ss. In this context, we argue that RNA folding acts doubly, by preventing the spliceosomal targeting of HAGs included in a structure, and by promoting the use of one located downstream.

In silico search for secondary structures

To test our model we built a dataset of all annotated *S.cerevisiae* introns in SGD (282, www.yeastgenome.org), and predicted possible RNA secondary structures between the BS and the 3'ss (see Methods, data can be viewed on http://regulatorygenomics.upf.edu/Yeast_Introns/index.html). From this analysis we conclude the following. First, more than one third (113) of *S.cerevisiae* introns have the potential to form a structure in this region. Interestingly, this group includes all those with a BS-3'ss distance larger than 45 nucleotides (**Fig. 1**, black and dark grey**)**. Second, the effective BS-3'ss distance distribution of those introns, calculated by subtracting the number of nt contained in the structure, is not significantly distinct (p = 3.4 x10^{-1}) from that of introns without structures (**Fig. 1**, light grey). Importantly, the effective BS-3'ss distance is predicted to be never larger than 45 nt. This suggests that 45 nt from BS to 3'ss may be the maximum distance for efficient splicing in *S. cerevisiae*, or indeed in yeasts, since this correlation is also observed for other species (**Sup. Fig. 2** and **Sup. Data**). Importantly as well, these conclusions cannot be reproduced with a similar sample of random sequences (**Sup. Data**).

To strengthen our *in silico* analyses, we computed independently the accessibility of a given AG between the BS and the 3'ss. We defined accessibility as the likelihood of a 3'ss not to be base-paired (in any fold). The data (**Sup. Fig. 3**, **Sup. Data**, and our website) indicate that annotated 3'ss are in many cases predicted to be significantly more accessible than cryptic AGs, in agreement with the folding predictions.

Validation of the secondary structures *in vivo*

To validate our predictions we expanded our *in silico* approach by identifying mutations predicted to disrupt the structure in the pre-mRNA and assessing their effect on 3'ss usage *in vivo*. Subsequently, we asked whether additional mutations, expected to restore the pre-mRNA structure, would stimulate

proper 3'ss selection. This strategy was tested with the *RPS23B* intron, where the 3'ss is at 60 nt from the BS. We predict that the BS-3'ss region folds, occluding an unused 3'ss (**Fig. 2a**). If this is correct, we expect that disruption of this structure releases this additional AG, making it available for splicing. To test this, the *RPS23B* intron was inserted into a *CUP1* reporter gene (widely used for splicing assays[21]), and splicing of several folding variants was monitored by primer extension analyses (**Fig. 2**). As expected, in the case of the wt *RPS23B* intron (*RPS23B* **1**), splicing goes exclusively to the annotated splice site (**Fig. 2c**, lane 1). The spliceosome is therefore oblivious to the AG that is sequestered within the predicted structure. The situation changes radically when the structure is disrupted by a set of 5 mutations (*RPS23B* **2**), and the alternative AG becomes the sole target of the spliceosome (**Fig. 2**, lane 2). Importantly, the wt situation is restored when the complementary mutations to the first set are introduced to reinstate the structure (*RPS23B* **3**; **Fig. 2c**, lane 3). In this case, splicing is undistinguishable from that of the wt, consistent with our predictions. The loss of splicing to the annotated 3'ss, when the structure is disrupted, could be attributed to a preference by the spliceosome for the first AG from the BS, as has been proposed[7]; or to an excessive (>45 nt) distance of the second AG from the BS. To assess this, an additional construct was made, in which the structure is disrupted as in *RPS23B* **2** but an additional mutation, AG to AU, is introduced to abrogate the upstream 3'ss (*RPS23B* **4**). In this case no mRNA is formed (**Fig. 2c**, lane 5). Interestingly, splicing is blocked after the first step, as shown by both accumulation of first step lariat intermediate in a *dbr1*Δ strain (**Fig. 2c**, lane 6), and lack of accumulation of pre-mRNA. These observations are consistent with the possibility that the annotated 3'ss cannot be reached by the spliceosome in the absence of a structure that brings it into proximity of the BS. We conclude that the structure that we predicted for *RPS23B* is formed *in vivo* and allows for the proper selection of the 3'ss.

A BS-3'ss distance greater than 45 nt decreases splicing efficiency

To test another of our structure predictions, and to refine the BS-3'ss distance requirements, we analyzed the *VMA10* intron. This intron of 162 nt has an interval of 105 nt between the BS and its annotated 3'ss. Our *in silico* analysis predicts a 63 nt structure (**Fig. 3a**) that includes two additional AGs, one at the top of the structure, and one at the very 3' end of it. Splicing of the wt construct (*VMA10* **1**) is as expected and goes to the annotated 3'ss (**Fig. 3c**, lane 1). Upon disruption of the predicted fold by the introduction of five mutations (*VMA10* **2**), splicing to the annotated AG is lost, and is replaced by weak splicing to the first AG (**Fig. 3c**, **lane 2**). We attribute the inefficiency of splicing to the large distance (57 nt) of this AG to the BS in the absence of a secondary structure. When the complementary mutations are introduced (*VMA10* **3**), the structure is restored and splicing is as in the wt (**Fig. 3c**, lane 3), consistent with our hypothesis. However, the possibility remained that the *VMA10* structure, in addition to bringing the 3'ss closer and occluding additional AGs, is required for proper *VMA10* splicing. To address this question, a construct was made in which the predicted 63 nt structure was precisely deleted (*VMA10* **4**). Notably, splicing of this pre-mRNA is undistinguishable from that of the wt (**Fig. 3c**, compare lanes 1 and 4), consistent with our model. We next verified that the distance between the BS and the 3'ss is the determining factor that impedes splicing of construct *VMA10* **2**. We made additional constructs based on *VMA10* **2**, in which the distance between the BS and the first AG is decreased from 57 nt (*VMA10* **2**) to 51 (*VMA10* **5**) or 45 (*VMA10* **6**, **Fig. 3d**). The result is that reducing this distance gradually restores splicing to wt levels at 45 nt (lanes 2-4). From these experiments we conclude that an AG has to be located at a maximum distance of about 45 nt from the BS to be efficiently used, which correlates well with our *in silico* findings (see above).

Rules of 3'ss selection

Both *VMA10* and *RPS23B* predicted structures are supported by our results *in vivo*. If we assume that the remaining predictions are valid, then 99 out of the 141 HAGs that are located between the BS and the annotated 3'ss are predicted to fall within a structure (see http://regulatorygenomics.upf.edu/Yeast_Introns/index.html), occluding them from the spliceosome as shown above. Out of the remaining 42 HAGs, 25 are located at a distance of 9 nt or less from the BS, whereas the shortest natural occurrence is of 10 nt. As for the rest (16), they are all AAGs. Therefore, we made two additional predictions. First, the minimal distance between the BS and the 3'ss is of 10 nt, and therefore AGs located closer to the BS cannot be used by the spliceosome. Second, the yeast spliceosome selects C/UAG over AAG, as reported in other systems[7]. We tested these predictions *in vivo* with constructs based on the *DMC1* intron. This intron contains an AAG at 16 nt from the BS and the annotated UAG is 12 nt further downstream. In case of the wt construct (*DMC1* **1**, **Fig. 4**, lane 1) the spliceosome essentially selects the annotated 3'ss, therefore bypassing the AAG. When the AAG is mutated to UAG (*DMC1* **2**) or when the two splice sites are inverted (*DMC1* **3**), splicing changes radically compared to the wild type (**Fig. 4**, lanes 1-3). In the first case, both splice sites are used to an equal extent, whereas UAG is again used exclusively in the latter construct. Therefore, in the wt, the selection of the second AG is likely due to the fact that the AAG is a relatively weak splice site, and not to the sequence context. However, when the distance of the UAG in the construct *DMC1* **3** was decreased from 16 to 9 nt (*DMC1* **4**, **Fig. 4**, lane 4), splicing again changed, and the weak splice site AAG was used exclusively. This shows that there is a clear minimal distance between the BS and the 3'ss, because a strong 3'ss that is located close to the BS cannot compete even with the weak AAG. Taken together, these results support our predictions and enable us to reveal a certain number of rules that are followed by the spliceosome in order to select a suitable 3'ss. To be used, a 3'ss has to be located within a window of 10 to 45 nt from the BS, either naturally or

with the help of a RNA fold. If there are two splice sites within the window, but outside of a structure, both will be used, unless one is stronger than the other, in which case there is a strong selection bias towards the stronger one.

We took advantage of the *DMC1* construct with two active UAGs (*DMC1* **2**) to determine whether the presence of an RNA fold between them alters 3'ss selection. For this we introduced the *RPS23B* stem between the two 3'ss (**Fig. 4b**). Remarkably, our data show that there is no change in the ratio of usage of 3'ss (**Fig. 4b**, lane 5), indicating that the spliceosome is able to probe for potential 3'ss upstream as well as downstream from the stem (but not inside). Consistent with our model, disrupting the stem (construct *DMC1* **6**) renders the downstream AG inaccessible (**Fig. 4b**, lane 6) because it is then too far from the BS.

Minimal distance requirement between BS and secondary structure

Given the critical role we propose for RNA structures in 3'ss selection, we next asked whether a fold can occur anywhere between the BS and the selected 3'ss. For this we made additional constructs based on *VMA10*, and decreased the distance between the BS and the structure from 20 nt in the wt, to 14 (*VMA10* **7**), 8 (*VMA10* **8)** and finally 5 (*VMA10* **9**, **Fig. 5**). Strikingly, when the distance to the stem is decreased to 8 nt, only half of the splicing goes to the annotated 3'ss, and the other half to the second AG (previously at the base of the stem, lane 3). When the distance is further decreased to 5 nt, splicing exclusively goes to this alternative splice site (lane 4). This indicates that the spliceosome does not tolerate a structure in close proximity of the BS, and causes it to melt from the bottom up, thus freeing the alternative 3'ss that was before trapped within the structure. This explains why in *VMA10* **9** the alternative 3'ss is used and why the annotated AG is not, because it is pushed outside of the spliceosome's scope. This strongly suggests that a structure cannot be formed at less than about 8 nt form the BS, and this was taken into account in the *in silico* structure predictions.

Secondary structure and biological fitness

Our results show that there is no apparent effect on *VMA10* splicing when the RNA structure that brings the 3'ss into spliceosomal range is removed (**Fig. 3c**, *VMA10* **1** and *VMA10* **4**). Since this structure covers about 40% of the intron, it seems a rather wasteful way to mediate 3'ss selection. In consequence, we investigated whether a secondary structure in *VMA10*, rather than a short distance between the BS and 3'ss, provides the cell with a biological advantage. To address this, we deleted the 63 nt that correspond to the secondary structure in the genomic copy of *VMA10*, and put this strain to compete by co-cultivation with its wt counterpart. The two strains were mixed then grown, collected and diluted repeatedly during 10 days (see **Methods**) and the presence of the two strains was assessed by Southern blot (**Fig. 6**). Importantly, *VMA10* is expressed in the conditions of the experiments, because deletion of the gene results in slow growth ([22] and data not shown). If one strain had had even a slight biological advantage over the other, it would have quickly become prevalent[23]. Remarkably, this is not what we observe and both strains grow equally. This indicates that, in our conditions, the structure does not impose any penalty upon the cell's fitness, even though it makes the intron almost twice as large as it would be if the BS-3'ss distance was short. Nor does it confer it a biological advantage by, for example, allowing for a means of regulation by modulating splicing through the secondary structure.

DISCUSSION

Around one third of *S. cerevisiae* introns have HAGs, between the BS and the annotated 3'ss. Yet, in most cases known mechanisms cannot explain why these potential splice sites are disregarded. In this study we have delineated the reasons that explain the choice of the spliceosome. The process of selection of the 3'ss is a composite of several factors, framed by a limited window. Remarkably, this space can be greatly diversified by the folding of the pre-mRNA, which does not seem to affect spliceosome function. Thus, the

substrate has evolved the capability to make a suitable HAG available (**Fig. 1** and **3**), occluding if necessary others in the process (**Fig. 2**). The efficient reach of the spliceosome appears to be limited at 45 nt from the BS, a number established both *in silico* (**Fig. 1**) and experimentally (**Fig. 3**). This number is significantly lower than what had been seen in other studies, both *in vitro* in HeLa extracts[7] or *in vivo* in yeast[11].We attribute this to the *in vitro* conditions (in HeLa), as well as the possibility of RNA folding in the constructs used for these studies. The fact that the spliceosome's scope is limited to 45 nt, apparently independently of the substrate, is intriguing. It could either be that this is as far as the spliceosome has time to search for a suitable splice site, before substrate discarding by Prp16[24] or, more likely, Prp22[25,26]. Alternatively, the 45 nucleotides could represent a physical distance beyond which the spliceosome active centre is not able to reach. Further research will be necessary to distinguish between these two models. Either way, the results obtained here argue against an exclusively scanning mechanism to explain 3'ss selection. Indeed, when two equivalent 3'ss are located within the window we have defined, both are used to equal extents (**Fig. 4**). It therefore appears that the spliceosome considers any splice site that is at its disposal within 10 to 45 nt from the BS. The fact that the active centre of the spliceosome shows tolerance to a diversity of RNA folds to perform the second step of catalysis is remarkable. But there is a number of nucleotides downstream from the BS where no functional 3'ss, nor RNA structures, are allowed. This indicates a structural requirement by the spliceosome to have access to about 9 nt downstream from the BS to be able to catalyze the exon ligation. These observations are consistent with the *in vitro* substrate requirements for spliceosome assembly described recently[27].

Out of the 141 HAGs located before the different 3'ss, 99 are sequestered in structures, which explains why they are not used. The remaining 42 HAGs are all either AAG or located at 9 nt or less from the BS. To explain why they are not used, we show that the spliceosome is oblivious to AAGs in the presence

of the more efficient splice sites UAG and CAG (**Fig. 4** lanes 1 to 3 and data not shown) which had been described also in the human a-tropomyosin transcript[7]. This favoured selection may be explained by the fact that an interaction between intron position -3 and G50 of U6 snRNA[28,29] is required for the second step of splicing. To assess the lack of selection of 3'ss found close to the BS, we show that even strong splice sites that are located at 9 nt from the BS become too weak to compete even with a weak splice site located downstream (**Fig. 4** lanes 2 and 4), consistent with previous reports[8].

With this set of rules, we are able to explain the splice site choice of nearly all introns. The only exception is *REC102*. Its intron contains three possible HAGs upstream of the 3'ss, two of which are contained in a secondary structure, but one of them, located only 5 nucleotides upstream of the splice site, is not. This and the annotated are both AAG and too proximal to be discriminated by a poly-U tract. However there is a selection mechanism, and splicing goes only to the annotated AG even in a NMD defective strain (data not shown).

Our data suggest that some RNA folds, such as in *VMA10*, could be removed without penalty on splicing (**Fig. 3**, lanes 1 and 4) or, more importantly, cellular fitness (**Fig. 6**). It is surprising that the insertion of 63 nt in a short intron of (otherwise) 99 nt has no impact; however, 63 extra nucleotides is probably a relatively low cost for an efficient mechanism of 3'ss selection. This is especially true when compared with higher eukaryotes where exons sizes are minimal compared to introns that for the better part will be degraded just after removal from the pre-mRNA. In addition, these structures offer an opportunity for regulation. This may not be the case for *VMA10*, or at least not in the conditions tested, but it cannot be excluded for others.

It is not the first time that secondary structures have been found to influence splicing in yeast. It has been shown for several introns, that structures could

help splicing by bringing the 5' and 3'ss into closer proximity[30] (reviewed in[31]). It had also been shown in the actin gene of another yeast species, *Kluyveromyces lactis*, that RNA folding is involved in splicing[32]. In Drosophila[33] or in human[31], regulation of alternative splicing by RNA folding has been demonstrated. However, we have shown that rather than being an oddity encountered in disparate transcripts, RNA secondary structures critical to splicing are widespread in the genome. Just between the BS and 3'ss of *S.cerevisiae*, over a third of genes harbor a secondary structure. Such structures will predictably be even more prominent in metazoans, with much larger introns and less defined BS sequences. These introns offer countless opportunities for secondary structure formation, 3'ss selection, and splicing regulation. That some of these RNA folds can be modulated by splicing factors or act as a riboswitches[34] is a testament to their evolutionary relevance.

FIGURES AND FIGURE LEGENDS

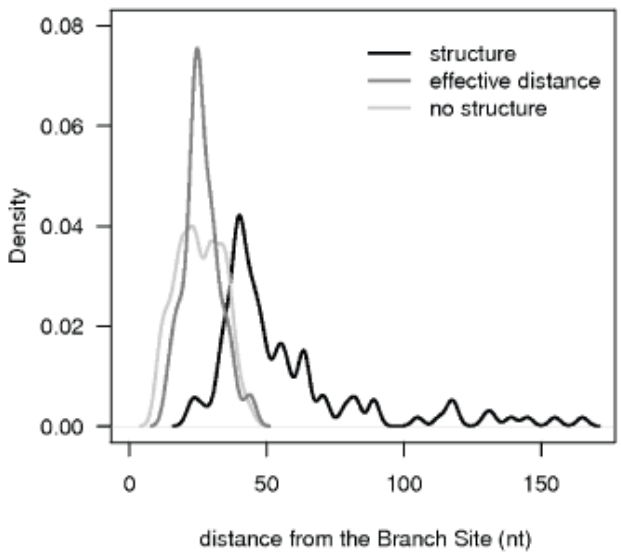


Figure 1 Meyer et al. Distribution of BS to 3'ss distances in *S.cerevisiae* introns. Introns are separated in two categories, those with a secondary structure between the BS and 3'ss (black line) and those without (light gray line). The dark gray line corresponds to the distribution of the former when the number of nucleotides comprised in the structure are removed from the BS to 3'ss distance. When the structures are taken into account, the BS to 3'ss distance never exceeds 45 nt.

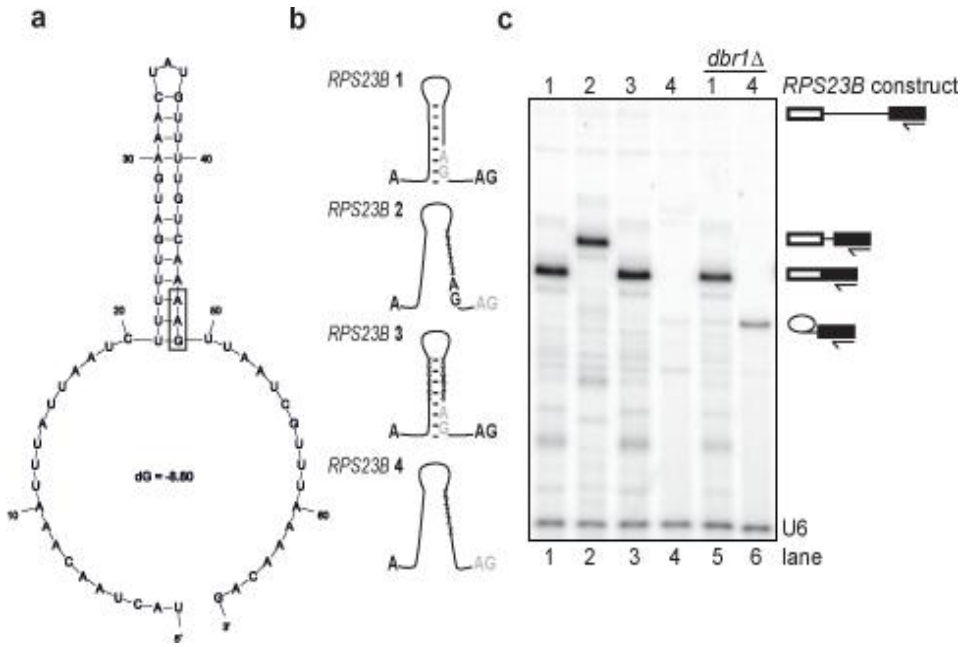


Figure 2 Meyer et al. *RPS23B* secondary structure and splicing pattern. (**a**) Predicted structure between the BS and 3'ss of *RPS23B*. (**b**) Schemes of the different constructs analyzed by primer extension in (**c**). In the wt construct (*RPS23B* **1)**, splicing goes to the annotated AG, whereas it is directed towards the alternative when the structure is disrupted by introduction of mutations (*RPS23B* **2)**. When the complementary mutations are introduced (*RPS23B* **3)** the wt splicing pattern is restored. In an open stem, when the alternative splice site is mutated (*RPS23B* **4)**, there is an accumulation of lariat intermediate in a *dbr1*Δ strain (lane 6).

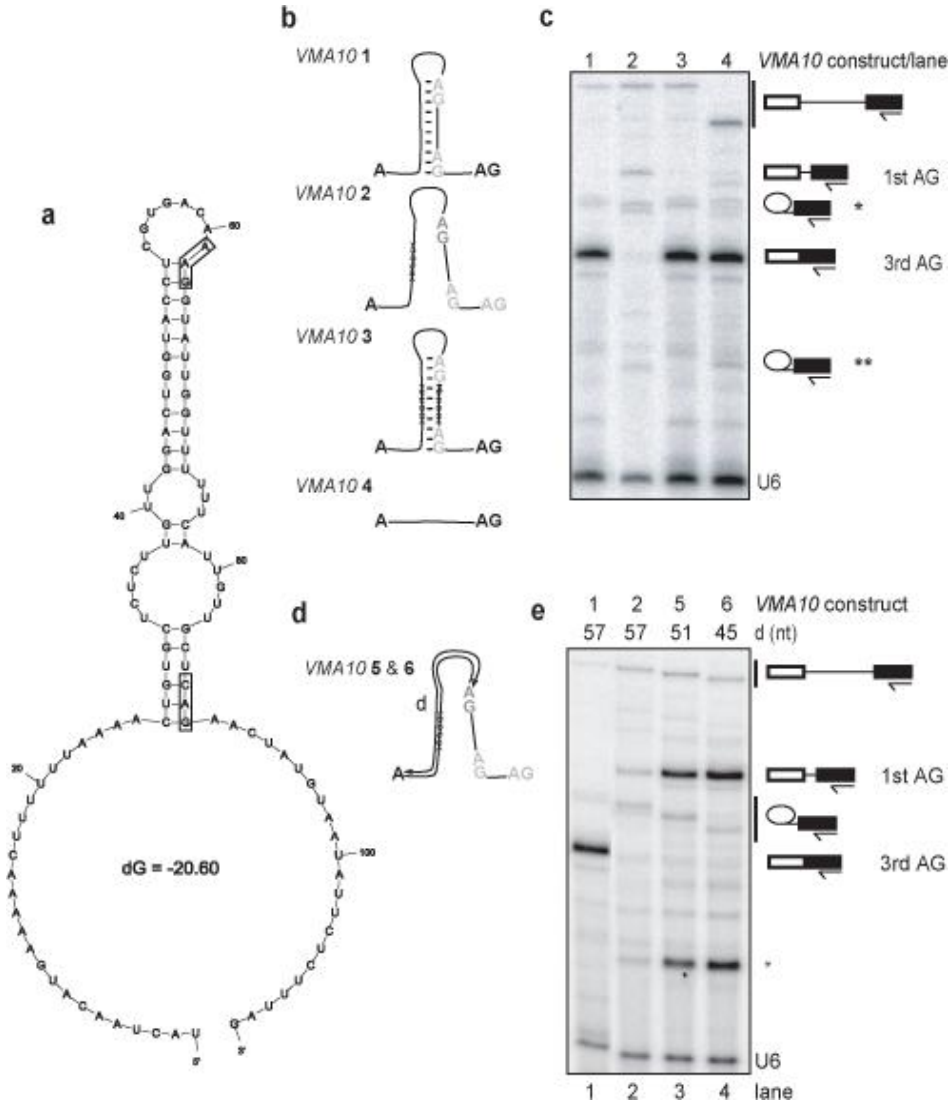


Figure 3 Meyer et al. *VMA10* secondary structure, splicing and effect of distance. The secondary structure predicted between the BS and 3'ss of *VMA10* is depicted in (**a**). The different constructs of which the splicing pattern was analyzed by primer extension are shown in (**b**) and (**d**). (**c**) The effect of mutations in the structure of *VMA10* is that splicing changes from going to the annotated splice site (*VMA10* **1**) to the 1st AG, but only weakly (*VMA10* **2**). The wt splicing pattern is restored when the complementary mutations are also introduced (*VMA10* **3**). When the 63 nucleotides of the

structure are deleted (*VMA10* **4**), there is no apparent difference between splicing of this construct and the wt. (**e**) The effect of distance between the BS and 1st AG in an open stem is tested. When this distance is reduced from 57 nt to 51 (*VMA10* **5**) and 45 (*VMA10* **6**), splicing becomes gradually more efficient. The * indicates an unaccounted for band.

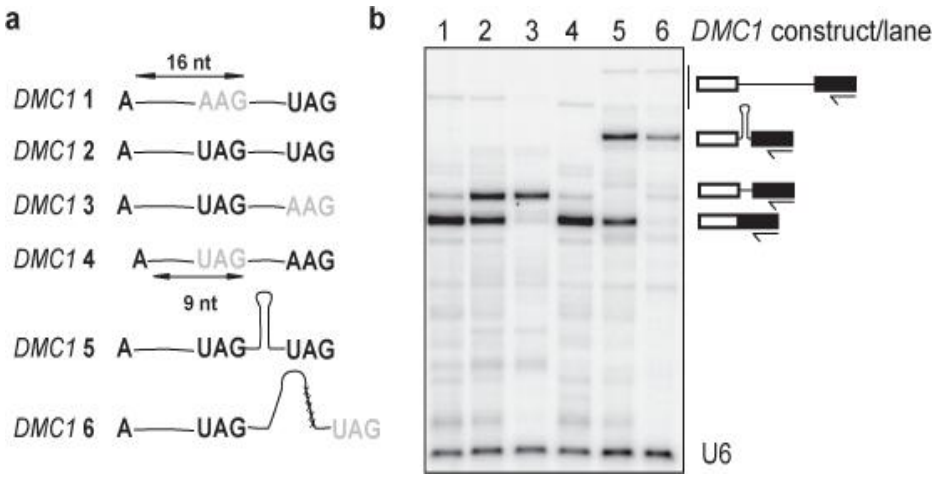


Figure 4 Meyer et al. Splice site strength and minimal distance between BS and 3'ss. The different constructs of which splicing was analyzed by primer extension are depicted in (**a**). (**b**) In the wt construct (*DMC1* **1**) mainly splicing to the annotated 3'ss is detected. When both splice sites are UAG ((*DMC1* **2**) both are equally used. Inverting the splice sites compared to the wt (*DMC1* **3**) also results in an inverted splicing pattern. Putting a strong splice at 9 nt form the BS (*DMC1* **4**) rather than 16 as in the wt makes it even weaker than an already weak splice site. Splicing of construct *DMC1* **5** indicates that introduction of the structure of *RPS23B* between two splice sites does not change the usage of either of them, unless the stem is disrupted (*DMC1* **6**), and the distance between the BS and the second splice site becomes greater than 45 nucleotides.

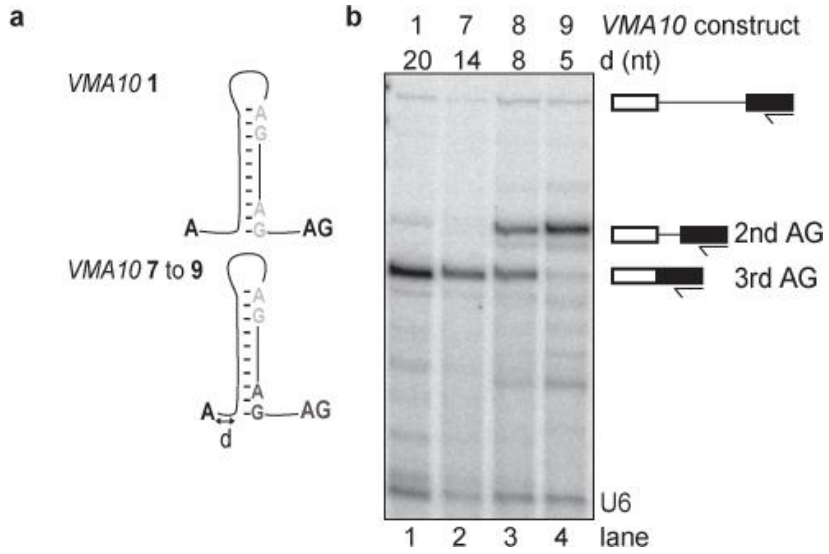


Figure 5 Meyer et al. Requirement of a minimal distance between the BS and the beginning of a structure. Splicing patterns of constructs depicted in (**a**) are shown in (**b**), as revealed by primer extension. When the distance between the BS and the secondary structure of *VMA10* is reduced from 20 (*VMA10* **1**) to 14 nt (*VMA10* **7**), the splicing pattern does not change. However, when this distance is further decreased to 8 nt (*VMA10* **8**) the splicing changes, and both, the annotated and 2nd splice sites are used. Reducing this distance by 3 additional nt (*VMA10* **9**) results in splicing only to the 2nd AG.

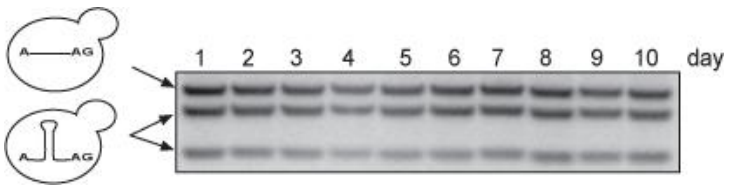


Figure 6 Meyer et al. Biological fitness with or without a structure. A wt strain was put in competition for biological fitness with a strain that differs from the wt only in a deletion of the 63 nt that correspond to the structure formed in the *VMA10* intron (**Fig. 3**). During 10 days a mixed culture of both strains was repeatedly grown, and diluted. The presence of the two strains was then assessed by southern blot analysis, which reveals that both strains are equally fit.

MATERIALS AND METHODS

S. cerevisiae intron dataset

We downloaded the annotation and genomic sequence of *S. cerevisiae* from the Saccharomyces Genome Database (SGD July 2009). We then extracted all introns from chromosomal genes (327) and kept only those that had length > 0nt, canonical splice sites (GT or GC at the 5'ss and AG at the 3'ss) and did not have any ambiguous nucleotide (N) in the sequence, obtaining a final set of 282 introns.

Branch site prediction

We looked for NNNTRACNN motifs in the 200 nt upstream of the 3'ss. If the intron was shorter than that, we scanned the whole intron. We defined the BS as the motif having the smallest Hamming distance to the TACTRACNN motif. When several motifs with the same Hamming distance where found, we selected the motif having the smallest basepairing energy to the U2snRNA (see below). If several 9-mers had the same basepairing energy, we selected the closest to the 3'ss.

Energy measurement

To calculate the basepairing energy between the BS and the U2snRNA we used the program RNAcofold from the Vienna package. The value obtained depends on the sequence composition, the length, and the number of nucleotides involved in the basepairing. In this case, we calculated the energy between all putative BS and GTGTAGTA, forcing the basepairs between the two sequences and leaving the A of the branch site unpaired.

Expected values

For each of the species we extracted the frequencies of each nucleotide in the BS-3'ss region for all the introns in the dataset discarding the first 7nt after the invariable A of the BS. The expected values of each triplet were then calculated assuming that nucleotide positions are independent.

Secondary structure prediction

For each intron, we recovered the sequence between the BS and the 3'ss, and discarding both signals. From this region, we further removed the first eight nucleotides after the BS A, as previous experiments show that these nucleotides cannot belong to a secondary structure. In the selected region, we did the secondary structure prediction using the program RNAfold from the Vienna package with default parameters. The predictions for all yeast species analyzed and the pair probabilities for each structure can be found at http://regulatorygenomics.upf.edu/Yeast_Introns/.

Effective distance measure

We defined the distance between the BS and any 3'ss as the number of nucleotides between the A of the BS and the 3'ss, including the last one. Using this definition, TACTAACACNNNN|TAG would be a distance of 10 nt. We defined the effective BS-3'ss distance as the linear distance (in nucleotides) between the BS and the 3'ss after removing the secondary structure. More specifically, we removed all the bases that were part of a structured region and each structured region was substituted by 2 bases, corresponding to the beginning and the end of the structured region.

Accessibility measurement

Accessibility is defined as the probability of a nucleotide not to beto another nucleotide. Thus, it is one minus the pair probability. We calculated pair probabilities using the program RNAfold. For each of the AGs (real and cryptic 3'ss), we measured the accessibility of the three nucleotides (HAG) using 4 different windows of increasing length size with lengths n, n+5, n+10, and n+15 respectively, where n is the BS-HAG distance (see above). The final value is the average of the values of the three nucleotides averaged over the four windows.

Strains and reporter plasmids

*S.cerevisiae*s trains are derived from BY4741. The reporter plasmids are based on pCC71{Collins, 1999 #112} where the actin intron has been replaced by the various versions of the different introns studied. As a result, the introns are flanked by 50 nt of actin leader, and the *CUP1* gene. Cloning strategies and oligonucleotides are available in **Supplementary Data**.

Primer extension

Performed as described in {Siatecka, 1999 #111}, on RNA from strain BY4741 *upf1*Δ unless stated otherwise, carrying the reporter plasmid. The oligonucleotides used are YAC6, complementary to *CUP1* and yU6-61 complementary to U6 which is used as a loading control {Xu, 2007 #113}.

Competition experiments and southern blot

Performed as in {Li, 1996 #30}. The strains used are BY4741 and a mutant with a genomic deletion of the 63 nucleotides corresponding to the structure of VMA10 intron shown in **Fig. 4**.

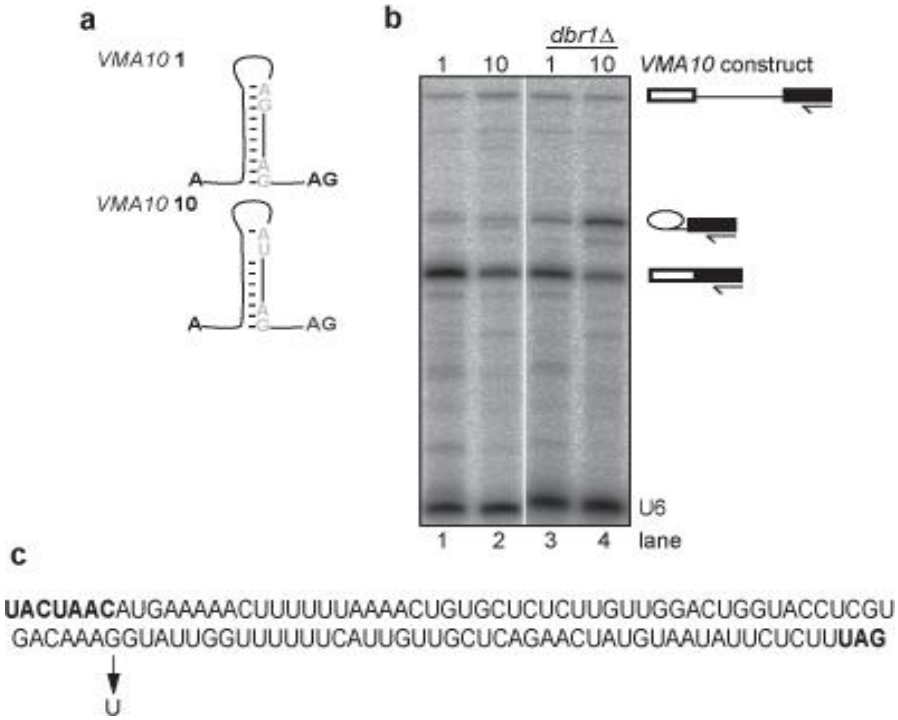
REFERENCES

1. Cooper, T.A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777-93 (2009).
2. Nilsen, T.W. The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* **25**, 1147-9 (2003).
3. Seraphin, B. & Rosbash, M. Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**, 349-58 (1989).
4. Zhuang, Y. & Weiner, A.M. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* **46**, 827-35 (1986).
5. Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature* **458**, 475-80 (2009).
6. Kielkopf, C.L., Lucke, S. & Green, M.R. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* **18**, 1513-26 (2004).
7. Smith, C.W., Chu, T.T. & Nadal-Ginard, B. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol Cell Biol* **13**, 4939-52 (1993).
8. Patterson, B. & Guthrie, C. A U-rich tract enhances usage of an alternative 3' splice site in yeast. *Cell* **64**, 181-7 (1991).
9. Goguel, V. & Rosbash, M. Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell* **72**, 893-901 (1993).
10. Crotti, L.B. & Horowitz, D.S. Exon sequences at the splice junctions affect splicing fidelity and alternative splicing. *Proc Natl Acad Sci U S A* **106**, 18954-9 (2009).
11. Cellini, A., Felder, E. & Rossi, J.J. Yeast pre-messenger RNA splicing efficiency depends on critical spacing requirements between the branch point and 3' splice site. *Embo J* **5**, 1023-30 (1986).
12. Luukkonen, B.G. & Seraphin, B. The role of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in Saccharomyces cerevisiae. *Embo J* **16**, 779-92 (1997).
13. Smith, D.J., Query, C.C. & Konarska, M.M. "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Mol Cell* **30**, 657-66 (2008).
14. Umen, J.G. & Guthrie, C. The second catalytic step of pre-mRNA splicing. *Rna* **1**, 869-85 (1995).
15. Spingola, M., Grate, L., Haussler, D. & Ares, M., Jr. Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *Rna* **5**, 221-34 (1999).
16. Plass, M., Agirre, E., Reyes, D., Camara, F. & Eyras, E. Co-evolution of the branch site and SR proteins in eukaryotes. *Trends Genet* **24**, 590-4 (2008).

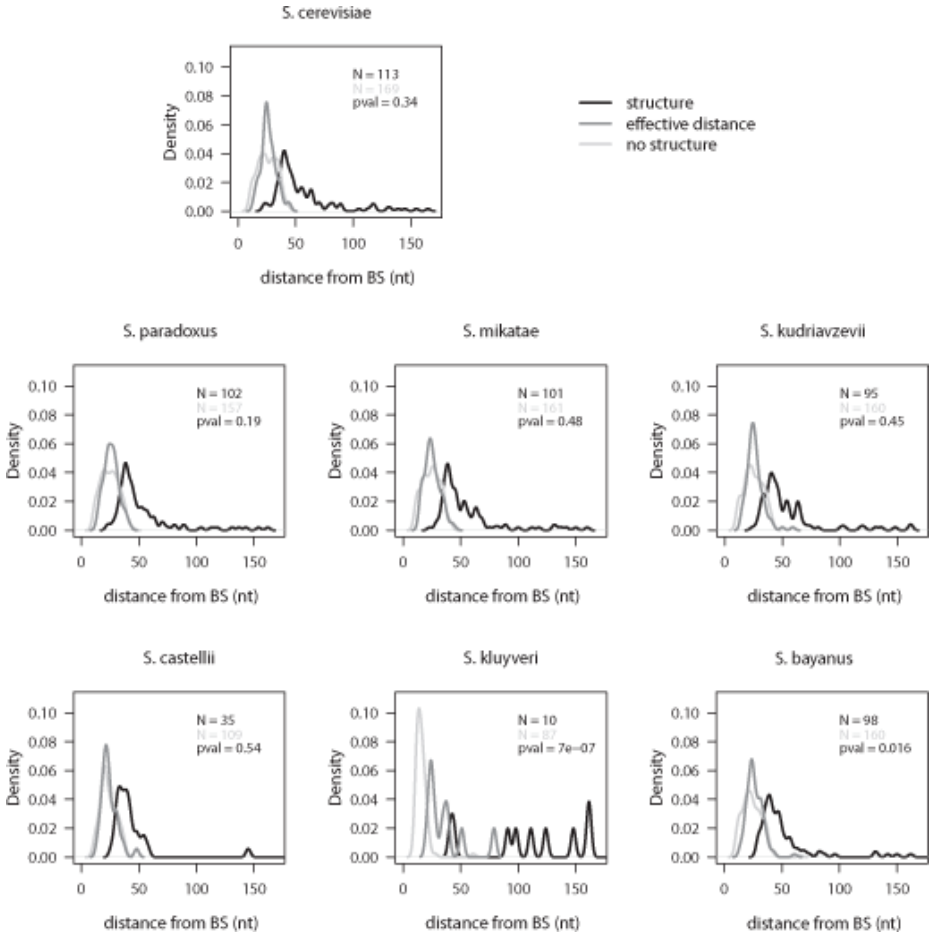
17. Davis, C.A., Grate, L., Spingola, M. & Ares, M., Jr. Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res* **28**, 1700-6 (2000).
18. Yassour, M. et al. Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. *Proc Natl Acad Sci U S A* **106**, 3264-9 (2009).
19. Langford, C., Nellen, W., Niessing, J. & Gallwitz, D. Yeast is unable to excise foreign intervening sequences from hybrid gene transcripts. *Proc Natl Acad Sci U S A* **80**, 1496-500 (1983).
20. Clark, T.A., Sugnet, C.W. & Ares, M., Jr. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* **296**, 907-10 (2002).
21. Lesser, C.F. & Guthrie, C. Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics* **133**, 851-63 (1993).
22. Supekova, L., Supek, F. & Nelson, N. The Saccharomyces cerevisiae VMA10 is an intron-containing gene encoding a novel 13-kDa subunit of vacuolar H(+)-ATPase. *J Biol Chem* **270**, 13726-32 (1995).
23. Li, B., Vilardell, J. & Warner, J.R. An RNA structure involved in feedback regulation of splicing and of translation is critical for biological fitness. *Proc Natl Acad Sci U S A* **93**, 1596-600 (1996).
24. Schwer, B. & Guthrie, C. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349**, 494-9 (1991).
25. Mayas, R.M., Maita, H. & Staley, J.P. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat Struct Mol Biol* **13**, 482-90 (2006).
26. Query, C.C. & Konarska, M.M. Splicing fidelity revisited. *Nat Struct Mol Biol* **13**, 472-4 (2006).
27. Fabrizio, P. et al. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* **36**, 593-608 (2009).
28. Collins, C.A. & Guthrie, C. Genetic interactions between the 5' and 3' splice site consensus sequences and U6 snRNA during the second catalytic step of pre-mRNA splicing. *Rna* **7**, 1845-54 (2001).
29. Konarska, M.M., Vilardell, J. & Query, C.C. Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol Cell* **21**, 543-53 (2006).
30. Rogic, S. et al. Correlation between the secondary structure of pre-mRNA introns and the efficiency of splicing in Saccharomyces cerevisiae. *BMC Genomics* **9**, 355 (2008).
31. Warf, M.B. & Berglund, J.A. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci* (2009).
32. Deshler, J.O. & Rossi, J.J. Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron. *Genes Dev* **5**, 1252-63 (1991).

33. Graveley, B.R. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**, 65-73 (2005).
34. Cheah, M.T., Wachter, A., Sudarsan, N. & Breaker, R.R. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447**, 497-500 (2007).

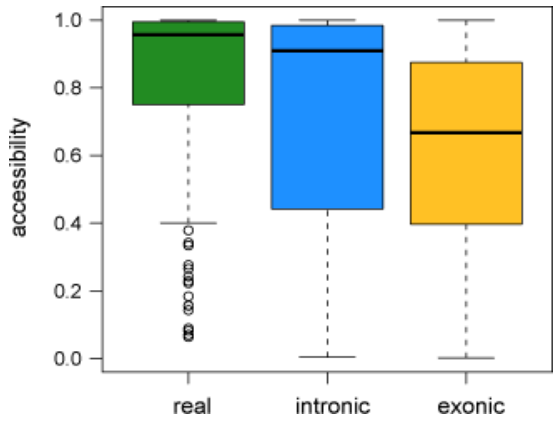
SUPPLEMENTARY FIGURES AND LEGENDS



Supplementary Figure 1 Meyer et al. A mutation 48 nt upstream of the 3'ss impairs splicing. The splicing pattern of the two constructs based on the *VMA10* introns (**a**) is shown in (**b**). The mutation of AAG to AAU (**c**) in the *VMA10* intron causes a large deficit in splicing. This is due specifically to impairment of the second step, as evidenced by 1st step lariat accumulation in a *dbr1*Δ strain.



Supplementary Figure 2 Meyer et al. Similar distribution of effective BS to 3'ss distances in different *Saccharomyces*. The same analysis as the one shown in **Fig. 1** for *S.cerevisiae*, was carried out for additional *Saccharomyces* species. The maximum BS-3'ss distance is very similar in those other spicies



Supplementary Figure 3 Meyer et al. Accessibility of HAGs. The probability for HAGs to be outside of a secondary structure is represented in a box-plot. The annotated 3'ss (green) is in average more likely not to be base-paired than HAGs found upstream (blue) or downstream (orange) of it.

SUPPLEMENTARY DATA

Strains

Deletion strains are from the YKO MATa Strain Collection (Open Biosystems).The wt strain (yJV131) and the strain with the deletion of the 63 nt that correspond to the secondary structure described in the *VMA10* intron (yJV132) where generated by transforming the *vma10*Δ strain with PCR JV1328 (5'-ataaagcaatcacaccttaaacaatacatttttttttttcctgcaatctccaaagtgtgcAAGGTATACAAAGCAGAATGGTA-3') -JV1329 (5'- CCGAATGCAATGTCAGTTTA) using genomic DNA as a template for the first and PCR JV1327 (5'-aaggtatacaaagcagaatggtatgtgccattacattacgtgtcaacacttctgtctctaacaagcgttcttactaacatgaaaaacttttttaaaaAACTATGTAATATTCTCTTTAGT-3') -JV1329 for the second.

Plasmids

All constructs used for this study were made by a BamHI/PacI digestion of pMM4c and a PCR product that contains a BamHI site followed by 50 nt of the actin leader, ATG, an intron followed by 14 to 21 nt of exon 2 and a PacI site. The oligonucleotides used for every construct are detailed bellow.All PCRs were made with genomic DNA of J47a as template unless stated otherwise, and upon cloning, all constructs were sequenced.

*RPS23B***1** (pMM42): JV1358 (5'-ctaggatccccggcgactcttttagatttttcacgcttcactgctttttttcttcccaaatgACAAGTATGTACAATTATAGAAGATTG-3') - JV1359 (5'-tacgttaattaaCTCTTCTTATAGTTGTTTTCGGCCCAACGGCTGTTT-3')

*RPS23B***2** (pMM43): JV1358-JV1360 (5'-tacgttaattaaCTCTTCTTATAGTTGTTTTCGGCCCAACGGCTGTTTTAAACGAaTAtCTTTacACAAtACATAGTTTCATCAAAAAGATT-3')

*RPS23B***3** (pMM44): JV1358-JV1361 (5'-tacgttaattaaCTCTTCTTATAGTTGTTTTCGGCCCAACGGCTGTTTTAAA

CGAaTAtCTTTacACAAtACATAGTaTCATgtAAAAGAaTAtTAAATTTG
TTAGTAAATG-3')

*RPS23B***4** (pMM52): JV1358-JV1383 (5'-
tacgttaattaaCTCTTCTTATAGTTGTTTTCGGCCCAACGGCTGTTTTAAA
CGAaTAtaTTTacACAAtACATAGTTTCATCAAAAAGATT-3')

VMA10 **1** (pMM18): JV1128 (5'-
ctaggatccccggcgactcttttagatttttcacgcttcactgctttttttcttcccaaatgATGGTATGTGC
CATTACATTAC-3') -JV1130 (5'-
tacgttaattaaTCCGTTTTTTTTGGGACTAAAG-3')

VMA10 **2** (pMM29): JV1128-JV1269 (5'-
tacgttaattaaTCCGTTTTTTTTGGGACTAAAGAGAATATTACATAGTTCTG
AGCAACAATGAAAAAACCAATACCTTTGTCACGAGcaACCAcTCCA
AgAAGAGAGgACAGTTTTAAAA-3')

VMA10 **3** (pMM30): JV1128-JV1270 (5'-
tacgttaattaaTCCGTTTTTTTTGGGACTAAAGAGAATATTACATAGTTCTG
AcCAACAATcAAAAAAgCAATtgCTTTGTCACGAGcaACCAcTCCAAg
AAGAGAGgACAGTTTTAAAA-3')

VMA10 **4** (pMM31): JV1128-JV1271 (5'-
tacgttaattaatccgtttttttgggactaaagagaatattacatagttTTTTAAAAAAGTTTTTCAT
GTTAG-3')

VMA10 **5** (pMM47): JV1128-JV1380 (5'-
tacgttaattaaTCCGTTTTTTTTGGGACTAAAGAGAATATTACATAGTTCTG
AGCAACAATGAAAAAACCAATACCTTTGTCACGAGcaACCAcTCCA
AgAAGAGAGgACAG-3') using as template PCR JV1128-JV1376 (5'-
caaccactccaagaagagaggacagAAAAGTTTTTCATGTTAGTAAGAAC-3')

VMA10 **6** (pMM48): JV1128-JV1380 using as template PCR JV1128-JV1377
(5'-caaccactccaagaagagaggacagTTTTCATGTTAGTAAGAACGC-3')

VMA10 **7** (pMM60): JV1128-JV1382 (5'-
tacgttaattaaTCCGTTTTTTTTGGGACTAAAGAGAATATTACATAGTTCTG
AGCAACAATGAAAAAACCAATACCTTTGTCACGAGGTACCAGTCC

AACAAGAGAGCACAG-3') using as template PCR JV1128-JV1402 (5'-gtaccagtccaacaagagagcacagAAAAGTTTTTCATGTTAGTAAGAAC-3')

VMA10 **8** (pMM59): JV1128-JV1382 using as template PCR JV1128-JV1401 (5'-gtaccagtccaacaagagagcacagTTTTCATGTTAGTAAGAACGC-3')

VMA10 **9** (pMM58): JV1128-JV1382 using as template PCR JV1128-JV1400 (5'-gtaccagtccaacaagagagcacagtttTGTTAGTAAGAACGCTTGTTAG-3')

DMC1 **1** (pMM63): JV1406 (5'-ctaggatccccggcgactcttttagatttttcacgcttcactgctttttcttcccaaatgGTATGTTATAATAACATTTTAAAACC-3') -JV1407 (5'-tacgttaattaaTCTTGTTGTTGACAAAACGGT-3')

DMC1 **2** (pMM65): JV1406-JV1409 (5'-tacgttaattaaTCTTGTTGTTGACAAAACGGTCTATAAAAGTTCCTaTCCAAATTATTAGTTAGTAAAAG-3')

DMC1 **3** (pMM66): JV1406-JV1410 (5'-tacgttaattaaTCTTGTTGTTGACAAAACGGTCTtTAAAAGTTCCTaTCCAAATTATTAGTTAGTAAAAG-3')

DMC1 **4** (pMM76): JV1406-JV1434 (5'-tacgttaattaatcttgttgttgacaaaacggtctttaaaagttcctaTATTAGTTAGTAAAAGAAAGGGG-3')

DMC1 **5** (pMM71): JV1406-JV1429 (5'-tacgttaattaatcttgttgttgacaaaacggtctataaaaattaacttttgacaaaacatagtttcatcaaaaagattaatGTTCCTaTCCAAATTATTAGTTAGTAAAAG-3')

DMC1 **6** (pMM72): JV1406-JV1430 (5'-tacgttaattaatcttgttgttgacaaaacggtctataaaaaatatctttacacaatacatagtttcatcaaaaagattaatGTTCCTaTCCAAATTATTAGTTAGTAAAAG-3')

Homologous introns datase

We used Galaxy to extract the homologous regions to the *S.* cerevisae introns in 6 *Saccharomyces* species (*S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, *S. castellii* and *S. kluyveri*). We extracted the genomic alignments for the 7 yeast species provided by UCSC and kept only those containing

canonical splice sites and no ambiguous nucleotides in the sequence. Subsequently, we built pairwise alignments between each of the putative introns and the *S. cerevisiae* homologous introns using PRANK and defined as homologous introns only those that contained the BS aligned with the *S. cerevisiae* BS. For each of the homologous introns obtained, we did independent BS predictions applying the same method used for *S. cerevisiae.*

Running title:

RPL30 regulation of splicing reveals distinct roles for Cbp80 in U1 and U2 snRNP co-transcriptional recruitment

Mireia Bragulat, Markus Meyer, Sara Macías[1], Maria Camats, and Josep Vilardell[*]

Centre de Regulació Genòmica, Dr. Aiguader 88, 08003 Barcelona, Spain

Key words: regulated splicing, Cbp80, 5'SS, BS, ChIP, U1 snRNP, U2 snRNP, *RPL30*, L30

*Corresponding autor:
Josep Vilardell.
Present address: IBMB
Baldiri Reixach 10-12
08028 Barcelona
Spain
Tel: +34.93.4034668
Fax: +34.93.4034979
e-mail: josep.vilardell@ibmb.csic.es

[1]Present address: Western General Hospital (Medical Research Council), Crewe Road, EH4 2XU Edinburgh, Scotland (UK)

ABSTRACT

Pre-mRNA splicing is catalyzed by the spliceosome, and its control is essential for correct gene expression. While splicing repressors typically interfere with transcript recognition by spliceosomal components, the yeast protein L30 blocks spliceosomal rearrangements required for the engagement of U2 snRNP (small ribonucleoprotein particle) to its own transcript *RPL30*. Using a mutation in the *RPL30* binding site that disrupts this repression, we have taken a genetic approach to reveal that regulation of splicing is restored in this mutant by deletion of the cap-binding complex (CBC) component Cbp80. Indeed, our data indicate that Cbp80 plays distinct roles in the recognition of the intron by U1 and U2 snRNP. It promotes the initial 5' splice site recognition by U1 and, independently, facilitates U2 recruitment, depending on sequences located in the vicinity of the 5' splice site. These results reveal a novel function for CBC in splicing and imply that these molecular events can be the target of a splicing regulator.

INTRODUCTION

Most eukaryotic pre-mRNAs need to be spliced before being translated. During pre-mRNA splicing, intervening sequences (introns) are precisely removed and the adjacent sequences (exons) are spliced together. This process takes place in a large complex ribonucleoprotein particle (RNP) called spliceosome (reviewed in (Wahl et al., 2009)). Multiple results support a co-transcriptional assembly of the spliceosome in metazoans (reviewed in (Bentley, 2005; Kornblihtt et al., 2004; Moore and Proudfoot, 2009)), while in budding yeast assembly starts during transcription but completion of spliceosome formation and splicing occurs mostly post-transcriptionally (Gornemann et al., 2005; Lacadie and Rosbash, 2005; Macías et al., 2008; Tardiff et al., 2006). Research in several systems supports a model for spliceosome assembly based upon an ordered, albeit dynamic, pathway for building the catalytic complex. In the first stage, U1 small nuclear RNP (snRNP) recognizes sequences at the 5' splice site (5'SS) and binds the nascent intron, initiating the formation of the spliceosome and committing the transcript to splicing. Recognition of the 5'SS by U1 is assisted by the cap-binding complex (CBC), an heterodimer of the factors Cbp20 and Cbp80 (Izaurralde et al., 1994). This role of CBC is especially important in introns with poor potential for base-pairing to U1 (Colot et al., 1996; Fortes et al., 1999a; Fortes et al., 1999b; Lewis et al., 1996). Subsequently, the 3' end of the intron, including the branch site (BS) and the 3' splice site (3'SS), is identified by the factors BBP and Mud2 in budding yeast, and SF1 and U2AF65 in metazoans. An interaction between these components and U1 snRNP has been shown in this complex (Abovich and Rosbash, 1997; Kent and MacMillan, 2002; Reed, 1997; Rutz and Seraphin, 1999), known as commitment complex or CC (complex E in metazoans). In the next step, a poorly understood remodeling of the CC takes place, likely to be promoted by members of the family of DExH/D box RNA helicases (Bleichert and Baserga, 2007; Cordin et al., 2006), such as Sub2 and Prp5. This process affects the accessibility of the BS and leads to the association of U2 snRNP

with the BS forming the pre-spliceosome or complex A in mammals (Parker et al., 1987; Wu and Manley, 1989; Zhuang and Weiner, 1989). Next, the pre-spliceosome is further remodeled and the tri-snRNP U4/U6.U5 engages in the nascent particle (complex B in mammals), and subsequent rearrangements involving the displacement of U1 and U4 snRNP lead to the formation of a catalytically active spliceosome (complex C) (Jurica and Moore, 2003; Villa et al., 2002; Will, 2006). In addition to its role in the formation of the commitment complex, the CBC has also been shown to promote association of U4/U6.U5 to the nascent spliceosome (Gornemann et al., 2005; O'Mullane and Eperon, 1998; Staley and Guthrie, 1999).

Spliceosome assembly is tightly regulated at multiple levels (Dreyfuss et al., 2002; Graveley, 2000; Hertel and Graveley, 2005; House and Lynch, 2008). Most known splicing factors act by modulating access to splicing signals, occluding or exposing particular RNA sequences, or by interfering with interactions between spliceosomal components. In *Saccharomyces cerevisiae*, genetic analyses have provided relevant data on splicing mechanisms (see (Brow, 2002) for a review). Despite its reduced number of introns and a relatively simple gene structure (Spingola et al., 1999), there is evidence indicating that the yeast spliceosome can distinguish between groups of transcripts (Pleiss et al., 2007). Splicing regulation is a key process during gene expression in yeast, where it has been studied in some detail providing important clues on how splicing can be modulated (Meyer and Vilardell, 2009).One example is the essential *RPL30* gene, encoding the ribosomal protein L30 which, when in excess, binds to its own transcript and stalls spliceosome assembly. L30 interacts with a kink-turn structure in the *RPL30* transcript that mimics the L30 rRNA binding site (Vilardell et al., 2000) (and references therein). Recent results in our laboratory show that L30 prevents association of U2 snRNP with the BS by a distinct mechanism likely to involve an interference with conformational changes that occur during spliceosome assembly, rather than with the recognition of splicing signals

themselves (Macías et al., 2008). We now follow a genetic approach to address the molecular mechanisms involved in this regulation. Taking advantage of silent *RPL30* mutants that affect splicing regulation by L30, we have sought mutations in *trans* that restore control of splicing. We have identified multiple mutations in the CBC component Cbp80, indicating that this factor has a wider role in spliceosome assembly than previously assumed. This includes a function in U2 recruitment that is the target of a splicing regulator.

RESULTS

C9 in *RPL30* is required by L30 to repress U2 snRNP recruitment

To undertake a genetic approach to dissect mechanisms involved in the repression of U2 snRNP recruitment by L30, we investigated several mutations in the *RPL30* transcript that are known to abolish control of splicing (Eng and Warner, 1991; Vilardell and Warner, 1994). Mutation C9U (9 relative to the start of transcription, Fig. 1A) weakens L30 binding *in vitro* and disrupts splicing repression (Fig. 1C, lanes 7-12; and (Vilardell et al., 2000)). C9U likely destabilizes the kink-turn RNA structure (Klein et al., 2001) recognized by L30. Consistent with this possibility, combination of C9U with a more stable version of the *RPL30* kink-turn (transcript "5A", with the large loop that is not required for L30 binding reduced to five adenines; Fig. 1B) restores L30 binding (transcript 5A C9U in Fig. 1C, lanes 19-24). To test the effect of 5A C9U *in vivo* we used the *LCUP1* reporter system, based on fusing *RPL30* exon 1 and intron to the *CUP1* ORF (Vilardell and Warner, 1997) (Fig. 2A). Northern analyses shown in Fig. 1D indicate that splicing of wt *LCUP1* transcripts is regulated by L30 (lane 5), while C9U mutants fail to be regulated (lane 6). *LCUP1* 5A transcripts are highly repressed under excess L30 (lane 3 *vs* 1), consistent with increased affinity to L30. Interestingly, regulation of splicing by L30 is not restored in *LCUP1* 5A C9U transcripts (lanes 4 *vs* 2). We conclude that both the stability and the

sequence of the *RPL30* kink-turn are important for regulation, and that the C9U mutation abolishes regulation even in transcripts where the kink-turn is stably bound by L30.

There are two alternate possibilities to explain the C9U phenotype of lack of repression. L30 is either dislodged during intron recognition, or remains associated with the transcript but fails to repress U2 snRNP recruitment. To distinguish between these possibilities, we determined the levels of co-immunoprecipitation of L30 with BBP, in the presence of either 5A C9U or wt transcripts. As a positive control we used a synthetic transcript in which splicing and L30 binding are compatible, as the 5'SS is located 12 nt downstream of the L30 binding site ("+12" RNA) (Macías et al., 2008). Fig. 1E shows that L30 binding is compatible with recognition of the BS by BBP in 5A C9U transcripts, as evidenced by the co-immunoprecipitation of MBP-L30 and BBP. Levels of co-immunoprecipitation are similar to those with wt and "+12" transcripts, (compare lane 5 *vs* 4 and 6), indicating that a C at position 9 is required for repression of splicing by L30 in RNA molecules simultaneously bound by L30 and BBP. Consistent with this possibility, U2 snRNP fails to co-immunoprecipitate with L30 in 5A C9U RNAs, while recognition of the 5'SS by U1 is not affected by L30 in these transcripts (Fig. 1F). We conclude that C9U blocks repression of splicing after intron recognition, and we decided to search for mutations that restore regulation of splicing by L30 to transcripts with this mutation.

Screen for regulation of splicing of C9U transcripts

Our genetic approach is based in two constructs based on the *LCUP* reporter (Fig. 2A, B). In the *LCUPIF* reporter the ORF includes the intron, and functional Cup1 protein can only be expressed from the pre-mRNA, thus *LCUPIF* splicing inhibition leads to increased copper tolerance (Methods and Fig. 2B). Consistent with our prediction, Fig. 2C shows that *LCUPIF* increases copper resistance (>0.1 mM) to a strain engineered to produce

excess L30 (yJV25), while *LCUPIF* C9U does not. This is further supported by the analysis of the splicing pattern of *LCUPIF* pre-mRNA transcripts, shown in Fig. 2E (lanes 1, 2). Therefore, we anticipated that cells with the *LCUPIF* C9U reporter would grow in copper concentrations higher than 0.1 mM if they bear a mutation enhancing L30 repression of splicing. To select such mutants, yJV25 cells carrying *LCUPIF* C9U were UV-irradiated and selected on 0.3 mM Cu^{2+} (Fig. 2D and Methods). Forty-six copper-resistant colonies were further tested (see Methods and Sup. Fig. S2) and six mutants showed splicing repression of this reporter (Fig. 2E, lanes 3-9), consistent with our screen selection. These mutants were called SLR (Supressors of Lack of Repression by L30). They all exhibit similar Cu^{2+} tolerance with *LCUPIF* C9U (Sup. Fig. S2).

To test more specifically the effects on splicing repression of our mutants, the aim of this work, we used the *LCUP* reporter containing the 5A C9U *RPL30* kink-turn, which is stably bound by the protein in wild type cells but remains unrepressed (Fig. 1 and Fig. 2A). Under excess L30 this reporter should lead to more tolerance to copper in wt than in SLR cells, in which splicing of 5A C9U transcripts is expected to be repressed. Indeed this was validated by copper tolerance and Northern analyses (Fig. 3B and Sup. Fig. S3). Importantly, as our SLR mutations are recessive (data not shown) introduction in these cells of the corresponding wt SLR allele should restore full splicing of the 5A C9U reporter, allowing their selection by copper tolerance (Fig. 2D). Using this approach, a YCp50 plasmid containing the *CBP80* gene was found to restore wt levels of copper tolerance to SLR5 and SLR7 (see Methods). Northern analyses further verified that in these cells the repression of splicing by L30 is similar to that in wt, as indicated in Fig. 2F (even lanes). We determined that the genomic copy of *CBP80* includes the mutation L157P in SLR5, and D291Stop in SLR7 (Sup. Fig. S4).

Deletion of *CBP80* suppresses the requirement for C9 by L30

The isolation of a *cbp80* allele with a non-sense mutation prompted us to verify the Cbp80 levels in SLR5, where the protein is not truncated. Cbp80 levels from wt, *cbp20*Δ, and SLR5 cells were assessed by Western blot. As Fig. 3A shows, L157P mutation leads to a drastic reduction of Cbp80 protein levels. This suggests that deletion of *CBP80* in yJV25 could produce an SLR phenotype. To assess this possibility we determined that the Cu^{2+} tolerance of yJV25 *cbp80*Δ cells, bearing a *LCUP* 5A C9U reporter, was indistinguishable from that of SLR5 (Fig. 3B), and we verified by Northern analysis that in *cbp80*Δ cells splicing of this reporter was repressed by L30 (Fig. 3C). Therefore we conclude that absence of functional Cbp80 is the cause of augmented splicing regulation by L30 on a 5A C9U transcript.

Next, to test the possible role of Cbp80 (or the CBC) on the regulation of splicing by L30 of wt transcripts (C at position 9), we analyzed splicing of a wt reporter in *cbp80*Δ, *cbp20*Δ, and *cbp80*Δ *cbp20*Δ (*cbc*Δ) strains. RNA from cells transformed with pLGFP (based on pLCUP, with the ORF of *CUP1* replaced by GFP (Macías et al., 2008)), under either normal or excess L30 conditions, was extracted and subjected to Northern analysis. There is a marked increase in *LGFP* pre-mRNA levels in all strains under excess L30 (Fig. 3D, even lanes), indicating that L30 effectively represses splicing of a wild-type *RPL30* kink-turn (C in position 9, Fig. 1A) in the *cbp80*Δ, *cbp20*Δ, *cbc*Δ backgrounds.

Depletion of the CBC complex affects mRNA levels of transcripts with non-consensus intronic ends (Fortes et al., 1999b; Lewis et al., 1996), and we asked whether deletion of just Cbp80 had the same effect. We measured the mRNA levels of *RPS13A*, *RPS11A*, *RPS11B*, and *RPL30* (transcripts with non-consensus intronic ends) in *cbp80*Δ, *cbp20*Δ, and *cbc*Δ strains by Northern analysis. Our data (Sup. Fig. S5) are consistent with previous reports showing a marked decrease in mRNA levels, especially those of *RPS11B*.

Thus, absence of a functional CBC affects processing of several genes and alters repression of *RPL30* splicing by L30.

Effects of Cbp80 deletion on co-transcriptional spliceosome assembly and L30 regulation

During control of splicing by L30, U2 recruitment is abolished and U1 association with the intron becomes stabilized (Macías et al., 2008). Our data, showing enhancement of L30 repression by deletion of Cbp80, raise the possibility of a novel role for CBC on U2 recruitment. CBC is known to act on U1 snRNP function during spliceosome assembly. First, on the initial recognition of the intron (Colot et al., 1996; Lewis et al., 1996; Tardiff et al., 2006). Subsequently, after U2 snRNP recruitment, on U1 release concomitant to U6 snRNP association (Gornemann et al., 2005; O'Mullane and Eperon, 1998). To determine the interaction between CBC and L30 we decided to monitor co-transcriptional splicing regulation by L30 in wt and *cbp80*Δ cells. For this we used the *RPL30-LacZ* reporter, based on *RPL30* but with the *LacZ* ORF as the second exon. This provides a convenient tool to assess co-transcriptional spliceosome assembly and its regulation by ChIP (Macías et al., 2008). The reporter was introduced in strains either producing or not excess L30 (from plasmid pMB73 (Macías et al., 2008)). ChIP analyses from these cells are shown in Fig. 4.

We first verified that deletion of Cbp80 does not lead to an increased co-transcriptional recruitment of L30, which could otherwise be consistent with enhanced regulation in *cbp80*Δ cells (Fig. 4B). Next we followed co-transcriptional engagement of U1 snRNP on *RPL30-LacZ* (ChIP of the U1 component Snu71-HTB (Macías et al., 2008)). In wt cells, a persistence of U1 on the transcript can be observed under excess L30 (Fig 4C, left, grey line). In a *cbp80*Δ strain, the maximum U1 ChIP signal is reduced, but it persists towards the 3' end of the gene (Fig. 4C, right, black line). Under excess L30,

ChIP of U1 shows a slight reduction and a similar persistence (Fig. 4C, right, grey line).

Co-transcriptional recruitment of U2 in wt cells (ChIP of the U2 component Lea1-HTB (Macías et al., 2008)) is strongly repressed by excess L30 (Fig. 4D, left, gray line). In *cbp80*Δ cells, U2 recruitment is diminished (Fig. 4D, right, black), with a further signal reduction under excess L30 (Fig. 4D, right, grey). Taking into account the persistence of the U1 ChIP profile and the low U2 recruitment observed in *cbp80*Δ cells, our data suggest that in the *RPL30* intron co-transcriptional engagement of U2 is dependent on Cbp80. Consistent with this, deletion of Cbp80 leads to some pre-mRNA accumulation in the absence of excess L30 (Fig. 3, lanes 1, 3).

Our ChIP results on L30 repression in wt cells are in agreement with our previous data (Macías et al., 2008). In addition, a persistence of U1 ChIP on the target gene has also been described in *cbc*Δ cells (Gornemann et al., 2005). However, our U2 ChIPs from *cbp80*Δ cells differ from what has been described for the CBC regarding co-transcriptional recruitment of U2 on *ECM33* transcripts, where it does not change (Gornemann et al., 2005). To assess this apparent discrepancy, we analyzed the co-transcriptional recruitment of U2 snRNP on the *ACT1* gene, which is not affected by L30 overexpression, in a *cbp80*Δ strain. Neither deletion of *CBP80* nor excess L30 have a significant effect on co-transcriptional recruitment of either U1 or U2 snRNP on the *ACT1* intron (Fig. 4F-G). This is consistent with previously published results (Gornemann et al., 2005) and indicates that the role of Cbp80 in co-transcriptional spliceosome assembly is transcript-specific.

Cbp80 and 5' splice site sequences impact U1 and U2 snRNP co-transcriptional recruitment

The persistence of U1 on the *RPL30* transcript in *cbp80*Δ cells (Fig. 4C) can reflect a limited co-transcriptional recruitment of U2 (Lacadie and Rosbash, 2005; Macías et al., 2008; Tardiff and Rosbash, 2006), suggesting a role for

Cbp80 in this recruitment. To further dissect the contribution of Cbp80 to U1 and U2 recruitments we investigated whether sequence features in the *RPL30* or *ACT1* introns could affect the role of Cbp80 in U1 and U2 co-transcriptional recruitment. A particular attribute of the *RPL30* intron is the evolutionarily conserved 5'SS sequence, GUCAGUAU, unique in yeast introns (Grate and Ares, 2002). This sequence can potentially form seven base-pairs with U1 snRNA, while the prototypical *ACT1* intron, with GUAUGUUC at its 5' end, only has the potential to form five (Fig. 5A). Thus we investigated the role of these sequences and Cbp80 on spliceosome assembly in the context of the *RPL30-LacZ* transcript. The mutants GUauGUuc, containing the first 8 nucleotides of *ACT1*, and GUauGUAU, with positions 3 and 4 as in *ACT1* (lowercase indicates changes in wt *RPL30* to *ACT1*), were compared to the wt version GUCAGUAU (Fig. 4C). These mutations have been shown not to affect *RPL30* splicing in wt cells (Sup. Fig. S6 and (Macías et al., 2008)).

The results, shown in Fig. 5, can be summarized as follows. In wt cells, U1 co-transcriptional recruitment is influenced not only by the potential base pairing between the intron and U1, but also by the particular 5'SS sequence (panel 5A, black). Thus, GUauGUAU (middle) produces a ChIP signal more than three times higher than that of wt *RPL30* (left). This difference is striking considering that the potential to form base-pairs with U1 is similar in both constructs (Fig. 5). Consistent with this, disrupting the potential base pairing of U1 to positions 7 and 8 (GUauGUuc) produces intermediate levels of U1 ChIP (right), despite having less potential for base-pairing with U1 than *RPL30* 5'SS (left), indicating that AU at positions 3 and 4 facilitate U1 recruitment. The corresponding U1 ChIPs in *cbp80*Δ cells indicate that Cbp80 is required to take full advantage of the increased recruitment afforded by the extra base pairing of U1 to intronic positions 7 and 8 (panel A, grey lines). Thus, deletion of Cbp80 has a greater effect on the intron with GUauGUAU than on the one with GUauGUuc at the 5'SS, to the point that both reporters

recruit similar levels of U1 in absence of Cbp80 (middle and right, respectively).

U2 co-transcriptional recruitment in wt cells (panel 5B, black) shows efficiencies that do not necessarily correlate with those of U1. While U2 associates better with a GUauGUAU intron (center) than with *RPL30* (left), the improvement does not match that of U1. Accordingly, U2 ChIPs on GUauGUuc (right) and GUauGUAU (middle) introns are similar, despite having distinct U1 ChIPs.

The absence of correlation between U1 and U2 recruitments is also apparent in *cbp80*Δ cells (panel 5B, grey lines). Interestingly, this is more marked in the GUauGUuc intron (left), where the ChIP profile remains closer to that of wt cells, compared to GUauGUAU (middle). Importantly, both introns display similar U1 ChIP profiles in *cbp80*Δ cells (panel A). These data indicate that the role of Cbp80 in U2 recruitment becomes more apparent in introns with increased base pairing to U1, consistent with the low recruitment of U2 in GUCAGUAU introns in *cbp80*Δ cells (left). We conclude that Cbp80 has roles in U1 as well as U2 co-transcriptional recruitment, and these become more evident when the interaction between U1 and the intron is hyperstabilized. We verified that the introduced mutations do not induce alterations in pol II recruitment (Sup. Fig. S7) and we ascertained that the ChIPs on *ACT1* remained similar across the samples (Sup. Fig. S8).

To verify the generality of our conclusions, we measured U1 and U2 co-transcriptional recruitment to other yeast transcripts with a similar 5' SS to that of *RPL30*. There are six yeast genes with GUACGUAU at the 5'SS (Grate and Ares, 2002), thus having the same potential to base-pair to U1. Most of them produced poor ChIP data in wt cells under our conditions (data not shown), except *RPS13A*. According to our model, deletion of *CBP80* should affect U1 recruitment, as Cbp80 is required to take advantage of the

extra base pairing of U1 to intronic positions 7 and 8. Consequently, U2 co-transcriptional recruitment could be further diminished in *cbp80*Δ cells due to a possible role for Cbp80 in U1 remodelling. Our ChIP data are shown in Fig. 6B and support these predictions. Interestingly, *RPS13A* has an altered BS, compared to the consensus (GACUAAC instead of UACUAAC), which may diminish the stability of the interaction between U1 and the intron (Berglund et al., 1997; Seraphin and Rosbash, 1991), and thus explaining the reduced requirement for Cbp80.

DISCUSSION

Data on the control of splicing by L30 support a novel strategy based on interfering with spliceosomal transitions during U2 snRNP association with the intron (Macías et al., 2008). In a genetic screen for mutations that alter this regulation we have identified Cbp80. This factor, together with Cbp20, form the CBC, which is linked to the recognition of the 5'SS during both early and late steps in spliceosome assembly((Gornemann et al., 2005) and references therein). Yet our genetic analysis suggests an additional role for CBC in U2 recruitment, and we have carried out experiments to test this hypothesis. In this manuscript we show that U2 snRNP recruitment is assisted by Cbp80, is influenced by the sequence next to the 5' SS, and can be targeted by splicing regulatory factors like L30.

A mutation in the exon 1 of RPL30 that restores U2 snRNP recruitment requires Cbp80

We have determined that the C9U *RPL30* mutation not only weakens binding to the L30 protein (Fig. 1C) but also abolishes regulation of splicing even under conditions where L30 binding is restored (Fig. 1D). It is possible that the spliceosomal machinery can compete with L30 binding when one of the stems of the *RPL30* kink-turn is destabilized. Consistent with this, our data indicate that L30 remains associated with the 5A C9U transcript during intron

definition (Fig 1E), but fails to co-immunoprecipitate with U2 (Fig. 1F). U2 recruitment involves the initial recognition of the BS by BBP, which also interacts with U1 snRNP (Abovich et al., 1994; Berglund et al., 1997). A conformational change is required to proceed with U2 binding, and it seems plausible that L30 is dislodged from C9U mutant RNAs at this step.

These results indicate that the effects of C9U on L30 binding and splicing regulation can be separated, and we took advantage of this set up to isolate mutants that restore splicing regulation in the context of the combined C9U and 5A mutations. Using this approach we have found that two *cbp80* alleles, L157P and D291Stop (Sup. Fig. S4), suppress the phenotype of the *RPL30* C9U mutant. These mutations lead us to verify that the deletion of *CBP80* leads to an SLR phenotype as well (Fig. 3B, 3C), without blocking repression of splicing of wild type (C9) *RPL30* transcripts (Fig. 3D).

Two roles have been proposed for CBC at different steps during spliceosome assembly, both related to the interaction between the 5'SS and U1. There is evidence that CBC promotes the recognition of the 5'SS by U1 snRNP (Colot et al., 1996; Fortes et al., 1999b; Lewis et al., 1996). CBC also plays a role in later steps during U6 snRNP association with the intron (Gornemann et al., 2005; O'Mullane and Eperon, 1998). Because L30 repression occurs after U1 is recruited but before 5' splice site recognition by U6 snRNP, the molecular relationship between Cbp80 and L30 is not clear. It is unlikely that the effect of *CBP80* deletion is caused simply by a reduction of splicing activity, since other mutations in splicing factors do not display the same phenotype (data not shown).

Therefore to delineate the molecular basis for the striking synergism between C9U and Cbp80, we have analyzed in detail U1 and U2 co-transcriptional spliceosome assembly in *cbp80*Δ cells. Our results show that Cbp80 plays a significant role in co-transcriptional U2 recruitment when the potential base

pairing between the intron and U1 snRNP is unusually strong. This helps to explain the genetic interaction of Cbp80 with L30 regulation, as *RPL30* is an effective U1 snRNP recruiter (Fig. 6 and (Tardiff et al., 2006)). Regarding this U1 recruitment, comparison of ChIP profiles from GUAUGUAU and GUCAGUAU introns reveals a critical role for intronic sequences in the vicinity of the 5' splice site, even when the number of potential base pairs to U1 snRNA does not change. This is in agreement with previous reports indicating that initial recognition of an intron by U1 is not limited by base pairing (Du and Rosbash, 2001), and that GUAUGU could be favored over GUCAGU by the U1 snRNP factor U1-C (Du and Rosbash, 2002). Interestingly, however, there is no apparent change in the splicing efficiency of both transcripts, or in their regulation (Sup. Fig. S6). Therefore, this increased recruitment does not necessarily result in more mRNA accumulation, which will be determined by further steps downstream in the processing pathway.

The effective recruitment of U1 to the *RPL30* intron depends on both its sequence and Cbp80 (Fig. 5A), since in *cbp80*Δ cells U1 ChIP on GUAUGUAU and GUAUGUUC introns are equivalent (Fig. 5A). Thus, our results indicate that positions 3-4, and 7-8, have different effects on U1 recruitment. They also reveal that the contribution of positions 7-8 are mostly dependent on Cbp80 function, in contrast to that of positions 3-4.

Our ChIP data indicate that U2 recruitment does not always correlate with that of U1, arguing that co-transcriptional association of U2 does not simply follow that of U1. Moreover, it can be a limiting factor for splicing, subjected to regulation, as in the *RPL30* intron. For example, while mutation of GUAUGUAU to GUAUGUUC in *cbp80*Δ cells does not have a significant effect on U1 co-transcriptional recruitment, there is a clear reduction in the U2 ChIP on a AU intron (Fig. 5 panel B). This suggests that in *cbp80*Δ cells hyperstabilized U1 binding interferes with U2 recruitment. Consistent with

this hypothesis, our U1 and U2 ChIP analyses on the *RPS13A* transcript that has a 5'SS similar to *RPL30* (GUACGUAU) indicate a clear reduction of U2 recruitment in *cbp80*Δ cells. It is worth noting that *RPS13A* has a non-consensus BS (Fig. 6A), which may interfere also with U2 binding. Yet the effect of Cbp80 is still noticeable, suggesting that it takes place before the recognition of the branch site by U2 snRNP (Fig. 6). This situation is reminiscent of the effect of hyperstabilized U1 binding to the 5'SS (Chen et al., 2001; Staley and Guthrie, 1999) on U4/U6.U5 tri-snRNP recruitment, and similarly points towards a U1 snRNP remodeling prior to U2 recruitment.

There are genetic and molecular data supporting this U1 remodeling. Genetic interactions between U1 snRNP, CBC, and the BS, have been observed (Fortes et al., 1999a) . In addition, mutations in U1 predicted to weaken the interaction with the 5'SS allow for some U2 recruitment in absence of ATP (Liao et al., 1992). Furthermore, recent findings on the structure of the human U1 snRNP bound to the 5'SS (Pomeranz Krummel et al., 2009) argue in favor of this possibility as well. The factor U1-C stabilizes the interaction U1-5'SS while being subjected to long-range protein connections within the particle (Pomeranz Krummel et al., 2009). Perhaps remodeling of the bridging between U1 snRNP and the BS affects the binding of U1 to the 5'SS, an interaction clearly influenced by CBC (Gornemann et al., 2005). In this context, the genetic interaction between Cbp80 and L30 could be explained by their antagonistic roles on spliceosome assembly (Sup. Fig. S9). L30 inhibits U2 recruitment, stabilizing a particular conformation of U1 on the intron, while Cbp80 promotes U2 recruitment and a likely U1 remodeling. In a C9U mutant, L30 binding may not be stable enough to repress unless Cbp80 is absent. Thus, Cbp80 would be facilitating the conformational change that competes with L30 binding or that L30 inhibits.

Cbp80 and regulated splicing

Given its multiple roles in RNA processing (including an indirect effect on histone ubiquitination (Hossain et al., 2009)), deletion of Cbp80 can have diverse effects on different transcripts, depending on when its function is more critical. Thus, while early spliceosome events on *ACT1* are not severely affected in *cbp80*Δ cells (Fig. 4); microarray analyses (Clark et al., 2002) indicate strong changes in *ACT1* splicing in these cells, suggesting an effect on later stages of splicing. Consistent with this, microarray data from *cbp80*Δ cells correlate well with those from cells with deletions of factors involved either in U2 recruitment (Mud2/U2AF, Msl1/U2B), or U4/U6.U5 tri-snRNP function (Snu66; Fig. 2 of (Clark et al., 2002)). This is also in agreement with the ChIP data showing a role for Cbp80 in early and late stages of spliceosome assembly (Gornemann et al., 2005). Furthermore, the same analyses indicate that some of the most affected transcripts in *cbp80*Δ cells are subjected to regulated splicing, and do not have consensus branch site sequences. One example is the*YRA1* transcript, with highly regulated splicing in a scheme that includes the link between splicing and RNA export (Dong et al., 2007; Preker and Guthrie, 2006), processes both affected by CBC. Therefore, the involvement of Cbp80 in steps involving the recognition of the 3' end of the intron by U2 snRNP could be targeted as well by several other strategies of regulation.

Our findings may have implications in mammalian systems as well, with higher variability in the splicing signals and more regulated splicing than in yeast. CBC is part of the first exon definition complex (Berget, 1995), and as such it may play a relevant role in splicing decisions involving alternative 5’ and 3’ splicing sites in this intron. Moreover, as CBC has been shown to affect co-transcriptional events downstream of the first exon (Wong et al., 2007), it is not unconceivable that the described modulation of U1 and U2 co-transcriptional recruitment affects a variety of splice site choices along mammalian transcripts.

FIGURES AND FIGURE LEGENDS

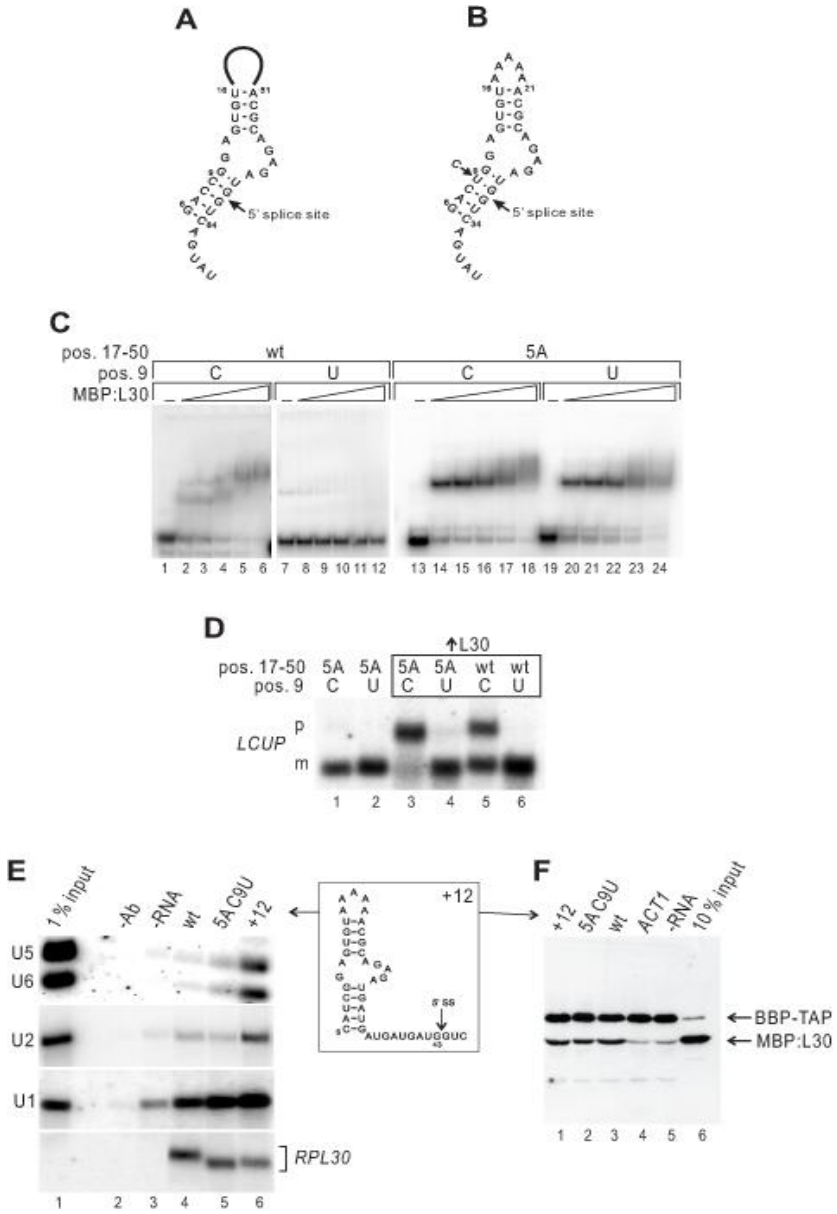


Figure 1 Bragulat et al. L30 binds the *RPL30* C9U transcript. (**A**) Schematic representation of the wt *RPL30* kink-turn (long arrow denotes the 5' splice site, just after the initiation codon). Nucleotides 17-50 are not

required for L30 binding. (**B**) Schematic representation of the 5A C9U *RPL30* kink-turn. Nucleotides 17-50 in the wt transcript (**A**) are replaced by 5 adenines. This stabilizes the kink-turn and favors L30 binding (Mao et al., 1999). The C9U mutation is indicated. (**C**) Effect of mutations in the 5' UTR of *RPL30* in L30 binding. *RPL30* transcripts (0.5 picomoles, nucleotides 1-123) were incubated with buffer (lanes 1, 7, 13, 19) or with increasing amounts of MBP:L30 (50, 100, 200, 500, 1000 ng; lanes 2-6; 8-12; 14-18; 20-24, respectively). Transcripts were wt (lanes 1-6), C9U (7-12), 5A (13-18), 5A-C9U (19-24). Reactions were subsequently analyzed in a 6% non-denaturing acrylamide gel. (**D**) Both C9U and 5A-C9U transcripts fail to accumulate pre-mRNA under *in vivo* conditions of L30 excess. W303 (lanes 1-2) or yJV25 (producing excess L30, see Methods; lanes 3-6) cells were transformed with pLCUP plasmids with a kink-turn bearing the indicated mutations (see panel A). RNA was extracted, subjected to Northern analysis and probed with *LCUP* sequences. Under normal conditions 5A and 5A-C9U transcripts splice efficiently (lanes 1-2). Under excess L30, wt (lane 5) and 5A (lane 3) transcripts are repressed, with 5A displaying a strong inhibition of splicing. Mutation C9U disrupts this repression, both in the wt (lane 6) and the 5A (lane 4) transcripts. (**E**) BBP recognizes the branch site of the *RPL30* 5A C9U intron in presence of MBP:L30. Indicated *RPL30* transcripts (nt 1-347) were incubated under splicing conditions (BBP-TAP extracts) with MBP:L30. Reactions were immunoprecipitated with IgG. Proteins that co-immunoprecipitated with BBP-TAP were analyzed by Western blotting, as indicated on the left. Transcripts were *ACT1* (lane 3); *RPL30* (lane 4); *RPL30* 5A C9U (lane 5); +12, a synthetic positive control based on *RPL30*, bound by L30 but not repressed, with the RNA binding motif separated by 12 nucleotides from the 5'SS, schematized at the right (lane 6). Lane 1, 10% of input; lane 2, no transcript added. (**F**) snRNP co-immunoprecipitation with L30. *RPL30* transcripts (nt 1-347) were incubated under splicing conditions and ATP with MBP:L30. Reactions were immunoprecipitated with anti-MBP, and pelleted RNA was subjected to Northern analysis to detect *RPL30*, U1

and U2 snRNA, as indicated. Lane 1, *RPL30* +12 (scheme on the left); lane 2, *RPL30* 5A C9U; lane 3, *RPL30*; lane 4, no transcript added; lane 5, no antibody added; lane 6, 1% of the input.

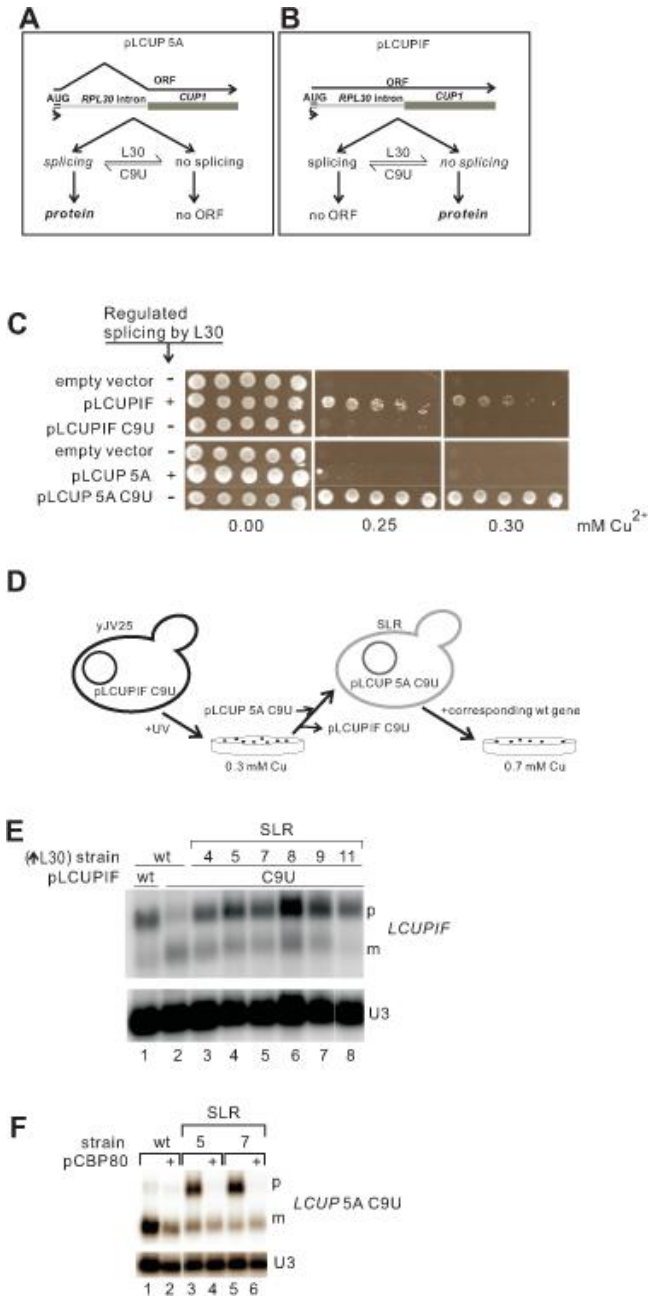


Figure 2 Bragulat et al. Screen for synthetic enhancers of L30 repression of splicing. (**A**)(**B**) Reporter plasmids based on the fusion between *RPL30* exon1 and intron with the *CUP1* ORF. *LCUP* RNAs (**A**) need to be spliced to encode the protein, while transcripts *LCUPIF* (**B**) produce Cup1 protein only

when unspliced. (**C**) Phenotype, identified as growth in medium containing copper, of cells with a constitutive excess of L30 (yJV25) and transformed with either pLCUPIF (upper panel) or pLCUP 5A (bottom panel). Under repression conditions ("+" rows), pLCUPIF confers copper resistance while pLCUP-5A does not. When L30 repression is abolished by the C9U mutation ("-" rows), pLCUP confers tolerance while pLCUPIF does not. Serial 1/5 dilutions where spotted in each case. (**D**) Screen strategy to select mutations that restore inhibition of splicing by L30 on a C9U transcript. Corresponding Northern analyses are shown in panels D and E. Strain yJV25 with the plasmid pLCUPIF-C9U (Cu-sensitive) was UV-irradiated and SLR mutants were selected on plates containing 0.3 mM copper. Colonies showing *LCUPIF* C9U pre-mRNA accumulation (panel D) were cured of the plasmid and transformed with pLCUP 5A C9U, rendering them again Cu-sensitive, because of increased repression, unless the *slr* mutation is complemented or suppressed. Thus, cells were transformed with a YCp50-based wt genomic library and the transformants selected on 0.7 mM copper, and pCBP80 was identified (panel F). (**E**) SLR mutants repress splicing of a *LCUPIF* C9U transcript, as seen by Northern analysis of RNA from SLR mutant cells transformed with pLCUPIF C9U (lanes 3-8). As a control, RNA extracted from yJV25 cells bearing a pLCUPIF wt (lane 1) or C9U (lane 2) were loaded in the same gel. U3 was used as loading control. (**F**) SLR5 and SLR7 are suppressed by *CBP80*. A library plasmid containing *CBP80*, isolated following the strategy depicted in (**D**) using SLR5, restores the wt phenotype in SLR7 mutants. SLR mutants (lanes 3-6) and the yJV25 strain (lanes 1-2) were transformed with the plasmid pLCUP 5A C9U (odd lanes) and pCBP80 (even lanes). RNA was extracted and subjected to Northern analyses, as indicated. U3 was used as loading control.

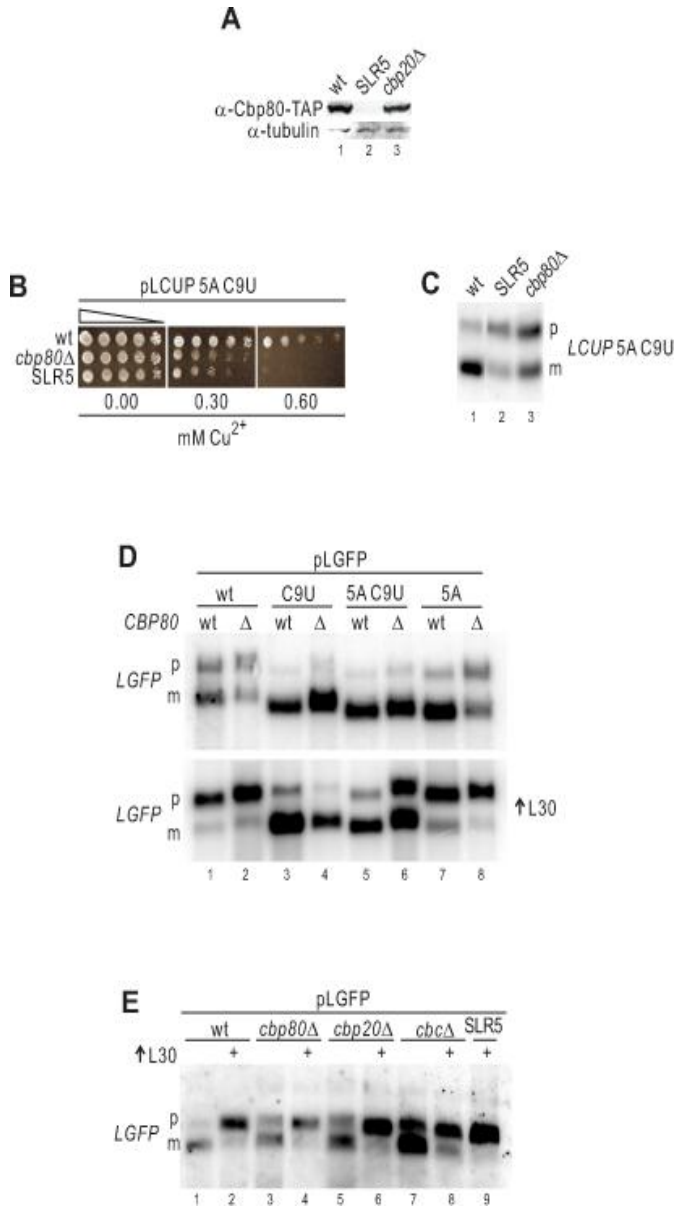


Figure 3 Bragulat et al. Deletion of *CBP80* is synthetic with L30 repression of splicing. (**A**) Decreased levels of Cbp80-TAP in SLR5 mutants. Cbp80 protein was TAP-tagged at the C-terminus in yJV25 (wt, lane 1), SLR5 (lane 2), and *cbp20*Δ (Y02074). Extracts were subjected to Western analyses, as indicated. Tubulin was used as loading control. (**B**) Deletion of *CPB80* produces the same phenotype as that of SLR5. 1/5-serial dilutions of wt

(yJV25, top), *cbp80*Δ (yJV35, middle) and SLR5 (bottom) cells transformed with pLCUP 5A C9T were spotted on copper-containing media, as indicated. Copper sensitivity denotes splicing repression of the *LCUP* 5A C9U transcript. (**C**) Deletion of *CBP80* leads to repression of *LCUP* 5A C9U splicing by L30. Northern analysis of RNA from wt (yJV25, lane 1), SLR5 (lane 2) and *cbp80*Δ yJV42, lane 3) cells. Positions of precursor (p) and mature (m) *LGFP* 5A C9U are indicated on the right. (**D**) Regulation of wt *RPL30* transcripts in *cbp80*Δ, *cbc20*Δ, *cbc*Δ, and SLR5 cells. Northern analysis of RNA extracted from strains transformed with the pLGFP plasmid, as indicated at the top, under excess L30 (pMB73, lanes 2, 4, 6, 8 and SLR5 in lane 9) or no excess (lanes 1, 3, 5, 7). Positions of precursor (p) and mature (m) *LGFP* are indicated on the left. pLGFP contains the GFP ORF replacing that of Cup1 from pLCUP (Vilardell and Warner, 1997). pMB73 encodes L30 without establishing the auto-regulatory loop (Macías et al., 2008).

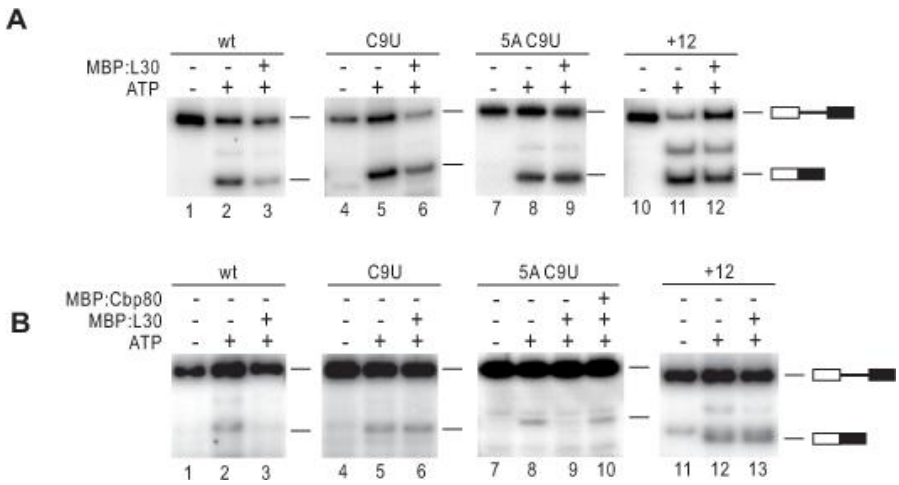


Figure 4 Bragulat et al. Co-transcriptional U1 and U2 recruitment on the *RPL30-LacZ* and *ACT1* genes in wt and *cbp80*Δ cells. The horizontal axis shows the distance in nt from the start codon. The vertical axis indicates the signal relative to that of the promoter (first primer pair, or PP1). The black bar indicates intron position. The ChIP profiles correspond to wt (left panels in each pair) or *cbp80*Δ cells (right panels), under normal conditions (black lines) or under L30 excess (grey lines). (**A**)(**E**) Scheme showing the positions of the PCR primers used for the ChIP analyses of the *RPL30-LacZ* (**A**) or *ACT1* (**E**) introns, relative to the translation start. (**B**) ChIP against L30 (L30-TAP). In both cases there is L30 excess (pMB73, see Methods). Black line in wt cells, grey in *cbp80*Δ cells.

In the following panels, ChIP profiles of the indicated proteins are shown, performed on *RPL30-LacZ* or *ACT1*, respectively. Sup. Fig. S8 shows another *ACT1* ChIP on *CBP80* and *cbp80*Δ cells. (**C**)(**F**) ChIP against U1 snRNP (Snu71-HTB). (**D**)(**G**) ChIP against U2 snRNP (Lea1-HTB).

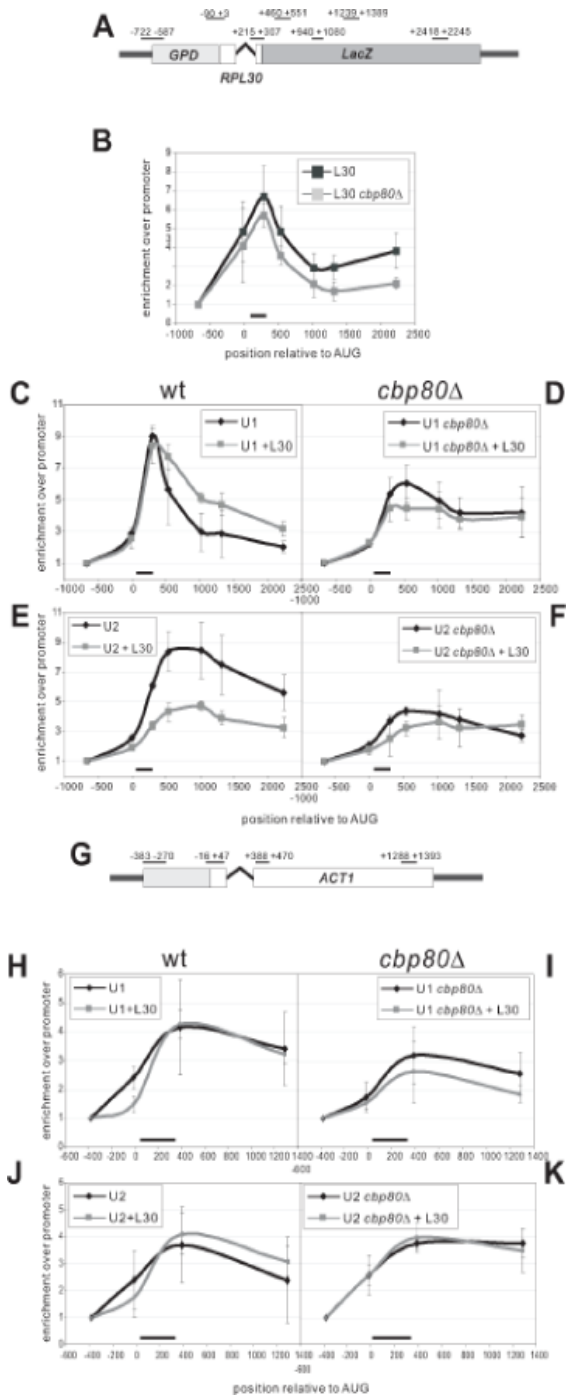


Figure 5 Bragulat et al. Effect of mutations in the *RPL30* intron and Cbp80 on co-transcriptional recruitment of U1 and U2 on the *RPL30-LacZ* gene Horizontal axes show the distance in nucleotides from the start

codon. Vertical axes indicate the signal relative that of the promoter (first primer pair, or PP1). The black bar indicates intron position. The ChIP profiles correspond to wt (black lines) or *cbp80*Δ cells (grey lines). Panels indicate different intronic 5' ends, and on the top of each there is a schematic representation of the possible base-paring with U1 snRNA. GUCAGUAU panels are based on data from Fig. 4. (**A**) ChIP profiles of U1 snRNP (Snu71-HTB). (**B**) ChIP profiles of U2 snRNP (Lea1-HTB).

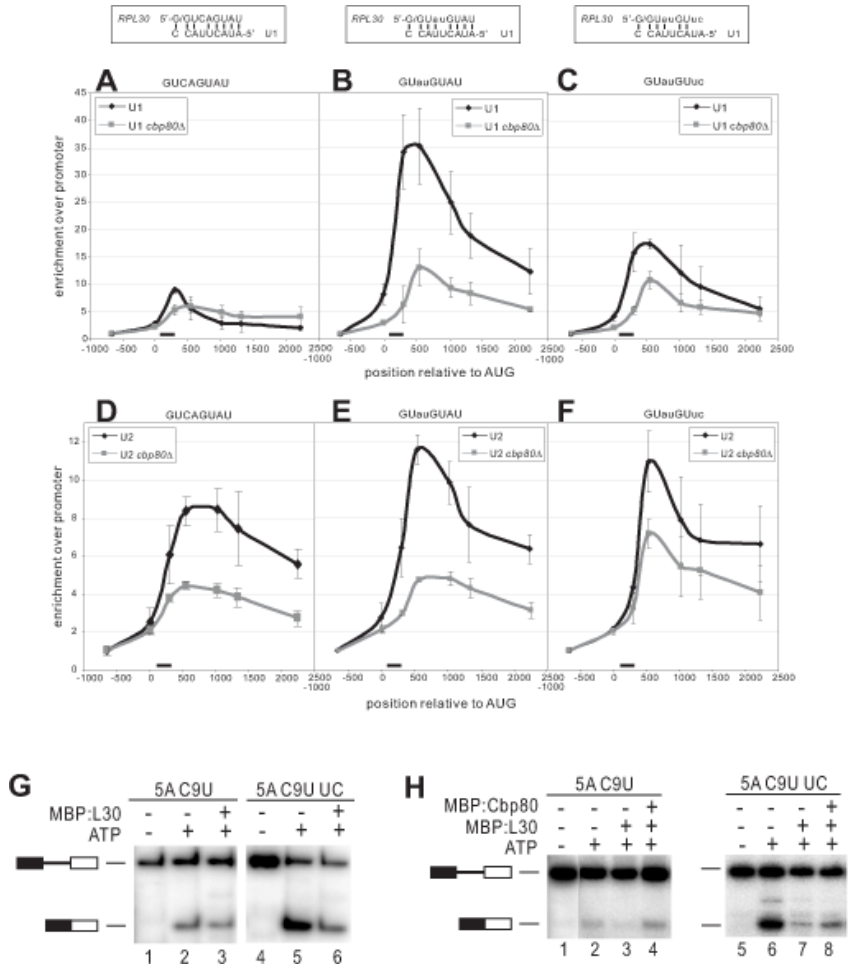


Figure 6 Bragulat et al. Effect of Cbp80 on the co-transcriptional recruitment of U1 and U2 on *RPS13A*. Graphs for the *RPL30* intron are based on data from Fig. 4. (**A**) Sequences at the 5' splice site (5'SS) and branch site (BS) of *ACT1*, *RPS13A*, and *RPL30*. Bases with potential base-pairing to U1 or U2 snRNAs are indicated in bold. (**B**) U1 snRNP (Snu71-HTB, light grey) and U2 snRNP (Lea1-HTB, dark grey) levels on exon 2 of the indicated transcripts (bottom) were normalized to those at the promoter.

EXPERIMENTAL PROCEDURES

Yeast Strains and Plasmids

Yeast strains and plasmids used in this study are described in Supplemental Table 1.

Screen for synthetic enhancers of L30 repression of splicing

10^8 yJV25 cells containing pLCUPIF C9U were UV-irradiated at different doses up to 200J/m^2. Mutants were selected on plates containing 0.3 mM copper. A total of 46 copper-resistant clones were retransformed with fresh pLCUPIF C9U plasmid, and the growth in copper and processing of the *LCUPIF* transcript verified. Six mutants (Fig. 2) were finally selected based on pre-mRNA accumulation. To identify the mutation, mutant SLR5 was cured of pLCUPIF C9U, and transformed with the reporter plasmid pLCUP 5A-C9U and a YCp50-based genomic library. SLR5 represses splicing of the *LCUP* 5A-C9U transcript and it is sensitive to 0.7 mM Cu^{2+}, unless it acquires the wt *SLR5* gene, which restores unregulated splicing of a C9U transcript. With this approach sixty colonies were selected by growth in 0.7 mM Cu^{2+}. Of those, 12 did not contain plasmids bearing the *CUP1* gene. Reintroduction of those plasmids into SLR5 cells and subsequent Northern analyses allowed the isolation of two independent plasmids containing the *CBP80*/*STO1* gene. The genomic copy of *CBP80* in SLR5 was sequenced, and thus verified as an allele of *STO1/CBP80* (L157P). Fig. 2D shows a scheme of this strategy. Subsequently, the *CBP80* genes of the remaining SLR mutants were sequenced and SLR7 was found to have another *STO1* allele, with a D291Stop mutation. ORFs of *RPL30*, *MUD13/CBP20*, and *SNR19* (U1), *SNR20* (U2) were verified in all SLR mutants by sequencing, and found to be wt.

Northern analyses, immunoprecipitations and *in vitro* assays

Splicing extracts, Western and Northern assays were performed as described in (Macías et al., 2008). Immunoprecipitations were done as in (Macías et al., 2008), where heparin was not used in the washes. Anti-tubulin antibody was from Sigma. Recombinant MBP:L30 was prepared as in (Vilardell and Warner, 1994). Mobility gel-shift assays were performed as in (Vilardell and Warner, 1994).

ChIP

Chromatin immunoprecipitation and quantitative PCR assays were performed as described in (Macías et al., 2008), using the same primer pairs. Error bars are based on at least two independent biological replicas. *RPL30-LacZ* mutants described in Fig. 5 were generated by *in vivo* gap-repair cloning, via co- transformation of pJV44 digested with Bam/Spe and a PCR fragment containing the corresponding mutations in the *RPL30* intron. Positive colonies were sequenced for verification. ChIPs in Fig. 6 were performed with primers pairs in the promoter and in the exon 2 (see Supplemental Table S2 for a list of used primers).

ACKNOWLEDGMENTS

We are grateful to C. Query and J. Valcárcel for helpful discussions and critiques to the manuscript; and F. Gebauer, R. Méndez, and J. Warner for critiques to the manuscript. This research has been supported by the Spanish Ministry of Science (MEC) (BMC and BFU grants), a Marie Curie Contract (#510183), and Agaur. S. M. and M. M. are supported in part by a Training of Researchers (FI) fellowship (Agaur) and. J.V. is supported by a Ramón y Cajal/IP3 contract (MEC).

REFERENCES

Abovich, N., Liao, X. C., and Rosbash, M. (1994). The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. Genes Dev *8*, 843-854.

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell *89*, 403-412.

Bentley, D. L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol *17*, 251-256.

Berget, S. M. (1995). Exon recognition in vertebrate splicing. J Biol Chem *270*, 2411-2414.

Berglund, J. A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell *89*, 781-787.

Bleichert, F., and Baserga, S. J. (2007). The long unwinding road of RNA helicases. Mol Cell *27*, 339-352.

Brow, D. A. (2002). Allosteric cascade of spliceosome activation. Annu Rev Genet *36*, 333-360.

Chen, J. Y., Stands, L., Staley, J. P., Jackups, R. R., Jr., Latus, L. J., and Chang, T. H. (2001). Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Mol Cell *7*, 227-232.

Clark, T. A., Sugnet, C. W., and Ares, M., Jr. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science *296*, 907-910.

Colot, H. V., Stutz, F., and Rosbash, M. (1996). The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. Genes Dev *10*, 1699-1708.

Cordin, O., Banroques, J., Tanner, N. K., and Linder, P. (2006). The DEAD-box protein family of RNA helicases. Gene *367*, 17-37.

Dong, S., Li, C., Zenklusen, D., Singer, R. H., Jacobson, A., and He, F. (2007). YRA1 autoregulation requires nuclear export and cytoplasmic Edc3p-mediated degradation of its pre-mRNA. Mol Cell *25*, 559-573.

Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol *3*, 195-205.

Du, H., and Rosbash, M. (2001). Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA-pre-mRNA base pairing. RNA *7*, 133-142.

Du, H., and Rosbash, M. (2002). The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. Nature *419*, 86-90.

Eng, F. J., and Warner, J. R. (1991). Structural basis for the regulation of splicing of a yeast messenger RNA. Cell *65*, 797-804.

Fortes, P., Bilbao-Cortes, D., Fornerod, M., Rigaut, G., Raymond, W., Seraphin, B., and Mattaj, I. W. (1999a). Luc7p, a novel yeast U1 snRNP protein with a role in 5' splice site recognition. Genes Dev *13*, 2425-2438.

Fortes, P., Kufel, J., Fornerod, M., Polycarpou-Schwarz, M., Lafontaine, D., Tollervey, D., and Mattaj, I. W. (1999b). Genetic and physical interactions involving the yeast nuclear cap-binding complex. Mol Cell Biol *19*, 6543-6553.

Gornemann, J., Kotovic, K. M., Hujer, K., and Neugebauer, K. M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol Cell *19*, 53-63.

Grate, L., and Ares, M., Jr. (2002). Searching yeast intron data at Ares lab Web site. Methods Enzymol *350*, 380-392.

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197-1211.

Hertel, K. J., and Graveley, B. R. (2005). RS domains contact the pre-mRNA throughout spliceosome assembly. Trends Biochem Sci *30*, 115-118.

Hossain, M. A., Claggett, J. M., Nguyen, T., and Johnson, T. L. (2009). The cap binding complex influences H2B ubiquitination by facilitating splicing of the SUS1 pre-mRNA. RNA.

House, A. E., and Lynch, K. W. (2008). Regulation of alternative splicing: more than just the ABCs. J Biol Chem *283*, 1217-1221.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, I. W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell *78*, 657-668.

Jurica, M. S., and Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol Cell *12*, 5-14.

Kent, O. A., and MacMillan, A. M. (2002). Early organization of pre-mRNA during spliceosome assembly. Nat Struct Biol *9*, 576-581.

Klein, D. J., Schmeing, T. M., Moore, P. B., and Steitz, T. A. (2001). The kink-turn: a new RNA secondary structure motif. Embo J *20*, 4214-4221.

Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Munoz, M. J., and Nogues, G. (2004). Multiple links between transcription and splicing. RNA *10*, 1489-1498.

Lacadie, S. A., and Rosbash, M. (2005). Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. Mol Cell *19*, 65-75.

Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., and Mattaj, I. W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. Genes Dev *10*, 1683-1698.

Liao, X. C., Colot, H. V., Wang, Y., and Rosbash, M. (1992). Requirements for U2 snRNP addition to yeast pre-mRNA. Nucleic Acids Res *20*, 4237-4245.

Macías, S., Bragulat, M., Tardiff, D. F., and Vilardell, J. (2008). L30 binds the nascent RPL30 transcript to repress U2 snRNP recruitment. Mol Cell *30*, 732-742.

Mao, H., White, S. A., and Williamson, J. R. (1999). A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. Nat Struct Biol *6*, 1139-1147.

Meyer, M., and Vilardell, J. (2009). The quest for a message: budding yeast, a model organism to study the control of pre-mRNA splicing. Brief Funct Genomic Proteomic *8*, 60-67.

Moore, M. J., and Proudfoot, N. J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688-700.

O'Mullane, L., and Eperon, I. C. (1998). The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice sites. Mol Cell Biol *18*, 7510-7520.

Parker, R., Siliciano, P. G., and Guthrie, C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. Cell *49*, 229-239.

Pleiss, J. A., Whitworth, G. B., Bergkessel, M., and Guthrie, C. (2007). Transcript Specificity in Yeast Pre-mRNA Splicing Revealed by Mutations in Core Spliceosomal Components. PLoS Biol *5*, e90.

Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature *458*, 475-480.

Preker, P. J., and Guthrie, C. (2006). Autoregulation of the mRNA export factor Yra1p requires inefficient splicing of its pre-mRNA. RNA *12*, 994-1006.

Reed, R., and Palandjian, L. (1997). Spliceosome assembly. In Frontiers in Molecular Biology: Eukaryotic mRNA processing, A. R. Krainer, ed. (Oxford, IRL Press), pp. 103-129.

Rutz, B., and Seraphin, B. (1999). Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. RNA *5*, 819-831.

Seraphin, B., and Rosbash, M. (1991). The yeast branchpoint sequence is not required for the formation of a stable U1 snRNA-pre-mRNA complex and is recognized in the absence of U2 snRNA. Embo J *10*, 1209-1216.

Spingola, M., Grate, L., Haussler, D., and Ares, M., Jr. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. RNA *5*, 221-234.

Staley, J. P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Mol Cell *3*, 55-64.

Tardiff, D. F., Lacadie, S. A., and Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. Mol Cell *24*, 917-929.

Tardiff, D. F., and Rosbash, M. (2006). Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly. RNA *12*, 968-979.

Vilardell, J., and Warner, J. R. (1994). Regulation of splicing at an intermediate step in the formation of the spliceosome. Genes Dev *8*, 211-220.

Vilardell, J., and Warner, J. R. (1997). Ribosomal protein L32 of Saccharomyces cerevisiae influences both the splicing of its own transcript and the processing of rRNA. Mol Cell Biol *17*, 1959-1965.

Vilardell, J., Yu, S. J., and Warner, J. R. (2000). Multiple functions of an evolutionarily conserved RNA binding domain. Mol Cell *5*, 761-766.

Villa, T., Pleiss, J. A., and Guthrie, C. (2002). Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? Cell *109*, 149-152.

Wahl, M. C., Will, C. L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Will, C. L. L., R. . (2006). Spliceosome structure and function. In The RNA World, Third Edition, R. F. Gesteland, Cech, T. R. &, and J. F. Atkins, eds. (Cold Spring Harbor, NY., Cold Spring Harbor Laboratory Press), pp. 369-400.

Wong, C. M., Qiu, H., Hu, C., Dong, J., and Hinnebusch, A. G. (2007). Yeast cap binding complex (CBC) impedes recruitment of cleavage factor IA to weak termination sites. Mol Cell Biol.

Wu, J., and Manley, J. L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev *3*, 1553-1561.

Zhuang, Y., and Weiner, A. M. (1989). A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev *3*, 1545-1552.

DISCUSSION

Splicing is investigated actively and hundreds of research articles are published in this field. An enormous amount has been learned about its fundamental mechanism, its regulation, its implications in various disease and the inner workings and composition of the spliceosome. However, in spite of over 30 years of efforts by scientists worldwide, many interrogations remain to be addressed.

1. Picking out the 3'ss

One important question that is still unresolved is how exactly the poorly defined 3'ss is selected amidst other sequences that appear not to be different from the one favored by the spliceosome. This is particularly intriguing in light of the fact that splicing is an extremely precise process, and alternative splice isoforms in *S.cerevisiae* are rare (Yassour et al., 2009). This is in stark contrast with the topology of many introns. In the extreme case of the *S.cerevisiae BUD25*, the 3'ss is the 9th HAG situated at 130 nucleotides from the BS. The proposed mechanism of scanning along the intron (Smith et al., 1993; Smith et al., 1989) clearly fails to explain selection of the 3'ss in this case. Additionally, the distance constraint between the BS and 3'ss (Cellini et al., 1986; Luukkonen and Seraphin, 1997) is violated. This case is by no means the only example of failure to explain selection in this organism because there are in total nearly 150 HAGs between the BS and the 3'ss scattered in about 90 introns. So far, no mechanism that could account for the oversight of all these HAGs by the spliceosome has been put forward.

1.a A structure for 3'ss selection in the *VMA10* intron

We devised a screen based on the *CUP1* reporter system (Lesser and Guthrie, 1993) in an effort to find factors that mediate selection of the 3'ss. The intron of *VMA10* was fused to *CUP1*, in frame with the CAG upstream of the annotated 3'ss (see **Fig. 11**). After mutagenesis, cells carrying this construct were grown on high concentrations of copper to select for mutants that should

have favoured usage of the alternative splice site. However, this screen did not yield the anticipated result because all positive clones that were tested carried mutations that introduced new 3'ss close to the BS in the construct. This result suggests that there is not a single factor responsible for choosing the 3'ss and that the selection may be intrinsic to the intronic sequence. In addition, the fact that all new 3'ss appeared close to the BS either indicates a clear preference for proximity to this splicing signal or the existence of a region upstream of the annotated splice site where an HAG cannot be selected. Independently of this screen, one final curious observation was made when the AAG in the same intron was mutated to AAU. We noted a clear decrease of splicing to the annotated splice site UAG but no increase in splicing to CAG (**Sup. Fig. 1** Meyer et al.). All these observations led us to consider a mechanism of 3'ss selection operated by a secondary structure that sequesters alternative HAGs. In this scenario, the requirement for a factor that favours one splice site over another is bypassed and the selection directly depends on the intronic sequence itself. Additionally, upon formation, this structure reduces the distance between the BS and the 3'ss which explains how this intron can be spliced in spite of the large distance that separates these two cues. It also offers a clarification as to why new HAGs are introduced only near the BS. Finally, it predicts that the mutation of AAG to AAU could destabilize the secondary structure which would as a consequence reduce splicing efficiency. Upon *in silico* prediction and experimental validation, we concluded that a 63 nucleotide secondary structure is indeed formed in the *VMA10* intron (**Fig. 3** Meyer et al.**)**. It has the dual function of sequestering the alternative HAGs and shortening the BS to 3'ss distance.

1.b Positioning of the secondary structure

The secondary structure of the *VMA10* intron is not allowed to be formed at less than 9 nucleotides from the BS (**Fig. 5** Meyer et al.), but only the bottom part of the structure appears to be prevented when placed close to the BS. The top of the structure is thus still likely to be formed, but the mechanism

through which this is attained is as of yet unknown. The BS region is an area crowded by many factors. After initial recognition by BBP and Mud2, it is occupied by U2 snRNP and is later also contacted by the U5 snRNP protein Prp8 (Grainger and Beggs, 2005). The finding that a secondary structure is not permitted to abide close to the BS is therefore orderly. Indeed, it is probable that such a structure, combined with the spliceosomal factors, would surpass the spatial capacity in the vicinity of the BS. It will be interesting to investigate how this structure is prevented, and two possibilities spring to mind. On the one hand, there could be a hindrance already at the onset of spliceosome assembly. In this case, the structure would never be formed close to the BS. On the other hand, after formation at the beginning, the structure would be unwound later, when shortage of space hinders more spliceosomal components to join and the area has to be disencumbered (Fabrizio et al., 2009). Possibly, the easiest way to distinguish between these possibilities is by finding the factor(s) responsible for the prevention of the structure. In an effort to do this, we hypothesized that Mud2 may be implicated, by imposing a steric hindrance downstream of the BS and thus preventing secondary structure formation. However, Mud2 appears not to be involved, since the bottom of the structure of the *VMA10* intron placed close to the BS is still prevented from forming in a *mud2Δ* strain (data not shown). One possibility to elucidate the question would be to perform a genetic screen based on the *VMA10-CUP1* system. The *CUP1* ORF should be placed in frame with the annotated 3'ss in an intron where the distance between the BS and the beginning of the structure is diminished to 5 nt as in *VMA10* **9**. In this setup, the structure is not formed at the bottom and the CAG is the 3'ss. The rationale for this screen is that *cupΔ* cells bearing this construct should grow only at low concentrations of copper, unless a mutation that allows formation of the secondary structure is introduced. This would in turn allow for at least partial recovery of a wt splicing pattern and thus, increased resistance to copper. Mutations that allow for structure formation in the vicinity of the BS

could be selected for on a high copper concentration. This mutation could then be identified by complementation and characterized.

1.c The maximum BS to 3'ss distance for splicing

The last point of *VMA10* intron splicing that needed clarification was the fact that even though the secondary structure was disrupted, the now freed HAGs were not efficiently used as splice sites (**Fig. 3c** Meyer et al., lane 2). This lack of splicing is not due to a blockage of the process as a whole, but to a specific impairment of the second step. This was evidenced by the accumulation of the first step lariat in a *dbr1*Δ strain (data not shown) incapable of degrading this intermediate. The answer as to why the second step is impaired came from the comparison of the different BS to HAG distances. Although the annotated splice site is located at 105 nucleotides from the BS, when the 63 nucleotide of the secondary structure are taken into account, the effective distance decreases to 42 nucleotides. In the case of an open structure, this same distance is 57 nucleotides for AAG and 83 for CAG. When nucleotides were gradually removed between the BS and the open structure, it became evident that the distance was the key factor to explain the observed phenotype. Indeed, splicing to AAG became increasingly efficient as the distance between this splice site and the BS was brought from 57 nucleotides (*VMA10* **2**), to 51 (*VMA10* **5**) and then 45 (*VMA10* **6**) from the BS (**Fig. 3e** Meyer et al.). Splicing efficiency to the AAG that is placed at 45 nucleotides from the BS is comparable to splicing to UAG in the wt construct (**Fig. 3e** Meyer et al., compare lanes 1 and 4). This indicates that there is probably a limit to the reach of the spliceosome that is set at about 45 nucleotides. A splice site located further away eludes the spliceosome and therefore leads to inefficient splicing, or no splicing at all. This also explains why splicing to CAG did not improve because the shortest distance from the BS attained by this potential 3'ss is of 71 nucleotides in this experiment (*VMA10* **6**).

The question as to how this limit is defined is a fascinating one, and the answers can only be speculative so far. The idea that a distance of 45 nucleotides is limiting because the spliceosome cannot physically reach further is interesting, but implies that the RNA strand between the BS and 3'ss is rigid and cannot be bent to accommodate binding to the spliceosome. Unless proteins binding to this stretch of RNA can explain such a property it is hard to imagine how this result could come about given the flexibility of RNA. Another possibility is that of a kinetic limitation to the realm of the spliceosome. In this scenario, the time imparted to searching for a 3'ss would expire when a protein hydrolyses a ATP to prompts the next conformation. In normal conditions, the spliceosome does not have time to reach beyond about 45 nucleotides, but this could be extended if ATP hydrolysis were to be delayed. This idea could be tested rather easily since the number of proteins that hydrolyse ATP is limited and many mutants for Prp5, Prp16 and Prp22 exist (Burgess and Guthrie, 1993; Schneider et al., 2002; Xu and Query, 2007). It could therefore be tested if mutants in the ATPase domain of one of these proteins enhances splicing to the AAG that is at 57 nucleotides from the BS in the open stem construct (*VMA10* **2**). If this is indeed the case, it could mean that when ATP hydrolysis is impaired, the spliceosome has time to reach further downstream of the BS in search of a suitable HAG for splicing.

1.d 3'ss selection in *S.cerevisiae*

The secondary structure encountered in the *VMA10* intron sequesters 2 alternative HAGs which renders them oblivious to the spliceosome. We thus defined a mechanism for selection of the 3'ss. A number of further HAGs are encountered in additional introns, and we sought to determine if the 3'ss is selected in a similar way in these other cases. The BS to 3'ss regions of all other *S.cerevisiae* introns were inspected for possible secondary structures. The parameters for this bioinformatics search were set to be in accordance with what has been learned from the case of *VMA10*. Consequently, the 8 nucleotides after the BS were forced to remain unpaired (see previous

section). This approach revealed the possibility that the mechanism of 3'ss selection may be similar to the one of *VMA10* in other introns because many additional structures could be predicted (see http://regulatorygenomics.upf.edu/Yeast_Introns/index.html). The secondary structure foreseen for the intron of *RPS23B* was also validated experimentally (**Fig. 2** Meyer et al.), thus substantially increasing the confidence in the other predictions. Very interestingly, when the predicted secondary structures are taken into account, the maximum distance between BS and 3'ss is of 45 nucleotides (**Fig. 1** Meyer et al.). This finding coincides astonishingly well with the result observed for the *VMA10* intron (**Fig. 3e** Meyer et al.). It shows that the maximum of 45 nucleotide between BS and 3'ss is not restricted to this intron sequence and reinforces the idea of a limit to the reach of the spliceosome.

The various structures sequester over two thirds of alternative HAGs. The remaining 42 that are not predicted to be included in a structure are disregarded by the spliceosome because they lie too close to the BS (9 nucleotides or less, **Fig. 4** Meyer et al., lanes 1 and 4) or are weak splice sites (AAG rather than UAG or CAG, **Fig. 4** Meyer et al., lanes 1 to 3, and data not shown). The composite set of rules that govern 3'ss selection in *S.cerevisiae* can therefore be summarised as follows. The strongest HAG present outside of a secondary structure, at an actual or effective distance of 10 to 45 nucleotides downstream of the BS will be selected as the 3'ss (**Fig. 16**). It is of interest to note that when two suitable 3'ss are present in this window, they are used equally by the spliceosome, provided that they have the same strength (**Fig. 4** Meyer et al., lane 2 and data not shown). In addition, a secondary structure does not impose any penalty on the usage of a HAG situated downstream of it (**Fig. 4** Meyer et al., compare lanes 2 and 5).

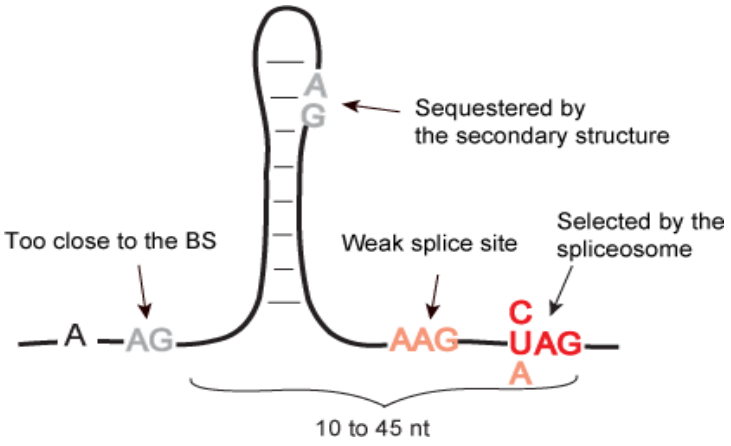


Figure 16. Illustrations of the rules that govern 3'ss selection in *S.cerevisiae.*

It is possible that the mechanisms of 3'ss selection described here are also applicable to higher eukaryotes. However, in those organisms, the BS is much more degenerate, and has not been mapped precisely in most cases. It makes a study similar to the one performed here much more arduous. Additionally, alternative splicing makes the task even more difficult, because unlike in *S.cerevisiae*, different 3'ss can be selected depending on various additional factors.

1.e Exceptions to the rule

With the set of rules enunciated above, the splicing of a majority of introns can be explained. In fact, the only exception is the intron of *REC102* that has 3 alternative HAGs between the BS and the 3'ss. The two first HAGs are masked by a secondary structure and therefore do not challenge the rules described in this work. However, both the annotated and the alternative that is not contained in a structure are AAGs, so splice site strength does not favour either of them. Additionally, they are set only 5 nucleotides apart which implies that their sequence context is very similar, and the presence of a poly-U tract, for example, that could strengthen the annotated 3'ss is therefore excluded. To make sure that this inconsistency was not due to a misannotation

or usage of both splice sites, we tested the splicing patter of this intron *in vivo*, by RT-PCR (data not shown). This experiment was performed on sporulating cells, because *REC102* is only expressed in this condition. In addition, the strain used was NMD defective due to the fact that splicing to the alternative AAG would lead to a transcript that contains an early stop codon. However, even in these conditions, we were able to identify only transcripts resulting from splicing to the annotated 3'ss. The mechanism whereby the 3'ss is selected in this case can thus not be implemented from our work. The answer may come from work in human cells were it was shown that in the case of tandem splice sites like here, the second splice site is usually favoured (Chua and Reed, 2001).

The only exception to the rules of 3'ss selection established here was presented in the previous paragraph; however, there is room for more caution. Indeed, some of the structures that we predicted are very weak. In most cases this is not relevant, because the weak structures do not cover any alternative HAG or are not required to bring the 3'ss within reach of the spliceosome. However, there is one notable exception. The 3'ss of *VPS75* (CAG) is preceded by 3 HAGs (UAG, AAG, UAG). All three are predicted to be included in a secondary structure, but this structure is weak. Upon testing the splicing pattern *in vivo* by RT-PCR on NMD defective strains (data not shown), we found that over half of the isoforms for this transcript resulted from splicing to the two UAGs (these isoforms were also found in Yassour et al., 2009). This indicates that the secondary structure is not formed. It would therefore be judicious to set a minimum threshold below which secondary structures have a low probability of being formed. However, it is difficult to decide where to set the limit, and it may depend on the sequence of the intron or on cellular conditions and external factors that facilitate or hinder secondary structures.

The last HAGs that have to be considered are those situated after the 3'ss. Although in most cases they are situated beyond 45 nucleotides after the BS, some are not. The most likely explanation as to why they are not selected is that additional secondary structures form to occlude these HAGs (**Sup. Fig. 3** Meyer et al.), especially since the GC content in exons is usually higher than in introns. This probably makes exons more prone to adopt secondary structures because GC base-pairing is stronger than AU or GU pairing. However, this hypothesis is hard to test. The difficulty resides in the troublesome task of predicting structures that could have this function. Indeed, for structures between the BS and the 3'ss, the nucleotides that can adopt a secondary structure are those situated between these two splicing signals. Additionally, the nucleotides of the BS and 3'ss have to be single stranded because they are used by the spliceosome. These restrictions establish a defined window within which a structure can be predicted. The case is drastically different in the case of HAGs downstream of the 3'ss. Indeed, it is difficult to decide which nucleotides should be taken into account to predict a secondary structure. An additional obstacle is that the transcription rate may have to be taken into account since the splicing process is at least partially co-transcriptional (Tardiff et al., 2006). The 3'ss might therefore already have been selected by the spliceosome before an alternative HAG placed downstream is transcribed.

1.f Incompatibility of scanning and 3'ss selection

The combination of the observation that in many introns, the 3'ss is not the first HAG after the BS, and the work presented here, disprove the mechanism of scanning along the intron (Smith et al., 1993; Smith et al., 1989), at least in the form it was described. It could be argued that a few amendments to the model would make it more suitable, by limiting the scope of the scanning to 45 nucleotides for example. But even so, several observations do not support the scanning mechanism. To start with, it is difficult to explain why an AAG would be weaker than a CAG or a UAG (**Fig. 4** Meyer et al., lanes 1 and 3

and data not shown). Indeed, in case of 3'ss selection by scanning, the first splice site would be chosen because, by definition, the scanning would stop at this splicing cue. It is hard to conceive how the spliceosome would then become aware of a stronger HAG located as much as 12 nucleotides downstream (**Fig. 4** Meyer et al.). A leaky scanning mechanism cannot be invoked either, because when an AAG is placed in front of a UAG, it is only scarcely used (**Fig. 4** Meyer et al., lane 1). This would mean that the scanning is very leaky on the AAG which could therefore not be used as an efficient splice site. Evidently this is not the case, because AAG is the natural 3'ss of at least 8 *S.cerevisiae* introns. It could be that the spliceosome goes back on itself in case it does not find another suitable HAG, but this seems rather convoluted. A similar claim can be made for the selection of two equally strong HAGs that are used to similar extents when present in the window of selection (**Fig. 4** Meyer et al., lane 2). If scanning were to be leaky, the first splice site would be used more than the second, because the percentage of leakiness would likely be the same for both. It is therefore much more likely that the spliceosome considers all HAGs in a window of 10 to 45 nucleotides downstream of the intron and uses the strongest one(s) it encounters. How this is achieved remains to be worked out.

1.g Biological relevance of a secondary structure

It can seem surprising that a deletion of nearly 40% of an intron (63 nucleotides out of 162) has no effect on splicing efficiency, or on biological fitness. This however is the case for the intron of *VMA10* (**Fig. 3c** Meyer et al., lane 4 and **Fig. 6** Meyer et al.), at least in the conditions that we tested. This result indicates that when the 3'ss is situated at an appropriate distance from the BS and not preceded by HAGs, the secondary structure is dispensable. It also suggests that transcription of an extra 63 nucleotides for every transcript is an acceptable biological cost for an efficient mechanism of 3'ss selection. Otherwise, selective pressure would have eliminated such structures and intervening HAGs. To put this biological cost into perspective,

introns in higher eukaryotes can span several hundreds of kilo bases. They are for the better part degraded after the splicing reaction is complete. Compared to this, the number of 63 extra nucleotides is minute.

In the case and conditions that we have tested, the secondary structure appears not to have a regulatory function. However, it cannot be ruled out that in other conditions, or in other instances, the cell takes advantage of the secondary structure to regulate production of a protein. Indeed, a system in which the excess of a protein triggers opening of the structure between the BS and 3'ss, and thus inefficient splicing of its own transcript, can be imagined.

1.h Secondary structures in splicing

A similar structure to the ones we found in this study had been encountered in the actin gene of another yeast, *Klyveromyces lactic* (Deshler and Rossi, 1991), but also in the *HGH* gene in human (Estes et al., 1992). This shows that our findings may be of general relevance. In addition to those examples, several instances of RNA folding with a direct role in aiding, inhibiting or modulating splicing have also been described. For instance, exon 10 of the gene *Tau* has been associated with frontotemporal dementia and parkinsonism, and is contained in a stem-loop in mouse and human. The destabilization of this structure leads to an increase in exon 10 inclusion (Grover et al., 1999; Jiang et al., 2000; Varani et al., 1999). Another example is the fruit fly pre-mRNA *Dscam* that has several possible exons 4 and 6, but only one of each is included in the mRNA, possibly through the competition between the formation of different secondary structures (Graveley, 2005; Kreahling and Graveley, 2005). A further instance of a secondary structure has been encountered in the duck hepatitis B virus, where an RNA fold encompasses both splice sites which leads to intron retention (Loeb et al., 2002). Finally, it has been shown in *S.cerevisiae* that a secondary structures in *RPS17B* and *YRA1* between the 5'ss and the BS makes splicing more efficient by effectively shortening the intron (Charpentier and Rosbash, 1996; Libri et

al., 1995; Preker and Guthrie, 2006). In summary, it is not the first time that RNA folding has been implicated in splicing. However, contrarily to the examples listed above, we show that instead of being an oddity encountered in disparate transcripts, secondary structures are widespread and govern the splicing outcome of a wide array of introns. Only in the short segment between the BS and 3'ss in *S.cerevisiae*, one third of introns are likely to harbor a secondary structure. But yeast introns are generally short in comparison to their metazoan counterparts. The latter are therefore much more prone to forming secondary structures, which may be involved for example in alternative splicing regulation. It will therefore become increasingly important to view transcripts not as linear entities, but to consider the secondary structures that they form in order to understand their splicing patterns. As our understanding of splicing advances, it may even become evident that not only secondary, but also tertiary structures play an important role in splicing. Advances in our knowledge of the folding capacities of RNA, and in computer modeling of secondary and tertiary structures will be a great asset to further understand the mechanisms of splicing.

2. Cbp80 and recruitment of the U2 snRNP

Selection of the 3'ss has been described in the previous section, but other aspects of 3' end identification remain to be explained. One of them is how L30 prevents the recognition of the BS by U2 snRNP. A genetic interaction between the regulation by L30 of its own transcript and the large subunit of Cbp80 has been found (**Fig. 2F** Bragulat et al.). This protein has been implicated in a variety of processes during the nuclear peregrinations of the transcripts. It helps the recruitment of U1 snRNP and its remodeling after recognition of the 5'ss (Colot et al., 1996; Fortes et al., 1999b; Lewis et al., 1996; Pomeranz Krummel et al., 2009). Later, it also facilitates the recruitment of the tri-snRNP U4/U6·U5 by promoting displacement of U1 snRNP (Gornemann et al., 2005; O'Mullane and Eperon, 1998), and is likely

to also mediate the connection between splicing and export, at least in some transcripts (Dong et al., 2007; Preker and Guthrie, 2006). Yet, we argue that this protein has at least another function that had not been described before. Based on previous results from our laboratory, inhibition of *RPL30* splicing by L30 is achieved through interference of a critical remodeling step of the forming spliceosome (Macias et al., 2008). This interference is compromised when the L30-*RPL30* system is destabilized (5A C9U), but can be partially recovered by the deletion of Cbp80 (**Fig. 3** Bragulat et al. D lower panel, lanes 1,2,5 and 6). Because Cbp80 does not have a function in destabilizing L30 on the *RPL30* transcript, this finding indicates an antagonistic role of L30 and Cbp80. Consequently, since L30 inhibits CC2 remodeling in conditions of regulation, Cbp80 must promote it.

The nature of the remodeling promoted by Cbp80 and hindered by L30 is not yet clear. It is possible that the remodeling step targeted by these two proteins is the same as the one promoted by Prp5, to secure docking of U2 snRNP onto the BS (Xu and Query, 2007). Although this hypothesis is somewhat undermined by the finding that mutations in Prp5 that affect U2 snRNP recruitment have no effect on splicing regulation by L30 (Sara Macías, unpublished results).It is therefore possible that there is another remodeling step that has not been described so far. Another unknown is the mechanism by which hindrance and promotion are achieved. It has to be worked out, for example, which element(s) L30 and Cbp80 target to achieve regulation, and if they are the same for both. It may be achieved by looking for elements of the CC2 or the pre-spliceosome that can interact with either of the two proteins. Additionally, several mutants that show decreased inhibition by L30, as well as further SLR strains, have been generated. Tracking down the mutations that procure these phenotypes may be of great help to further progress in the elucidation of splicing regulation by L30, and in finding out the nature of the remodeling promoter by Cbp80.

CONCLUSIONS

1. A secondary structure forms in the BS to 3'ss region of many introns.
2. Alternative 3'ss can be occluded from the spliceosome by secondary structures.
3. A secondary structure cannot be formed at a distance of less than 9 nucleotides from the BS
4. CAG and UAG are 3'ss of similar strengths, whereas AAG is weaker.
5. The largest distance between the BS and the 3'ss that allows for efficient splicing is of about 45 nucleotides.
6. The maximum distance, effective or natural, that separates the BS from the 3'ss in *S.cerevisiae* is of 45 nucleotides.
7. Even a strong AG situated at 9 nt or less from the BS cannot compete with a 3'ss that is situated downstream of it.
8. An HAG has to be situated in a window of 10 to 45 nucleotides after the BS, in order to be used efficiently.
9. Selection of the 3'ss is now explainable for all *S.cerevisiae* introns except one.
10. At least in the case of *VMA10*, the secondary structure can be removed without affecting splicing efficiency or biological fitness.
11. RNA secondary structures are widespread and have to be taken into account in order for splicing patterns to be understood.
12. Cbp80 favors a rearrangement of the nascent spliceosome that facilitates U2 snRNP recruitment and drives splicing forward.
13. The sequence of the *RPL30* 5'ss is important for the role of Cbp80 in spliceosomal rearrangement.

REFERENCES

Abovich, N., Liao, X.C., and Rosbash, M. (1994). The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. Genes Dev *8*, 843-854.

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell *89*, 403-412.

Ansari, A., and Schwer, B. (1995). SLU7 and a novel activity, SSF1, act during the PRP16-dependent step of yeast pre-mRNA splicing. Embo J *14*, 4001-4009.

Arenas, J.E., and Abelson, J.N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. Proc Natl Acad Sci U S A *94*, 11798-11802.

Arning, S., Gruter, P., Bilbe, G., and Kramer, A. (1996). Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. Rna *2*, 794-810.

Bartels, C., Klatt, C., Luhrmann, R., and Fabrizio, P. (2002). The ribosomal translocase homologue Snu114p is involved in unwinding U4/U6 RNA during activation of the spliceosome. EMBO Rep *3*, 875-880.

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. Embo J *26*, 1737-1748.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell *89*, 781-787.

Berglund, J.A., Fleming, M.L., and Rosbash, M. (1998). The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. Rna *4*, 998-1006.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Luhrmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. Nature *452*, 846-850.

Boon, K.L., Auchynnikava, T., Edwalds-Gilbert, G., Barrass, J.D., Droop, A.P., Dez, C., and Beggs, J.D. (2006). Yeast ntr1/spp382 mediates prp43 function in postspliceosomes. Mol Cell Biol *26*, 6016-6023.

Brenner, T.J., and Guthrie, C. (2005). Genetic analysis reveals a role for the C terminus of the Saccharomyces cerevisiae GTPase Snu114 during spliceosome activation. Genetics *170*, 1063-1080.

Burgess, S.M., and Guthrie, C. (1993). A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. Cell *73*, 1377-1391.

Caspary, F., and Seraphin, B. (1998). The yeast U2A'/U2B complex is required for pre-spliceosome formation. Embo J *17*, 6348-6358.

Cellini, A., Felder, E., and Rossi, J.J. (1986). Yeast pre-messenger RNA splicing efficiency depends on critical spacing requirements between the branch point and 3' splice site. Embo J *5*, 1023-1030.

Chao, J.A., and Williamson, J.R. (2004). Joint X-ray and NMR refinement of the yeast L30e-mRNA complex. Structure *12*, 1165-1176.

Charpentier, B., and Rosbash, M. (1996). Intramolecular structure in yeast introns aids the early steps of in vitro spliceosome assembly. Rna *2*, 509-522.

Chen, J.Y., Stands, L., Staley, J.P., Jackups, R.R., Jr., Latus, L.J., and Chang, T.H. (2001). Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Mol Cell *7*, 227-232.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol *10*, 741-754.

Chua, K., and Reed, R. (2001). An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. Mol Cell Biol *21*, 1509-1514.

Collins, C.A., and Guthrie, C. (2000). The question remains: is the spliceosome a ribozyme? Nat Struct Biol *7*, 850-854.

Colot, H.V., Stutz, F., and Rosbash, M. (1996). The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. Genes Dev *10*, 1699-1708.

Company, M., Arenas, J., and Abelson, J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. Nature *349*, 487-493.

Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. Cell *136*, 777-793.

Crotti, L.B., and Horowitz, D.S. (2009). Exon sequences at the splice junctions affect splicing fidelity and alternative splicing. Proc Natl Acad Sci U S A *106*, 18954-18959.

Dabeva, M.D., and Warner, J.R. (1993). Ribosomal protein L32 of Saccharomyces cerevisiae regulates both splicing and translation of its own transcript. J Biol Chem *268*, 19669-19674.

Datta, B., and Weiner, A.M. (1991). Genetic evidence for base pairing between U2 and U6 snRNA in mammalian mRNA splicing. Nature *352*, 821-824.

Davis, C.A., Grate, L., Spingola, M., and Ares, M., Jr. (2000). Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. Nucleic Acids Res *28*, 1700-1706.

de la Cruz, J., Kressler, D., and Linder, P. (1999). Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. Trends Biochem Sci *24*, 192-198.

Deshler, J.O., and Rossi, J.J. (1991). Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron. Genes Dev *5*, 1252-1263.

Dong, S., Li, C., Zenklusen, D., Singer, R.H., Jacobson, A., and He, F. (2007). YRA1 autoregulation requires nuclear export and cytoplasmic Edc3p-mediated degradation of its pre-mRNA. Mol Cell *25*, 559-573.

Eng, F.J., and Warner, J.R. (1991). Structural basis for the regulation of splicing of a yeast messenger RNA. Cell *65*, 797-804.

Estes, P.A., Cooke, N.E., and Liebhaber, S.A. (1992). A native RNA secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms. J Biol Chem *267*, 14902-14908.

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell *36*, 593-608.

Fedorova, O., and Zingler, N. (2007). Group II introns: structure, folding and splicing mechanism. Biol Chem *388*, 665-678.

Fortes, P., Bilbao-Cortes, D., Fornerod, M., Rigaut, G., Raymond, W., Seraphin, B., and Mattaj, I.W. (1999a). Luc7p, a novel yeast U1 snRNP protein with a role in 5' splice site recognition. Genes Dev *13*, 2425-2438.

Fortes, P., Kufel, J., Fornerod, M., Polycarpou-Schwarz, M., Lafontaine, D., Tollervey, D., and Mattaj, I.W. (1999b). Genetic and physical interactions involving the yeast nuclear cap-binding complex. Mol Cell Biol *19*, 6543-6553.

Frank, D., and Guthrie, C. (1992). An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. Genes Dev *6*, 2112-2124.

Frank, D., Patterson, B., and Guthrie, C. (1992). Synthetic lethal mutations suggest interactions between U5 small nuclear RNA and four proteins required for the second step of splicing. Mol Cell Biol *12*, 5197-5205.

Goguel, V., and Rosbash, M. (1993). Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. Cell *72*, 893-901.

Gornemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol Cell *19*, 53-63.

Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. Rna *11*, 533-557.

Graveley, B.R. (2005). Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell *123*, 65-73.

Griffiths, A., Miller, J., Suyuki, D., Lewontin, R., and Gelbart, W. (1999). Introduction to Genetic Analysis, 7th ed. (New York, NY: W.H Freeman & Co.).

Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K., and Hutton, M. (1999). 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. J Biol Chem *274*, 15134-15143.

Gu, M., and Lima, C.D. (2005). Processing the message: structural insights into capping and decapping mRNA. Curr Opin Struct Biol *15*, 99-106.

Hagiwara, M., and Nojima, T. (2007). Cross-talks between transcription and post-transcriptional events within a 'mRNA factory'. J Biochem *142*, 11-15.

Halic, M., Becker, T., Frank, J., Spahn, C.M., and Beckmann, R. (2005). Localization and dynamic behavior of ribosomal protein L30e. Nat Struct Mol Biol *12*, 467-468.

Hausner, T.P., Giglio, L.M., and Weiner, A.M. (1990). Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. Genes Dev *4*, 2146-2156.

Hilliker, A.K., Mefford, M.A., and Staley, J.P. (2007). U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. Genes Dev *21*, 821-834.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717-728.

Jiang, Z., Cote, J., Kwon, J.M., Goate, A.M., and Wu, J.Y. (2000). Aberrant splicing of tau pre-mRNA caused by intronic mutations associated with the inherited dementia frontotemporal dementia with parkinsonism linked to chromosome 17. Mol Cell Biol *20*, 4036-4048.

Jones, M.H., Frank, D.N., and Guthrie, C. (1995). Characterization and functional ordering of Slu7p and Prp17p during the second step of pre-mRNA splicing in yeast. Proc Natl Acad Sci U S A *92*, 9687-9691.

Juneau, K., Nislow, C., and Davis, R.W. (2009). Alternative splicing of PTC7 in Saccharomyces cerevisiae determines protein localization. Genetics *183*, 185-194.

Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol Cell *12*, 5-14.

Kao, H.Y., and Siliciano, P.G. (1996). Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. Mol Cell Biol *16*, 960-967.

Kim, S.H., and Lin, R.J. (1996). Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. Mol Cell Biol *16*, 6810-6819.

Kistler, A.L., and Guthrie, C. (2001). Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. Genes Dev *15*, 42-49.

Konarska, M.M., Grabowski, P.J., Padgett, R.A., and Sharp, P.A. (1985). Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature *313*, 552-557.

Konarska, M.M., and Sharp, P.A. (1987). Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. Cell *49*, 763-774.

Konarska, M.M., Vilardell, J., and Query, C.C. (2006). Repositioning of the reaction intermediate within the catalytic center of the spliceosome. Mol Cell *21*, 543-553.

Kramer, A. (1992). Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a presplicing complex. Mol Cell Biol *12*, 4545-4552.

Kreahling, J.M., and Graveley, B.R. (2005). The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the Drosophila Dscam pre-mRNA. Mol Cell Biol *25*, 10251-10260.

Kress, T.L., Krogan, N.J., and Guthrie, C. (2008). A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. Mol Cell *32*, 727-734.

Lacadie, S.A., and Rosbash, M. (2005). Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. Mol Cell *19*, 65-75.

Lamond, A.I., Konarska, M.M., Grabowski, P.J., and Sharp, P.A. (1988). Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. Proc Natl Acad Sci U S A *85*, 411-415.

Lareau, L.F., Brooks, A.N., Soergel, D.A., Meng, Q., and Brenner, S.E. (2007). The coupling of alternative splicing and nonsense-mediated mRNA decay. Adv Exp Med Biol *623*, 190-211.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. (1980). Are snRNPs involved in splicing? Nature *283*, 220-224.

Lesser, C.F., and Guthrie, C. (1993). Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. Genetics *133*, 851-863.

Lewis, J.D., Gorlich, D., and Mattaj, I.W. (1996). A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing. Nucleic Acids Res *24*, 3332-3336.

Li, B., Vilardell, J., and Warner, J.R. (1996). An RNA structure involved in feedback regulation of splicing and of translation is critical for biological fitness. Proc Natl Acad Sci U S A *93*, 1596-1600.

Libri, D., Stutz, F., McCarthy, T., and Rosbash, M. (1995). RNA structural patterns and splicing: molecular basis for an RNA-based enhancer. Rna *1*, 425-436.

Liu, H.X., Goodall, G.J., Kole, R., and Filipowicz, W. (1995). Effects of secondary structure on pre-mRNA splicing: hairpins sequestering the 5' but not the 3' splice site inhibit intron processing in Nicotiana plumbaginifolia. Embo J *14*, 377-388.

Liu, L., Query, C.C., and Konarska, M.M. (2007). Opposing classes of prp8 alleles modulate the transition between the catalytic steps of pre-mRNA splicing. Nat Struct Mol Biol *14*, 519-526.

Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Kramer, A., and Sattler, M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. Science *294*, 1098-1102.

Loeb, D.D., Mack, A.A., and Tian, R. (2002). A secondary structure that contains the 5' and 3' splice sites suppresses splicing of duck hepatitis B virus pregenomic RNA. J Virol *76*, 10195-10202.

Luukkonen, B.G., and Seraphin, B. (1997). The role of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in Saccharomyces cerevisiae. Embo J *16*, 779-792.

Macias, S., Bragulat, M., Tardiff, D.F., and Vilardell, J. (2008). L30 binds the nascent RPL30 transcript to repress U2 snRNP recruitment. Mol Cell *30*, 732-742.

Madhani, H.D., and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell *71*, 803-817.

Mayas, R.M., Maita, H., and Staley, J.P. (2006). Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat Struct Mol Biol *13*, 482-490.

Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. Nature *402*, 838-841.

Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature *365*, 364-368.

Newman, A.J., and Norman, C. (1992). U5 snRNA interacts with exon sequences at 5' and 3' splice sites. Cell *68*, 743-754.

Nilsen, T.W. (1998). RNA-RNA interactions in nuclear pre-mRNA splicing. RNA Structure and Function (Cold Spring Harbor Laboratory Press).

Nilsen, T.W. (2003). The spliceosome: the most complex macromolecular machine in the cell? Bioessays *25*, 1147-1149.

O'Day, C.L., Dalbadie-McFarland, G., and Abelson, J. (1996). The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. J Biol Chem *271*, 33261-33267.

O'Mullane, L., and Eperon, I.C. (1998). The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice sites. Mol Cell Biol *18*, 7510-7520.

Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. Cell *108*, 439-451.

Parker, R., Siliciano, P.G., and Guthrie, C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. Cell *49*, 229-239.

Patterson, B., and Guthrie, C. (1991). A U-rich tract enhances usage of an alternative 3' splice site in yeast. Cell *64*, 181-187.

Peled-Zehavi, H., Berglund, J.A., Rosbash, M., and Frankel, A.D. (2001). Recognition of RNA branch point sequences by the KH domain of splicing factor 1 (mammalian branch point binding protein) in a splicing factor complex. Mol Cell Biol *21*, 5232-5241.

Pena, V., Rozov, A., Fabrizio, P., Luhrmann, R., and Wahl, M.C. (2008). Structure and function of an RNase H domain at the heart of the spliceosome. Embo J *27*, 2929-2940.

Perriman, R., and Ares, M., Jr. (2000). ATP can be dispensable for prespliceosome formation in yeast. Genes Dev *14*, 97-107.

Perriman, R., Barta, I., Voeltz, G.K., Abelson, J., and Ares, M., Jr. (2003). ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA. Proc Natl Acad Sci U S A *100*, 13857-13862.

Perriman, R.J., and Ares, M., Jr. (2007). Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. Genes Dev *21*, 811-820.

Pikielny, C.W., Rymond, B.C., and Rosbash, M. (1986). Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes. Nature *324*, 341-345.

Pleiss, J.A., Whitworth, G.B., Bergkessel, M., and Guthrie, C. (2007). Rapid, transcript-specific changes in splicing in response to environmental stress. Mol Cell *27*, 928-937.

Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature *458*, 475-480.

Preker, P.J., and Guthrie, C. (2006). Autoregulation of the mRNA export factor Yra1p requires inefficient splicing of its pre-mRNA. Rna *12*, 994-1006.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501-512.

Query, C.C., and Konarska, M.M. (2006). Splicing fidelity revisited. Nat Struct Mol Biol *13*, 472-474.

Query, C.C., Moore, M.J., and Sharp, P.A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev *8*, 587-597.

Raghunathan, P.L., and Guthrie, C. (1998). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr Biol *8*, 847-855.

Rain, J.C., Rafi, Z., Rhani, Z., Legrain, P., and Kramer, A. (1998). Conservation of functional domains involved in RNA binding and

protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. Rna *4*, 551-565.

Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res *12*, 1231-1245.

Ritchie, D.B., Schellenberg, M.J., Gesner, E.M., Raithatha, S.A., Stuart, D.T., and Macmillan, A.M. (2008). Structural elucidation of a PRP8 core domain from the heart of the spliceosome. Nat Struct Mol Biol *15*, 1199-1205.

Roy, J., Kim, K., Maddock, J.R., Anthony, J.G., and Woolford, J.L., Jr. (1995). The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. Rna *1*, 375-390.

Ruby, S.W., and Abelson, J. (1988). An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. Science *242*, 1028-1035.

Ruskin, B., Zamore, P.D., and Green, M.R. (1988). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell *52*, 207-219.

Rymond, B.C., and Rosbash, M. (1985). Cleavage of 5' splice site and lariat formation are independent of 3' splice site in yeast mRNA splicing. Nature *317*, 735-737.

Schneider, S., Hotz, H.R., and Schwer, B. (2002). Characterization of dominant-negative mutants of the DEAH-box splicing factors Prp22 and Prp16. J Biol Chem *277*, 15452-15458.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nat Struct Mol Biol *16*, 990-995.

Schwer, B., and Gross, C.H. (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. Embo J *17*, 2086-2094.

Schwer, B., and Guthrie, C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. Nature *349*, 494-499.

Schwer, B., and Guthrie, C. (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. Embo J *11*, 5033-5039.

Seraphin, B., Kretzner, L., and Rosbash, M. (1988). A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. Embo J *7*, 2533-2538.

Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell *59*, 349-358.

Siliciano, P.G., and Guthrie, C. (1988). 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. Genes Dev *2*, 1258-1267.

Silverman, E., Edwalds-Gilbert, G., and Lin, R.J. (2003). DExD/H-box proteins and their partners: helping RNA helicases unwind. Gene *312*, 1-16.

Small, E.C., Leggett, S.R., Winans, A.A., and Staley, J.P. (2006). The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. Mol Cell *23*, 389-399.

Smith, C.W., Chu, T.T., and Nadal-Ginard, B. (1993). Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. Mol Cell Biol *13*, 4939-4952.

Smith, C.W., Porro, E.B., Patton, J.G., and Nadal-Ginard, B. (1989). Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. Nature *342*, 243-247.

Smith, D.J., Query, C.C., and Konarska, M.M. (2008). "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. Mol Cell *30*, 657-666.

Sontheimer, E.J., and Steitz, J.A. (1993). The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science *262*, 1989-1996.

Spingola, M., Grate, L., Haussler, D., and Ares, M., Jr. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. Rna *5*, 221-234.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315-326.

Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Mol Cell *3*, 55-64.

Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell *9*, 31-44.

Strauss, E.J., and Guthrie, C. (1991). A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. Genes Dev *5*, 629-641.

Tardiff, D.F., Lacadie, S.A., and Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. Mol Cell *24*, 917-929.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcarcel, J., and Guigo, R. (2009). Nucleosome positioning as a determinant of exon recognition. Nat Struct Mol Biol *16*, 996-1001.

Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. (2008). Crystal structure of a self-spliced group II intron. Science *320*, 77-82.

Tsai, R.T., Fu, R.H., Yeh, F.L., Tseng, C.K., Lin, Y.C., Huang, Y.H., and Cheng, S.C. (2005). Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev *19*, 2991-3003.

Umen, J.G., and Guthrie, C. (1995). Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. Rna *1*, 584-597.

Umen, J.G., and Guthrie, C. (1996). Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3' splice site selection. Genetics *143*, 723-739.

Valadkhan, S. (2007). The spliceosome: a ribozyme at heart? Biol Chem *388*, 693-697.

Valadkhan, S., and Manley, J.L. (2001). Splicing-related catalysis by protein-free snRNAs. Nature *413*, 701-707.

Varani, L., Hasegawa, M., Spillantini, M.G., Smith, M.J., Murrell, J.R., Ghetti, B., Klug, A., Goedert, M., and Varani, G. (1999). Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. Proc Natl Acad Sci U S A *96*, 8229-8234.

Vilardell, J., Chartrand, P., Singer, R.H., and Warner, J.R. (2000). The odyssey of a regulated transcript. Rna *6*, 1773-1780.

Vilardell, J., and Warner, J.R. (1994). Regulation of splicing at an intermediate step in the formation of the spliceosome. Genes Dev *8*, 211-220.

Villa, T., and Guthrie, C. (2005). The Isy1p component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. Genes Dev *19*, 1894-1904.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008a). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476.

Wang, Q., Zhang, L., Lynn, B., and Rymond, B.C. (2008b). A BBP-Mud2p heterodimer mediates branchpoint recognition and influences splicing substrate abundance in budding yeast. Nucleic Acids Res *36*, 2787-2798.

Wassarman, D.A., and Steitz, J.A. (1992). Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. Science *257*, 1918-1925.

White, S.A., Hoeger, M., Schweppe, J.J., Shillingford, A., Shipilov, V., and Zarutskie, J. (2004). Internal loop mutations in the ribosomal protein L30 binding site of the yeast L30 RNA transcript. Rna *10*, 369-377.

Wu, J.A., and Manley, J.L. (1991). Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA. Nature *352*, 818-821.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature *402*, 832-835.

Wyatt, J.R., Sontheimer, E.J., and Steitz, J.A. (1992). Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. Genes Dev *6*, 2542-2553.

Xu, D., Nouraini, S., Field, D., Tang, S.J., and Friesen, J.D. (1996). An RNA-dependent ATPase associated with U2/U6 snRNAs in pre-mRNA splicing. Nature *381*, 709-713.

Xu, Y.Z., and Query, C.C. (2007). Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. Mol Cell *28*, 838-849.

Yang, K., Zhang, L., Xu, T., Heroux, A., and Zhao, R. (2008). Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins. Proc Natl Acad Sci U S A *105*, 13817-13822.

Yassour, M., Kaplan, T., Fraser, H.B., Levin, J.Z., Pfiffner, J., Adiconis, X., Schroth, G., Luo, S., Khrebtukova, I., Gnirke, A., *et al.* (2009). Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. Proc Natl Acad Sci U S A *106*, 3264-3269.

Zamore, P.D., and Green, M.R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc Natl Acad Sci U S A *86*, 9243-9247.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609-614.

Zhang, D., and Rosbash, M. (1999). Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. Genes Dev *13*, 581-592.

Zhong, X.Y., Wang, P., Han, J., Rosenfeld, M.G., and Fu, X.D. (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. Mol Cell *35*, 1-10.

Zhuang, Y., and Weiner, A.M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell *46*, 827-835.