Charles University in Prague
Faculty of Science

Department of Cell Biology

Regulation of pre-mRNA splicing in *S. cerevisiae*: where RNA cooperates with proteins

PhD Thesis

Ondřej Gahura

2011

Supervisor: Doc.RNDr. František Půta, CSc.

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Prague, December 2011 Ondřej Gahura

ACKNOWLEDGMENTS

I would like to thank my supervisor Dr. František Půta and Dr. Petr Folk for the guidance of my work and for the support they have been providing me for years. I would like to acknowledge all lab members, especially Anna Valentová and Vanda Munzarová, who were involved in the project and who substantially contributed mainly to the synthetic lethal screen.

Nejvíc bych chtěl poděkovat svým rodičům a Elišce za lásku a podporu, bez které by tyto řádky a práce nemohly vzniknout.

This work was supported by the Grant Agency of the Charles University grant 398811, the Czech Science Foundation grant 204/02/1512, and the research projects MSM0021620858 and LC07032 of the Czech Ministry of Education, Youth and Sports.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

3´ss	3´ splice site
5´ss	5´ splice site
5-FOA	5-fluoroorotic acid
γ^{32}P-ATP	gamma-^{32}P-adenosin triphosphate
BBP	branch point binding protein
BCIP	5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt
BP	branch point
dBP	distant branch point
ddH_2O	double distilled water
DMF	dimethylformamide
EDTA	ethylenediaminetetraacetic acid
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
GFP	green fluorescent protein
HIV1	human immunodeficiency virus 1
hnRNP	heterogeneous nuclear ribonucleoprotein particle
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
LB	liquid broth
MFE	minimal free energy
NBT	nitro-blue tetrazolium chloride
nt	nucleotide
nts	nucleotides
NTC	Nineteen related complex (Prp19 associated complex)
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
P-TEFb	positive transcriptional elongation factor
pY-tract	polypyrimidine tract
qRT-PCR	quantitative reverse transcription polymerase chain reaction
RACE	rapid amplification of cDNA ends

RNApol II	RNA polymerase II
RNP	ribonucleoprotein particle
SD	synthetic dropout
SDS	sodium dodecylsulphate
SGD	Saccharomyces genome database
SL	synthetic lethal; synthetic lethality
SKIP	Ski-interacting protein
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particle
SR proteins	serin arginin rich proteins
TAP	tandem affinity purification
TBE	Tris borate EDTA
TPR	tetratricopeptide repeat
ts	temperature sensitive
U2AF65	U2 auxiliary factor, 65 kDa subunit
U2AF35	U2 auxiliary factor, 35 kDa subunit
UTR	untranslated region
UV	ultraviolet radiation
X-Gal	5-bromo-4-chloro-indolyl-galactopyranoside
YPAD	yeast extract, peptone, adenine, dextrose

ABSTRACT

Removal of introns from protein coding transcripts occurs in two splicing reactions catalyzed by a large nuclear complex, spliceosome. The spliceosome is an extremely intricate and dynamic machine, wherein contributions of small RNA molecules and multiple proteins are coordinated to meet the requirements of absolute precision and high flexibility. For an intimate understanding of pre-mRNA splicing, it is necessary to unravel roles of individual components and to dissect the partial mechanisms.

In the first part of this work, we describe the role of the Prp45 splicing factor in *Saccharomyces cerevisiae*. Mapping of genetic interactions of a conditionally lethal allele *prp45*(1-169) suggests a relationship of Prp45 to the NTC complex and to the second transesterification. Two-hybrid assay and purification of spliceosomal complexes reveal a contribution of the Prp45 C-terminus in the Prp22 helicase recruitment and/or regulation. Numerous experiments with reporter substrates document the need of Prp45 for the efficient splicing of a specific subset of introns. Our observations suggest that the function of Prp45 in splicing is conserved in evolution.

The second part is devoted to the role of intron secondary structure in 3´ splice site (3´ss) recognition. We show that the stem-loop structures formed downstream of the branch point (BP) are required for the splicing of *COF1* and *UBC13* introns, which have extremely long distances between BP and 3´ss. Identified structures aid to efficient 3´ss recognition by bringing remote 3´ss to the BP proximity and by sequestering AG dinucleotides, which behave as potential cryptic 3´ss. Our analyses strongly suggest that the structure based mechanism of 3´ss recognition is employed in most introns with distant 3´ss in *Saccharomyces cerevisiae* and possibly in other *Saccharomycotina* yeasts.

ABSTRACT IN CZECH

Odstraňování intronů z transkriptů probíhá prostřednictvím sestřihu v reakci katalyzované velkým jaderným komplexem – spliceosomem. Sestřih je nesmírně komplikovaný a dynamický proces, v němž koordinované fungování pěti malých molekul RNA a řady proteinů zajišťuje splnění požadavků na extrémní přesnost a flexibilitu. Pro důkladné pochopení sestřihu pre-mRNA je nezbytné rozklíčovat role jednotlivých komponent spliceosomu a porozumět všem dílčím mechanismům.

První část práce se zabývá rolí sestřihového faktoru Prp45 v kvasince *Saccharomyces cerevisiae*. Mapování genetických interakcí alely *prp45*(1-169) ukazuje na vztah mezi Prp45, NTC komplexem a druhým sestřihovým krokem. Analýza interakcí pomocí dvouhybridního systému a purifikace sestřihových komplexů dokladuje roli C-koncové části Prp45 v regulaci a/nebo vyvazování helikázy Prp22 do spliceosomu. Experimenty s reportérovými substráty prokazují, že Prp45 je vyžadován pro efektivní sestřih určité skupiny intronů. Naše pozorování podporují hypotézu, že role Prp45 v sestřihu je konzervována v evoluci.

Druhá část práce je věnována studiu vlivu sekundárních struktur intronů na identifikaci 3´ sestřihových míst (3´ splice site; 3´ss). Ukázali jsme, že „stem-loop“ struktura tvořená sekvencí následující za místem větvení (branch point; BP) je nebytná pro sestřih intronů *COF1* a *UBC13*, které mají extrémně dlouhou vzdálenost mezi BP a 3´ss. Identifikované struktury přispívají k efektivnímu nalezení 3´ss jednak přiblížením vzdálených akceptorových sekvencí do blízkosti BP a jednak maskováním dinukleotidů AG, chovajících se jako kryptická sestřihová místa. Naše analýzy jednoznačně podporují hypotézu, že sekundární struktury jsou využívány při nacházení 3´ss ve většině intronů se vzdálenými 3´ss u kvasinky *S. cerevisiae*, a pravděpodobně také v jiných kvasinkových organismech ze skupiny *Saccharomycotina*.

INTRODUCTION

Protein coding sequences in genes are frequently interrupted by non-coding regions – introns. The elimination of introns is one of the prerequisites for the production of messenger RNAs translatable into functional proteins. A process of intron removal from transcripts is called splicing and is catalyzed by a huge multi-component machine – the spliceosome. Splicing occurs in the cell nucleus and is spatially, temporary, and functionally coupled to the transcription (Orphanides and Reinberg, 2002; Will and Luhrmann, 2011).

Chemically, splicing consists of two consecutive transesterifications. It produces spliced mRNA, which is transported to the cytoplasm, and excised intron lariats, which are nucleolytically degraded. Despite the apparent simplicity of the reaction, the splicing represents one of the most complex of cellular processes. The spliceosome itself consists of five small nuclear RNAs (snRNAs) and more than 100 proteins (Bessonov et al., 2008b; Fabrizio et al., 2009). The snRNAs are positioned in the active site of spliceosome, form base pairs with transcripts, mediate intron recognition and facilitate splicing catalysis (Valadkhan, 2010). Proteins harbor multiple structural, enzymatic and regulatory functions, and support the dynamic network of interactions (Will and Luhrmann, 1997; Jurica and Moore, 2003; Valadkhan and Jaladat, 2010). Therefore, the spliceosome is a ribonucleoprotein (RNP) machine, which exhibits protein-dependent ribozyme activity (Wahl et al., 2009). To complement the complexity of the machinery, the additional associated protein factors link the splicing with the preceding and subsequent stages of gene expression.

The spliceosome assembles on the pre-mRNA in a stepwise manner. During whole process multiple rearrangements within the spliceosome occur, including the repeated reorganizations of the catalytic center (reviewed in Madhani and Guthrie, 1994; Staley and Guthrie, 1998; Newman, 2008). The pre-spliceosomal complexes, first step, and second step competent spliceosomes differ in the RNA-RNA, RNA-protein and protein-protein interactions. Transitions between the particular stages are associated with the release of specific proteins and with the recruitment of others.

Splicing must meet the demands for great precision. On the other hand, as it significantly contributes to the regulation of gene expression, it has to be flexible

enough to efficiently respond to the regulatory signals. These two apparently counteracting requirements are together reflected by the immense complexity of the splicing machinery. The regulatory role of the splicing is best represented by alternative splicing, which can give rise to hundreds of protein variants from a single gene and thereby enhances the complexity of proteomes (Chen and Manley, 2009). Importantly, the species with a largely reduced alternative splicing (e.g. *Saccharomyces cerevisiae*) take also advantage of splicing to regulate gene expression (Meyer and Vilardell, 2009). Transcript-specific modulation of splicing efficiency in response to the stress plays an important role in the adjustment of message levels (Pleiss et al., 2007a).

Although many aspects of splicing, including the chemical principle of the reactions and the composition of the spliceosome, were elucidated over recent decades, a lot of questions remain to be answered. Among these are the rather ambiguous roles of numerous splicing factors. In this work, we present a study on the role of the essential protein Prp45 in *Saccharomyces cerevisiae*. In yeast, Prp45 was implicated only in splicing (Albers et al., 2003). In higher eukaryotes, homologs of Prp45 were reported to co-regulate transcription initiation and elongation (reviewed in (Folk et al., 2004). Therefore, it is one of promising candidates which might be involved in the coupling of transcription to splicing.

Another fundamental splicing-related question is how the splicing machinery recognizes the sequences to be excised from pre-mRNA. Introns must be removed from pre-mRNA with single nucleotide accuracy to prevent messages from frame-shifts, premature stop codons, and the production of aberrant protein forms. On the sequential level, introns are defined only by three conserved motifs: the 5′ splice site and 3′ splice site on the respective ends of intron, and the sequence flanking the branch point adenosine. The recognition of the 5′ splice site and branch point is based on the base pairing with the U1 and U2 snRNAs, respectively (Parker et al., 1987; Wu and Manley, 1989; Rosbash and Seraphin, 1991). However, the mechanism of the 3′ splice site recognition remains unsatisfactorily understood. In the second part of this work, we will document that the secondary structures downstream of the branch point are required for the proper recognition of the remote 3′ splice sites in *Saccharomyces cerevisiae*.

AIMS OF THE STUDY

I. Elucidate the role of protein Prp45 in pre-mRNA splicing in *Saccharomyces cerevisiae.*

II. Explain the mechanism of distant 3′ splice site recognition in *S. cerevisiae*.

1. LITERATURE REVIEW

1.1 Spliceosomal introns

Nascent eukaryotic RNAs are composed of two types of sequential elements – exons and introns. Introns are sequences, which are transcribed, but eventually removed during the process of RNA maturation. In protein coding genes, introns represent non-coding sequences. Introns can be divided into two categories – (i) self-splicing introns and (ii) spliceosomal introns (Toor et al., 2009). Self-splicing introns are ribozymes with an intrinsic catalytic activity which are removed from transcripts in an auto-catalyzed reaction. The folding into a conserved secondary structure, wherein the reactive groups are juxtaposed, enables the formation of an active catalytic site without protein assistance (Lehmann and Schmidt, 2003). In contrast, spliceosomal introns are excised in a reaction conducted by a large ribonucleoprotein complex – spliceosome (Wahl et al., 2009; Will and Luhrmann, 2011).

Self-splicing introns are prevalent in prokaryotes and in the organelles of fungi, algae and plants, but were only rarely identified in nuclear eukaryotic genomes (Nielsen and Johansen, 2009; Lambowitz and Zimmerly, 2011). Spliceosomal introns are ubiquitous - their presence is a common feature of all eukaryotic genomes. However, intron densities vary significantly among taxons. In vertebrates, virtually all genes contain at least one, but usually more, introns. "Intron poor" organisms including *Saccharomycotina* yeasts have introns in less than 5 % of their genes (Schwartz et al., 2008). Only 273 nuclear genes in *Saccharomyces cerevisiae* have the coding sequence interrupted by an intron. Moreover, only nine genes contain more than one intron (Saccharomyces Genome Database, SGD; www.yeastgenome.org; June 2011). Also, intron length distribution is phylogenetically dependent. Whereas human introns commonly reach up to thousands of nucleotides, *S. cerevisiae* have only single intron longer than 1000 nt, and the median length is 141 nt (www.genome.jouy.inra.fr/genosplicing). The intron densities and exon-intron architecture in various species are summarized in Table 1.1.

The ends of introns, which make the boundaries with the adjacent exons, are called the 5´splice site (5´ss; donor site) and 3´splice site (3´ss; acceptor site). The 5´ss and 3´ss are characterized by the invariant terminal dinucleotides, 5´-GU and

Table 1. General statistics pertaining to the genomes and intron-exon architecture of the 22 analyzed organisms

	Genome statistics						Intron-exon statistics							
	Number of genes	% Non-int genes	% Uni-int genes	% Multi-int genes	Non-int gene length	Uni-int gene length	Multi-int gene length	Intron number	Intron density	Intron length	Exon length	% Introns in uni-int genes	% Introns in multi-int genes	External exon length	Internal exon length
A. thaliana	26043	20.5	13.6	65.9	981	1193	2372	114286	5.6	99	145	3.1	96.9	315	117
C. parvum	3396	99.0	0.9	0.1	1350	929	3154	44	1.3	66	323.5	70.5	29.5	341	165
D. discoideum	13416	31.6	35.0	33.4	1020	1187	1801	17326	1.9	103	270	27.1	72.9	285	248
S. cerevisiae	5850	95.6	4.3	0.2	1266	872	487	258	1.2	148	233	93.3	6.7	240	94
C. glabrata	5181	98.5	1.5	0.0	1278	1007	1920	80	1.1	487	246	95.0	5.0	246	342
K. lactis	5331	97.6	2.3	0.0	1188	957	489	127	1.0	273	240	98.4	1.6	243	23
E. gossypi	4718	95.4	4.5	0.1	1254	789	480	218	1.1	64	232	95.5	4.5	235	94
D. hansenii	6893	95.2	4.5	0.2	1128	988	721	346	1.1	89	221	90.2	9.8	237	62
Y. lipolytica	6520	89.9	9.2	0.9	1209	1362	1341	721	1.1	212	246	83.3	16.7	268	63
N. crassa	5894	20.1	35.2	44.7	1101	1134	1762	10023	2.1	84	215	20.7	79.3	227	196
M. grisea	10302	24.2	27.0	48.9	1008	1143	1778	19463	2.5	94	215	14.3	85.7	250	172
A. fumigatus	9923	21.8	30.4	47.8	1098	1149	1691	18212	2.4	61	253	16.5	83.5	282	218
S. pombe	5083	55.3	20.0	24.6	1185	1110	1343	4734	2.1	56	172	21.5	78.5	204	138
U. maydis	6495	62.3	19.7	18.0	1662	1250	1602	4878	2	95	219	26.3	73.7	280	151
C. neoformans	6475	3.0	7.7	89.3	990	947	1805	33740	5.5	55	149	1.4	98.6	172	142
C. elegans	19246	3.4	9.7	87.0	389	657	2113	98695	5.3	65	150	1.9	98.1	140	154
D. melanogaster	13287	20.1	19.6	61.3	955	1189	3205	41145	3.9	72	264	5.9	94.1	369	217
Zebrafish	25111	3.1	5.8	91.2	1161	2981	14377	194221	8.3	946	128	0.7	99.3	169	123
Chicken	16949	2.3	8.1	89.7	966	1221	12836	167626	10.3	841	130	0.8	99.2	189	125
Dog	20106	9.7	8.5	81.7	963	1679	21749	172059	9.7	1271	129	1.0	99.0	191	124
Mouse	23071	12.8	6.6	80.6	957	2506	20508	177766	9	1323	133	0.8	99.2	279	124
Human	22120	8.0	7.3	84.6	1195	2270	25108	184145	9.2	1516	133	0.9	99.1	305	124

Genes are divided into three categories: Genes lacking introns (non–int genes), genes containing one intron (uni–int genes), and multi-intronic genes containing two or more introns (multi–int genes). All measures of length are medians. Intron density is the mean number of introns per spliced gene. All first or terminal exons were defined as external; all remaining introns were defined as internal.

Table 1.1. Intron-exon architecture in selected species. Adopted from Schwartz et al., 2008.

AG-3′, respectively. A minority of introns have the alternative 5′-AT and AC-3′ termini. These so-called "ATAC" introns represent a small fraction of the intervening sequences in plants, insects, or vertebrates. They employ a partially distinct splicing machinery with the alternative snRNAs (U11, U12, U4atac and U6atac; U5 is common with the standard spliceosome) which are absent from *S. cerevisiae* (reviewed in Patel and Steitz, 2003; Will and Luhrmann, 2005).

The adenosine nucleotide that is involved in a lariat structure formation (see below) is called the branch point (BP). All three splicing signals – the 5′ss, 3′ss and the sequence flanking BP adenosine - are essential for intron recognition (Fig. 1.1A). Their sequences are conserved within the species' intron set as well as in evolution. Yet, the extent of the conservancy varies among organisms (Fig. 1.1B). Intron poor species are generally characterized by highly conserved and strong splicing signals.

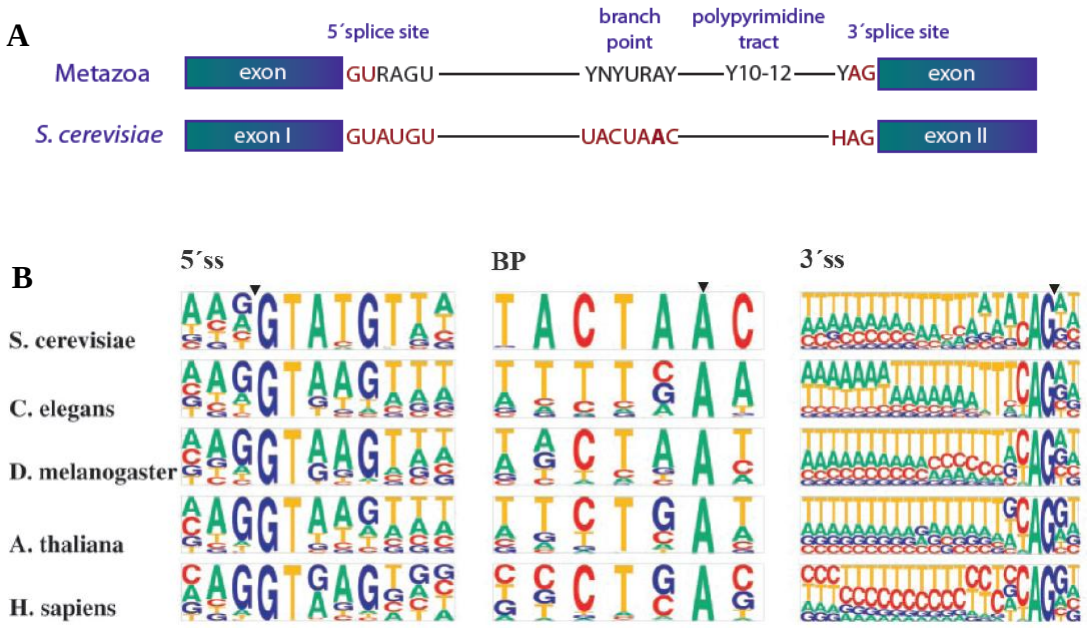


Figure 1.1. (A) Schematic representation of introns from Metazoa and *Saccharomyces cerevisiae*. (B) Splicing signal sequences in selected model species. The height of each letter is proportional to the frequency of the corresponding base at the given position. Positions of intron/exon boundaries and BP adenosine are marked by triangle. Adapted from (Lim and Burge, 2001).

For example, the BP in *S. cerevisiae* is flanked by an almost invariant UACUAAC sequence (the branch point adenosine is underlined). The other extreme can be observed in vertebrates, wherein the sequences of splicing signals are much more heterogeneous (Lim and Burge, 2001; Iwata and Gotoh, 2011). Efficient intron

recognition in the species with the weak splicing signals is dependent on the additional sequential motifs; a polypyrimidine tract (pY tract) downstream of BP and exonic or intronic splicing enhancers (ESE, ISE) or silencers (ESS, ISS) complement the information encoded in the primary splicing signals (Cartegni et al., 2002; Singh and Valcarcel, 2005). In *S. cerevisiae*, any signals with analogical functioning have not been identified.

1.2 Chemical principle of pre-mRNA splicing

Intron removal in the spliceosome-mediated reaction is, together with the 5´ end capping and the 3´ end polyadenylation, one of the processing steps which the RNA polymerase II transcripts undergo in the nucleus before being exported to the cytoplasm where they serve as templates for the protein synthesis on ribosomes.

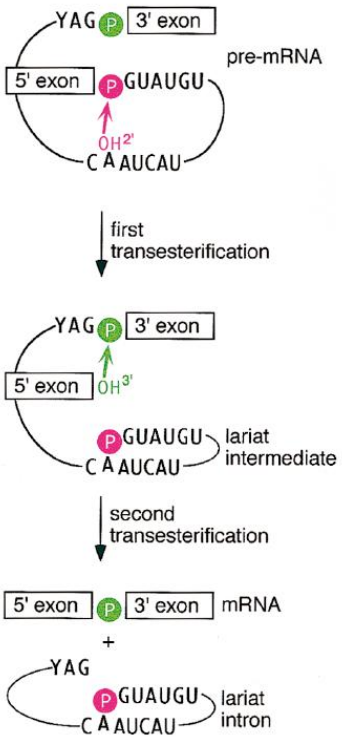


Figure 1.2. Splicing occurs in two consecutive transesterifications. The first transesterification reactants are shown in pink, the second transesterification reactants in green. Adopted from (Staley and Guthrie, 1998).

Splicing entails two consecutive transesterifications (Moore and Sharp, 1993; Fig. 1.2). When the spliceosome assembles on the RNA molecule, the first transesterification is initiated by a nucleophilic attack of the BP adenosine 2´-hydroxyl group to the phosphodiester bond in the 5´ss. The first step ends up with

a free 5′ exon and a lariat structure is formed by the intron and the 3′ exon. In this splicing intermediate product, the intron is circularized by a bond formed between the 5′ss and the BP nucleotide. In the second step, the active 3′-hydroxyl group of the 5′ exon attacks the phosphodiester bond on the intron-3′exon boundary (3′ss) and both exons are ligated. The spliced mRNA is released from the spliceosome and associates with the export machinery. The excised intron lariat is degraded and the spliceosomal components are recycled. The transesterifications themselves do not require any direct energy supply. However, splicing is an energy-consuming process, as the activities of several protein factors involved in the spliceosomal rearrangements are NTP-dependent (see section 1.7 Driving forces in splicing).

1.3 Composition of spliceosome

The spliceosome represents one of the cellular ribonucleoprotein (RNP) machines, in which different chemical properties of RNA and proteins are orchestrated to guarantee the specific function. Unlike most RNP complexes with a catalytic function, which are predominantly stable and contain ready-to-serve active sites, the spliceosome is an incredibly dynamic complex, in which the catalytic centre is formed and repeatedly remodeled in the substrate-dependent manner.

Spliceosome consists of five small uridine-rich RNAs, referred to as U1, U2, U4, U5 and U6 snRNAs. Lengths of snRNAs vary between 100 and 200 nt. A comparison of snRNAs from unrelated species reveals, that snRNA have an extremely phylogenetically conserved primary as well as secondary structure (Guthrie and Patterson, 1988; Mitrovich and Guthrie, 2007). Each snRNA is associated with a specific set of proteins which together form the ribonucleoprotein particle (snRNP; Brow, 2002). In addition, each of the U1, U2, U4, and U5 snRNPs contains a heptamer ring complex of seven Sm proteins (for review, see Will and Luhrmann, 2001). Besides the snRNP components, the additional non-snRNP proteins are employed in the splicing machinery during whole process. The total number of engaged proteins has been estimated by independent studies as between 100 and 300. This variability comes from the technique used and the conditions of purification. Although the splicing machinery is conserved in evolution, the composition is to some extent also species dependent. For example, *Saccharomyces cerevisiae* and other intron-poor organisms, which lack alternative splicing, do not

need some regulatory factors, e. g. SR or hnRNP proteins (Fabrizio et al., 2009). For the proteomic composition of yeast and human spliceosomes see Fig. 1.3.

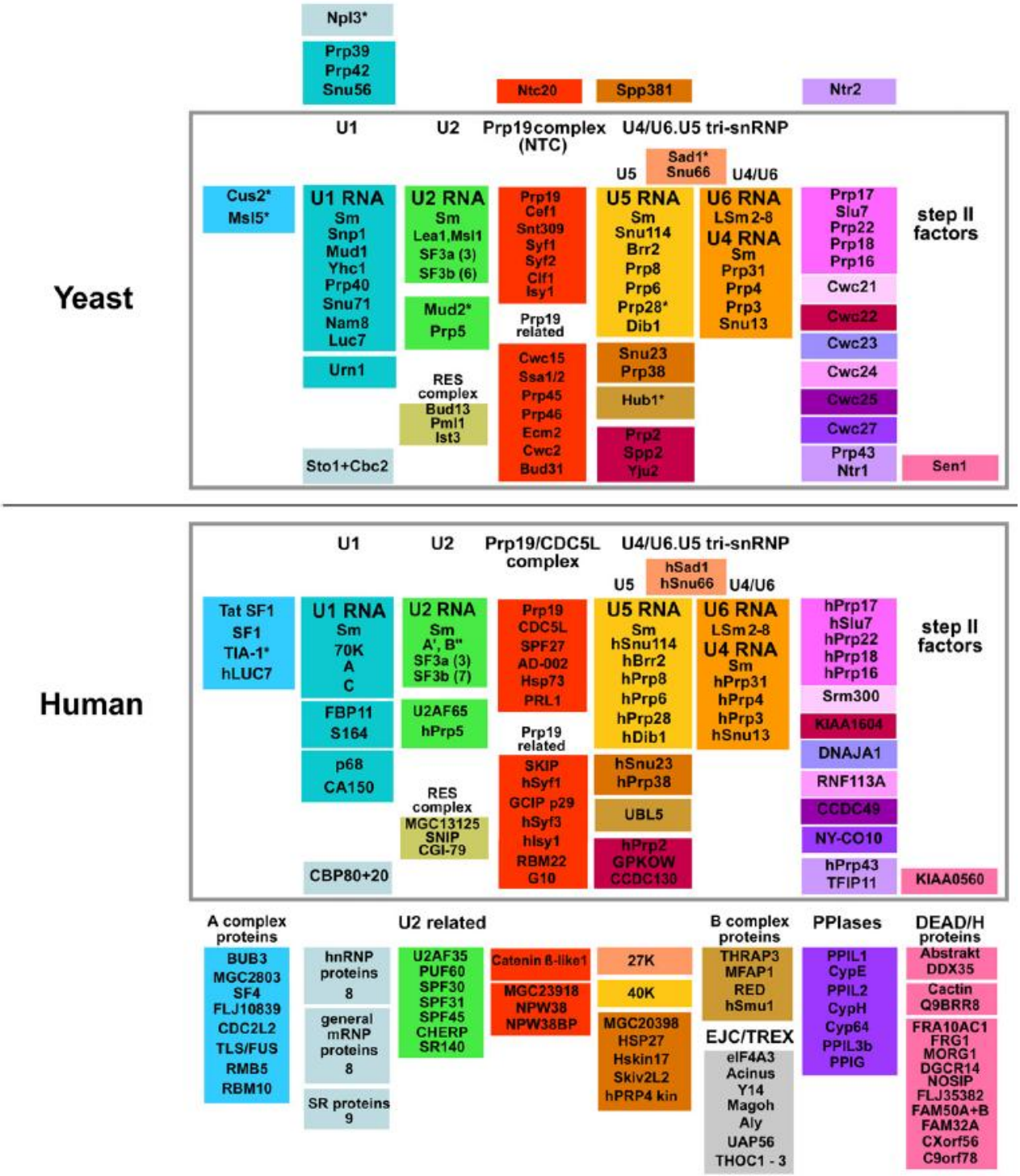


Figure 1.3. Comparison of spliceosomal composition in humans and *S. cerevisiae*. The proteins in rectangles have homologs in the opposite species. Adopted from Fabrizio et al., 2009.

1.4 Formation of prespliceosome

Stepwise spliceosome assembly is initiated by an ATP-independent recruitment of the U1 snRNP to the 5′ss (for summary of splicing cycle see Fig. 1.4). The 5′ terminus of the U1 snRNA base pairs with the 5′ss (Siliciano and Guthrie, 1988; Seraphin et al., 1988). In higher eukaryotes, the interaction is stabilized by U1 snRNP and SR proteins (reviewed in Konarska, 1999). Surprisingly, the U1 snRNP can associate with the 5′ss even in the absence of the sequence which is complementary to the 5′end of the intron (Du and Rosbash, 2001). Recently, a novel non-splicing function of the U1 snRNP has been reported in mammalian cells. The U1 snRNA was shown to base pair with 5′ss-resembling sequences in nascent transcripts and to prevent them from a premature termination (Kaida et al., 2010).

At the same time when the U1/5′ss interaction is being established, the BP sequence and downstream elements are also recognized. In mammals, SF1 and U2AF65 (large 65 kDa subunit of U2 auxiliary factor) cooperatively contact the BP and the polypyrimidine tract (pY-tract), respectively (Berglund et al., 1998; Singh et al., 2000). A small subunit, U2AF35, forms a stable heterodimer with U2AF65 and interacts with the AG dinucleotide of 3′ss, but only if it is located in the close proximity of the pY-tract (Merendino et al., 1999; Wu et al., 1999). In *Saccharomyces cerevisiae*, SF1 and U2AF65 are functionally substituted by the BBP (branch point bridging protein, encoded by *MSL5* gene) and Mud2 proteins, respectively (Abovich and Rosbash, 1997; Berglund et al., 1997). There is neither a homolog nor functional equivalent of U2AF35 in budding yeast (Wang et al., 2008; reviewed in Rymond, 2010). Interactions described here represent the initial contact of the splicing machinery with the 5′ss and BP/pY-tract/3′ss region. Although BBP was reported to contact the U1 snRNP protein Prp40 (Abovich and Rosbash, 1997), the recognition of 5′ and 3′ intron termini by the U1 snRNP and BBP/Mud2, respectively, seems to be rather independent. The resulting assembly is designated as either complex E or commitment complex (CC).

In the cascade of the spliceosomal building, U2 snRNP recruitment follows. The U2 snRNA base pairs with a BP-flanking sequence (Wu and Manley, 1989) displacing the BBP/SF1 protein. The branch point adenosine itself is excluded from the base pairing and is “bulged out”. This set up ensures a nucleophilic reactivity of its 2′-hydroxyl in the first transesterification (Query et al., 1994; Smith et al., 2009).

In *S. cerevisiae*, absolute complementarity between the almost invariant BP sequence and U2 snRNA is sufficient to maintain their mutual interaction. In mammalian cells, where the BP sequence is rather degenerate (YNCURAY), the weak base pairing requires stabilization by the additional oligomeric proteins, namely SF3a and SF3b (Gozani et al., 1996). BBP/SF1 and Mud2 displacement is driven by the activity of the ATP-dependent Sub2 helicase from a DExD/H box family (Kistler and Guthrie, 2001).

An additional ATPase activity is represented by another DExD/H box helicase – Prp5 (Perriman and Ares, Jr., 2000; Perriman et al., 2003). According the model, Prp5 removes the U2 snRNA binding protein Cus2. Its dissociation enables the remodeling of an internal U2 structure (stem IIc → stem IIa) and the generation of the splicing competent U2 snRNA conformation. The U2 snRNP containing complex is called complex A or prespliceosome.

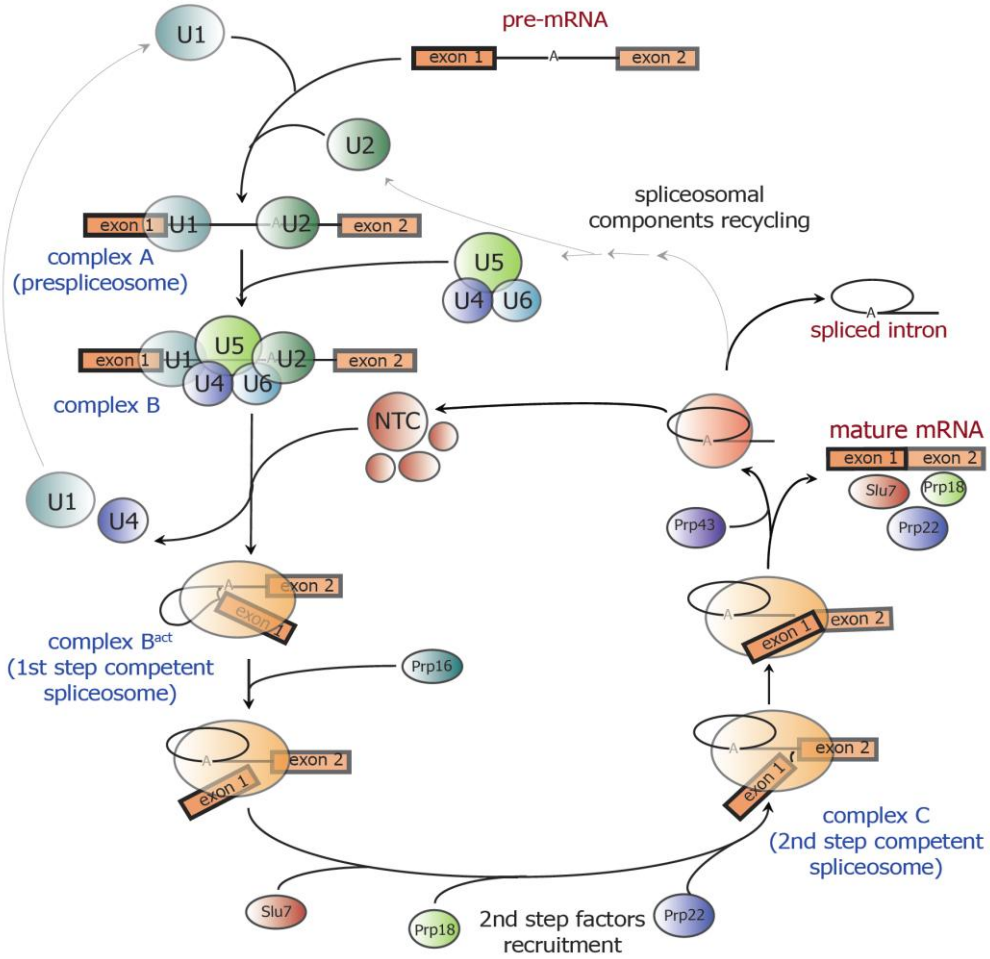


Figure 1.4. Splicing cycle. Stepwise recruitment of snRNPs on pre-mRNA is shown together with the nomenclature of rearranging splicing complexes. In addition, several splicing factors discussed in this work are depicted.

The early stages of spliceosomal assembly described here involve "across-intron" interactions between the complexes associated with both intron termini. However, in the typical mammalian gene, which contains multiple long introns (more than ~250 nt) separating relatively short exons, an alternative assembly pathway occur. The U1 snRNP bound to the 5´ss communicates with proteins interacting with splicing enhancers in the upstream exon and facilitate the recruitment of U2AF to the pY-tract/3´ss of the preceding intron, forming together "exon definition complex" (Berget, 1995; Fox-Walsch et al., 2005). For the subsequent steps of splicing reaction, the 5´ss must be correctly paired with the respective downstream 3´ss. The mechanism of this switch from the "across-exon" to the "across-intron" complex is at present only poorly understood. Importantly, this rearrangement contributes to the final localization of the intron-exon boundaries. Consequently, the exon inclusion/skipping events can occur during the switch from the exon definition complex to the across-intron assembly (Bonnal et al., 2008; Sharma et al., 2008).

1.5 Dynamic RNA-RNA interactions during spliceosome assembly and splicing

In ongoing events, the prespliceosome is recognized by a preformed U4/U6•U5 tri-snRNP particle and the complex B is formed. Two intermolecular helices mediating the original U4/U6 interaction are unwound and a new U2/U6 snRNA base pairing is established (Madhani and Guthrie, 1992; Sun and Manley, 1995). An intramolecular stem loop in the U2 snRNA downstream of the BP binding sequence is remodeled into a different, mutually exclusive stem structure. To achieve the catalytically active spliceosome – complex B^{act} – the additional extensive reformations in architecture occur (Fig. 1.5). The U1 and U4 snRNPs are released and the nucleotides 4 to 6 of introns are paired to the conserved sequence in the U6 snRNA. This interaction is mutually exclusive with the U1/5´ss interaction and seems to be responsible for the precise determination of the 5´ss phosphodiester bond to be cleaved (Sontheimer and Steitz, 1993). The U2/U6 and U6/5´ss interactions are crucial for juxtaposition of the BP and 5´ss and for the promotion of the first splicing reaction (for the review of dynamic RNA-RNA interaction, see Newman, 2008).

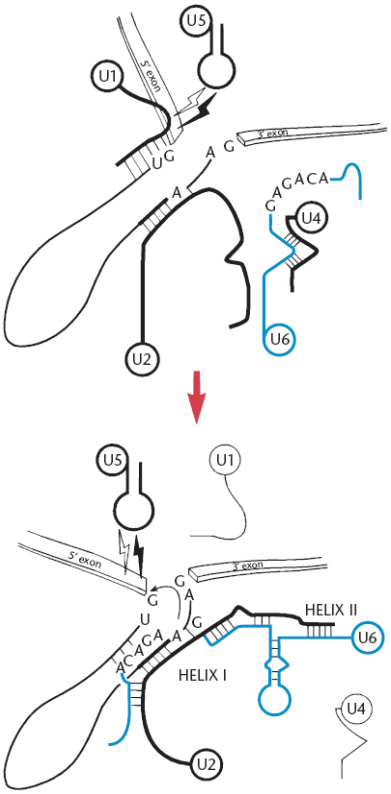


Figure 1.5. Changes in the RNA-RNA interactions during the activation of spliceosome (formation of B^{act} complex). Adapted from Newman, 2008.

The first transesterification ends up with an assembly called complex C. Multiple protein factors are afterwards recruited to achieve the second step competent spliceosome (Jones et al., 1995; James et al., 2002). A conserved loop of U5 snRNA with invariant nine nucleotides (GCCUUUUAC) interacts with intron adjacent sequences of both exons and promotes an alignment of the exons for the second transesterification. The loop interaction with the 5´ exon is established prior the first transesterification, while the 3´ exon is contacted during the transition into the second step conformation (Newman and Norman, 1992; Newman, 1997). Exon ligation is followed by release of spliced mRNA and the spliceosome disassembles.

1.6 Proteins in splicing

Spliceosomal protein composition was independently studied by several groups. The spliceosomal complexes were assembled *in vitro* on labeled intron containing RNA molecules and co-purified proteins were analyzed by mass spectrometry (Rappsilber et al., 2002; Zhou et al., 2002; Jurica et al., 2002). The improved techniques enabled the analysis of the spliceosomes stalled in the specific stages of assembly or catalysis (Makarova et al., 2004; Behzadnia et al., 2007; Bessonov et al., 2008a; Bessonov et

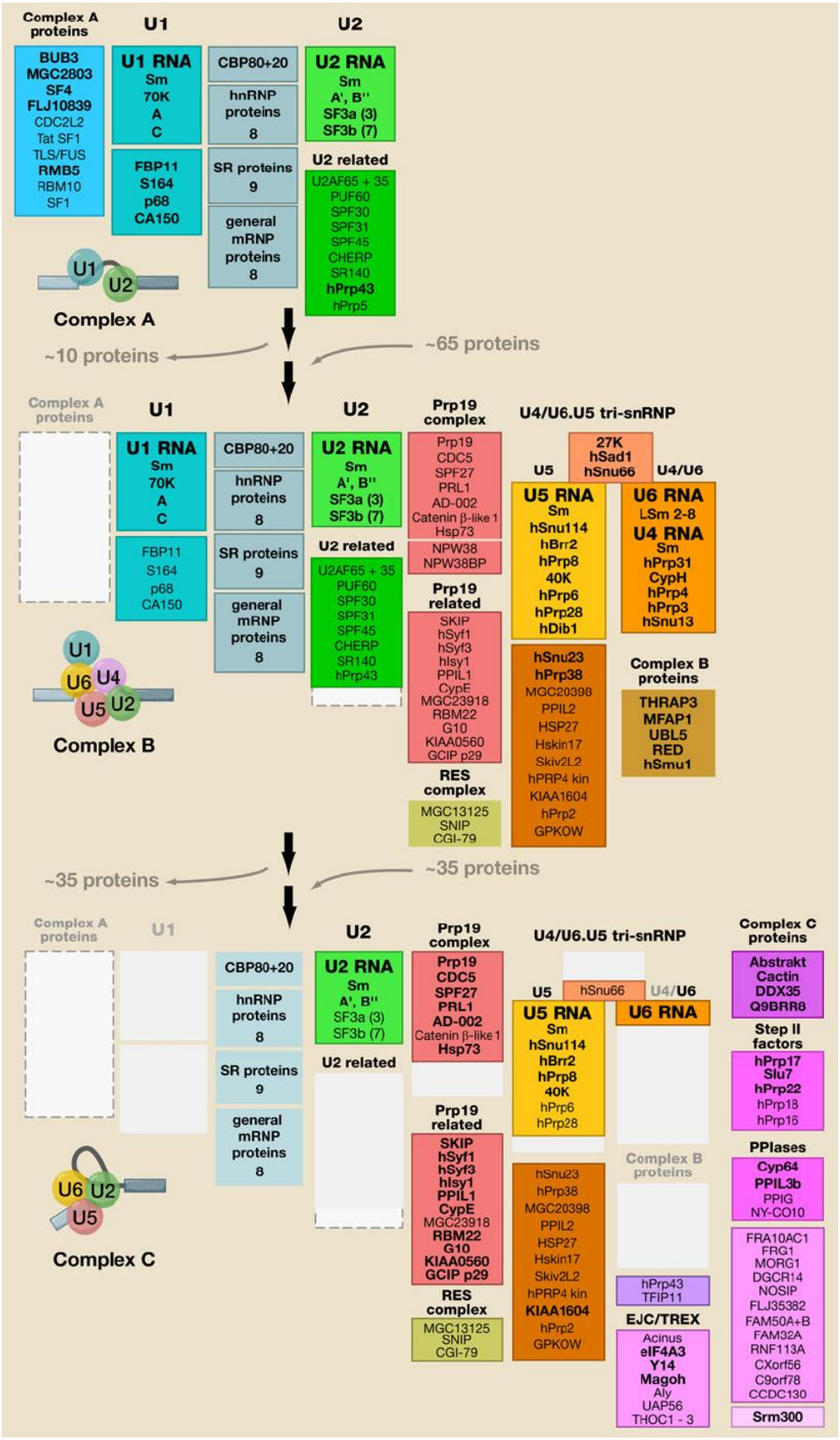


Figure 1.6. Dynamics of spliceosomal composition. Protein composition of human spliceosomal complexes A, B, and C, as determined by mass spectrometry (Behzadnia et al., 2007; Bessonov et al., 2008b). Adopted from Wahl et al., 2009.

al., 2010). While human spliceosomes are estimated to contain 150 to 300 proteins, the splicing complexes from *Saccharomyces cerevisiae* have a reduced composition to about 100 proteins. However, most proteins in yeast have orthologs in humans, showing a conservancy of the core splicing machinery (Fabrizio et al., 2009).

Dynamics in the protein composition goes hand in hand with the multiplicity of rearrangements, which the spliceosome undergoes during its assembly and splicing reaction. Protein factors are recruited in different stages of assembly. Some activities are present only transiently and leave the splicing machinery immediately after carrying out a specific task. During the activation of complex B ($B \rightarrow B^{act}$) concomitantly with the U1 and U4 snRNP dissociation, around 35 proteins leave the spliceosome and twelve other proteins are recruited. Accomplishing the first step and the transition into the second step conformation is accompanied with the release of two proteins and the association of nine second step factors (Fabrizio et al., 2009). A summary of protein dynamics in the spliceosome is shown in Fig. 1.6.

1.7 Driving forces in splicing

Spliceosome as a complex composed of both RNA and proteins is considered to be a ribozyme. Therein, neither proteins nor RNAs itself would be able to guarantee a function without cooperating each other. SnRNAs mediate the main interactions with a substrate and are crucial for its positioning within the catalytic centre. Proteins carry out variety of essential functions – structural, regulatory as well as enzymatic. There are multiple proteins forming the core of snRNP (Sm proteins; Will and Luhrmann, 2001) and the spliceosome itself (Prp8; Grainger and Beggs, 2005). Many factors are required for the stabilization of numerous weak RNA-RNA interactions, which are characteristic for spliceosome. Adjusting the stability of these interactions is one of the principles of how the splicing is regulated.

Most steps in spliceosomal assembly and rearrangements are powered by ATP-dependent enzymes from the family of DExD/H box helicases. Apart from splicing, the DExD/H box helicases are involved also in other stages of RNA metabolism, including RNA export, degradation or ribosomal biogenesis (reviewed in de la Cruz et al., 1999 and Pyle, 2008). In splicing, these proteins are employed in unwinding of double stranded RNA regions or interrupting RNA-protein interactions. Their activities provoke specific RNP remodeling events. There are eight splicing-related

DExD/H box helicases in *S. cerevisiae* – Sub2, Prp5, Prp28, Brr2, Prp2, Prp16, Prp22 and Prp43 (reviewed in Staley and Guthrie, 1998; for summary of see Fig. 1.7). All of them have an ortholog with the same function in mammals, demonstrating their importance for the splicing.

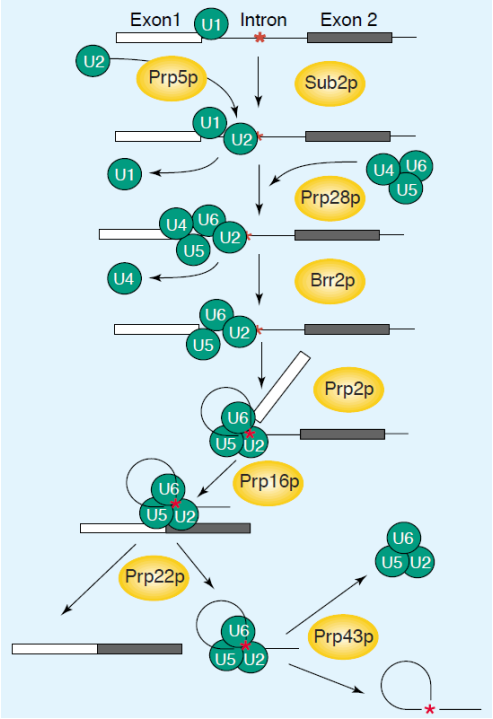


Figure 1.7. RNA helicases involved in pre-mRNA splicing. Adopted from (de la et al., 1999).

Sub2 and Prp5 are required in the early stages of spliceosomal development. Sub2 facilitates the release of Mud2 and promotes the U2 snRNP association (Kistler and Guthrie, 2001). Prp5 acts as an essential remodeling factor of internal U2 snRNA interactions (Perriman et al., 2003). As soon as all snRNPs are associated with pre-mRNA substrate, Prp28 disrupts the base pairing between the U1 snRNA and the 5´ss. Thus, it is a key player in the substitution of U1 by U6 snRNA on the intron 5´end (Staley and Guthrie, 1999; Chen et al., 2001c). Two additional ATPase/helicase activities are needed for the spliceosomal activation. Brr2 unwinds the U4/U6 snRNA helix, promoting the U4 snRNP dissociation from the spliceosome (Raghunathan and Guthrie, 1998b). The U4/U6 re-annealing during the U4/U6•U5 tri-snRNP recycling is conducted by the RNA binding protein Prp24 (Raghunathan and Guthrie, 1998a). The role of the other essential helicase, Prp2, is less understood. Its activity is also required for the U1 and U4 release and for the first step promotion (Kim and Lin, 1996).

Prp16 helicase is recruited to the spliceosome via its non-conserved N-terminal domain after the first transesterification (Wang and Guthrie, 1998). It mediates structural changes resulting in the second step conformation (Schwer and Guthrie, 1992). Prp16 can be crosslinked to the intron/3´exon boundary (McPheeters and Muhlenkamp, 2003). Prp22 exhibits two essential functions. Its ATP-independent activity is required for the second step. The ATP-dependent helicase activity facilitates spliced mRNA release after exon ligation (Company et al., 1991; Schwer and Gross, 1998; Wagner et al., 1998; Schwer, 2008). Spliceosome disassembly is further dependent on the second function of Brr2. As in the case of the U4/U6 snRNA unwinding Brr2 exhibits here an activity modulated by a small GTP protein, Snu116 (Small et al., 2006). Last helicase involved in splicing, Prp43, drives the dissociation of the excised intron lariat (Arenas and Abelson, 1997; Martin et al., 2002).

Importantly, some of helicases not only supply power for the rearrangements but also aid to splicing fidelity. The kinetics of ATPase/helicase mediated reaction ensures a proofreading of the splicing substrate quality. Suboptimal substrates, which are not converted rapidly enough within a helicase-controlled timeframe, are discarded from the complex and degraded (Fig. 1.8). This proofreading activity was reported for the Prp5, Prp16, Prp22 and Prp43 enzymes (Burgess and Guthrie, 1993; Konarska et al., 2006; Xu and Query, 2007; Mayas et al., 2006; Smith et al., 2009; reviewed in Query and Konarska, 2006). Recently, Prp16 and Prp43, helicases which were considered to act in the later stages of splicing, were reported to cooperate in the proofreading of the 5´ss cleavage (Koodathingal et al., 2010). Furthermore, Prp43 turned out to be the enzyme responsible for the rejection of suboptimal intermediates in the Prp22 dependent pathway (Mayas et al., 2010). The RNA helicases thus represent main activities guaranteeing splicing accuracy and preventing from the production of potentially deleterious aberrant messages.

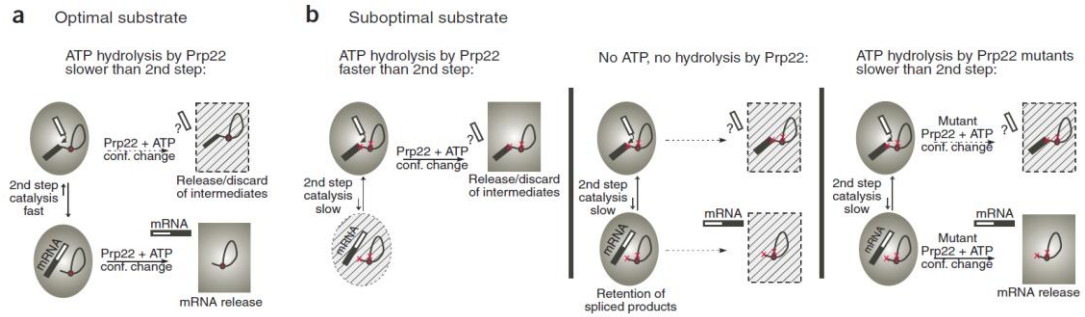


Figure 1.8. Model for Prp22 functioning in splicing fidelity. (A) Optimal substrates are spliced fast enough, mRNA is released. **(B)** The slow rate of the second step catalysis of suboptimal substrates provides Prp22 enough time to discard splicing intermediate. Absence of ATP prevents Prp22 from discarding intermediate or releasing mRNA. *prp22* mutants with decreased rate of ATP hydrolysis result in increased splicing efficiency of suboptimal substrates. Adapted from Query and Konarska, 2006.

1.8 Splicing-mediated gene expression regulation in *S. cerevisiae*

The cardinal contribution of splicing for gene expression regulation lies in alternative splicing (for review see Chen and Manley, 2009). However, some intron poor organisms, including *Saccharomyces cerevisiae*, which have usually only one intron per gene, almost lack the alternative splicing. The first recently documented observation of the genuine alternative 5´ss selection in the *SRC1* gene (Mishra et al., 2011) and of the intron retention in the two introns containing the *SUS1* gene (Hossain et al., 2009; Hossain et al., 2011) still seem to be exceptions rather than the rule. Yet, the splicing can also serve as a gene expression regulatory mechanism in species where the alternative splicing is absent. The splicing efficiency of specific substrates can be tuned up, resulting in the regulation of the message levels. Upon amino acid starvation or ethanol stress, the splicing efficiencies of many transcripts are specifically changed, as shown by splicing specific microarrays (Pleiss et al., 2007a).

Another example of posttranscriptional gene expression regulation by splicing was documented for the meiotic genes *AMA1*, *MER2*, *MER3*, and others. Some of these genes are transcribed constitutively, but their splicing occurs only after the induction of meiosis. The Mer1 protein, a meiosis-specific splicing enhancer, binds its intronic responsive element (Spingola and Ares, Jr., 2000) and by multiple mechanisms it activates splicing (Spingola et al., 2004; Scherrer, Jr. and Spingola, 2006). Among others, Mer1 stabilizes the interaction of U1 snRNA with the

suboptimal 5´ss of *MER2* and *MER3* introns, and is therefore required for spliceosome assembly (Nandabalan and Roeder, 1995).

Several proteins regulate the splicing of their own transcripts. L30, a ribosomal protein encoded by *RPL30* gene, binds a "kink-turn" structure formed by the 3´ end of the first exon and the 5´ss in the *RPL30* pre-mRNA. This binding does not prevent the *RPL30* pre-mRNA from the association with the U1 snRNP nor with the BBP/Mud2 dimer. Instead, it disables the recruitment of the U2 snRNP and the spliceosome formation (Macias et al., 2008). The inability to regulate its own splicing caused by *RPL30* mutation can be suppressed by multiple mutations in the cap binding complex (CBC) protein Cbp80, suggesting the role of the CBC in the U2 snRNP recruitment (Bragulat et al., 2010). For review of *S. cerevisiae* splicing regulation, see (Meyer and Vilardell, 2009).

1.9 SNW proteins

SNW proteins are essential nuclear proteins that seem to be widespread throughout the eukaryotes (identified as Snw1p in *Schizosaccharomyces pombe*, (Ambrozkova et al., 2001); SnwA in *Dictyostelium discoideum*, (Folk et al., 1996); Bx42p in *Drosophila melanogaster*, (Wieland et al., 1992); SNW/SKIP/NCoA-62 in human, (Dahl et al., 1998). The denomination "SNW" originates in a common and almost completely conserved SNWKN motif which occurs in the central part of the amino acid sequence (Fig. 3.19).

The SNW proteins are involved in various gene expression processes (reviewed in (Folk et al., 2004); Fig. 1.9). The human ortholog, SNW/SKIP, acts as a co-activator or co-repressor in transcription initiation. It was shown to interact with a number of transcription factors and signaling pathways' effector molecules, including Smad factors (Leong et al., 2001), nuclear steroid hormone receptors (Zhang et al., 2001), MyoD (Kim et al., 2001), NotchIC (Zhou et al., 2000), or LEF1 (Wang et al., 2010). SNW1/SKIP also interacts with the pRb antioncogene and inhibits its activity in the repression of transcription (Prathapam et al., 2002). In addition, SNW/SKIP affects an elongation phase of transcription. It interacts with P-TEFb and HIV1 Tat protein in the complex with the paused RNA pol II and, via the recruitment of additional factors (c-Myc, TRAPP), it contributes to the effective viral transcription (Bres et al., 2005; Bres et al., 2009).

Proteomic studies of active spliceosomal complexes and post-spliceosomal species revealed the association of SNW/SKIP with the splicing machinery (Makarov et al., 2002; Bessonov et al., 2008b). A truncation of SNW/SKIP resulted in the accumulation of unspliced transcripts from a SKIP-co-regulated reporter gene (Zhang et al., 2003), and SNW/SKIP overexpression altered the splicing of the HIV-1 transcript (Bres et al., 2005). Recently, SKIP was shown to be required for the splicing of p21(Cip1) transcripts upon a stress induction, presumably by recruiting U2AF65 (Chen et al., 2011).

Taken together, the SNW proteins are remarkable as potential coupling factors of transcription with co/post-transcriptional RNA processing.

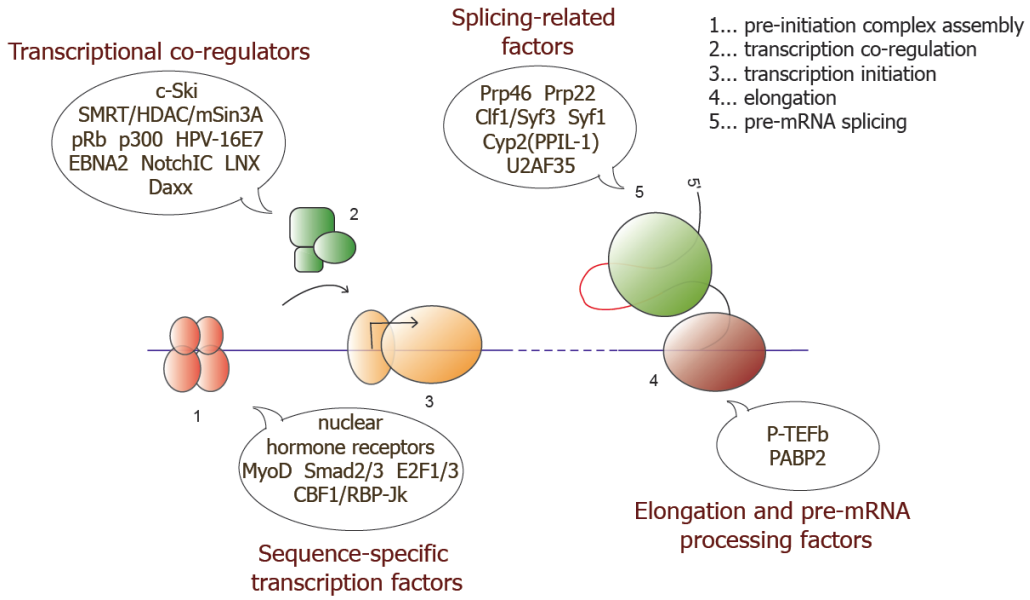


Figure 1.9. SNW proteins engagement in various stages of gene expression. The cartoon depicts SNW interactors that are categorized according to their function in the gene expression.

1.10 Prp45

Prp45 (pre-mRNA processing) is the SNW representative in *Saccharomyces cerevisiae*. Its gene was originally identified on chromosome 1 as an essential ORF with an unknown function (*FUN20*, function unknown; (Harris et al., 1992). *PRP45/FUN20* encodes a protein with the length of 379 amino acid residues and the molecular weight of 42 kDa. Unlike its orthologs from higher eukaryotes, it lacks the G-rich region containing the N-terminal part. The polyproline motif is reduced to only two proline residues, as compared to six to eight prolines in metazoal homologs. Sequence and motifs in Prp45 and other SNW proteins are illustrated in Fig. 3.19.

In contrast to human SNW/SKIP, Prp45 has been so far described to participate only in splicing. Prp45 was shown to be associated with the spliceosome during both splicing steps. As well as its counterpart from the evolutionary unrelated yeast *Schizosaccharomyces pombe*, SNW1, and human SNW/SKIP, Prp45 was found to co-purify with Cef1, Prp19, Cwc2, Syf1, Yju1, Clf1/Syf3, and Prp46, suggesting its relationship with the NTC complex (Rappsilber et al., 2002; Ohi and Gould, 2002; Jurica et al., 2002). The Beggs laboratory documented that Prp45 is present in the spliceosome prior to the recruitment of Prp2, the RNA helicase that initiates the first transesterification (Albers et al., 2003). After splicing is finished, Prp45 remains associated within the complex containing the U2, U5, and U6 snRNP, and the spliced intron lariat. Cells with downregulated *PRP45* transcription (under *GAL1* promoter in repressive conditions) accumulate unspliced transcripts. Prp45 was shown to interact in a two hybrid system with the Prp22 helicase and with the NTC-related Prp46 and Syf3 proteins. The interaction with Prp46 was confirmed by coimmunoprecipitation (Albers et al., 2003). Although authors suggested Prp45 in the regulation of Prp22 during mRNA release, the specific molecular function of Prp45 remains unknown.

In our laboratory, in an effort to search for the essential motifs, the *PRP45* gene was subjected to directed mutagenesis. Modified variants were used to test for growth complementation of *prp45-Δ* cells. Surprisingly, the deletion of the conserved SNW motif or its substitution by three alanines (SNW → AAA) turned out not to be deleterious. The viability of cells producing either of these variants was not distinguishable from the wild-type *PRP45* expressing cells. An analysis of truncated variants showed that both the N- and C-termini are dispensable for the viability. However, a simultaneous deletion of both ends is lethal (Martinkova et al., 2002). Importantly, the cells expressing the truncated variant with 168 N-terminal amino acid residues, *prp45*(1-168), are temperature sensitive. They grow at the same rate as wild-type cells at 30°C, but stop dividing within 24 hours from a shift to 37°C. For summary of complementation experiments see Fig. 1.10.

We used homology recombination to generate a chromosomal allele *prp45*(1-169). The resulting temperature sensitive strain is the main tool used in this study to unravel Prp45 role in splicing and will be described in section 3.1.1 Prp45(1-169).

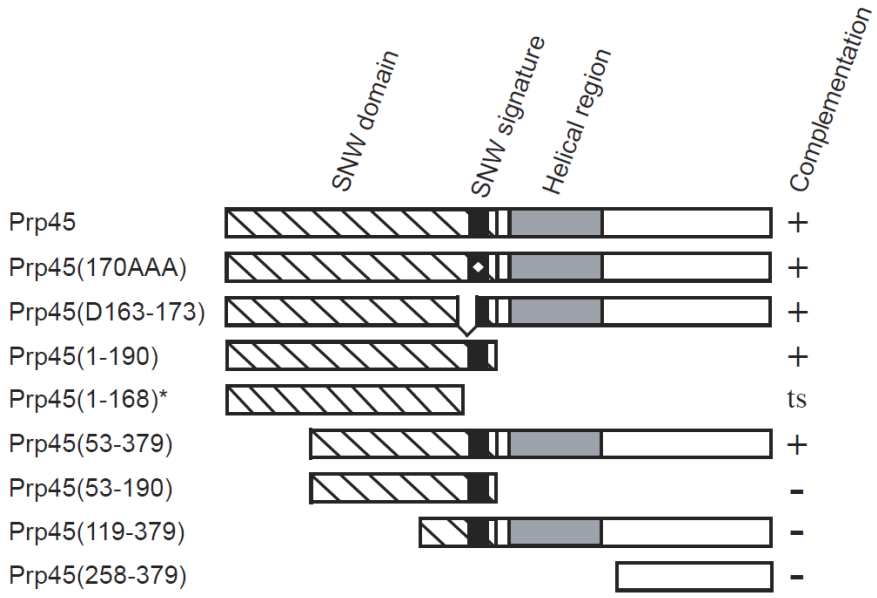


Figure 1.10. Schematic representation of truncated and manipulated Prp45 variants generated in our lab. Ability to complement *prp45*-Δ lethality is shown for each variant. "ts" stands for "temperature sensitive" complementation. **prp45*(1-168) is the variant expressed from plasmid. Corresponding genomic knock-in has one additional amino acid (*prp45*(1-169); see Results).

1.11 NTC complex

According to its co-purification with Cef1 (Ohi et al., 2002) and Clf1/Syf3 (Wang et al., 2003), Prp45 belongs to the group of "Prp19 related proteins". Prp19 is an essential splicing factor which together with other proteins forms a non-snRNP complex called NTC (Nineteen Complex). The NTC consists of Prp19p, Ntc85p/Cef1p, Ntc25p/Snt309p, Ntc31p/Syf2p, Ntc30p/Isy1p, Ntc20p, Ntc40p/Cwc2p, Ntc50p/Prp46p, Ntc77p/Clf1p/Syf3p and Ntc90p/Syf1p (Chen et al., 1998; Chen et al., 1999; Tsai et al., 1999; Chen et al., 2001a; Chen et al., 2002). The NTC seems to be required for the proper interaction of the U6 snRNA with the 5´ss. Moreover, after the NTC recruitment, a heptamer ring of Lsm proteins is released from the U6 snRNA 3´ end (Chan et al., 2003; Chan and Cheng, 2005). Additional functions of the NTC subunits are very likely.

Previously, the NTC complex was considered associated with spliceosome concomitantly or just after the release of U1 and U4 snRNPs (Tarn et al., 1994). However, all NTC components were recently co-purified with the complex B, which contains all snRNPs, including the U1 and U4 (Fabrizio et al., 2009). Prp19 related (or NTC related) proteins (besides Prp45, the group contains Prp46, Cwc2, Ecm2, Cwc15 and Bud31) are less stably associated with the NTC but are expected to enter the spliceosome at the same time as NTC (Ohi and Gould, 2002; Chen et al., 2002).

1.12 Pre-mRNA secondary structures in splicing

The ability of pre-mRNA to fold into intramolecular secondary structures was shown to impact both constitutive and alternative splicing in various species (for review see (Buratti and Baralle, 2004; Warf and Berglund, 2010). In general, secondary structures can mask sequences which are recognized by single strand RNA-binding proteins or snRNAs, form novel binding epitopes, or bring distant splicing signals into proximity, thus allowing their efficient recognition.

Sequestering splicing *cis* signals into the secondary structures is the most straightforward way to inhibit splicing by pre-mRNA folding. In the human *SMN2* gene, a secondary structure involving the 5´ss of exon 7 hinders the interaction with the U1 snRNA and leads to the exon exclusion (Singh et al., 2007). An extensive secondary structure encompassing both 5´ss and 3´ss of an intron in the duck hepatitis B virus is responsible for its inefficient splicing (Loeb et al., 2002). A stem-loop artificially placed in the close vicinity of the BP in the *S. cerevisiae ACT1* gene abolished its splicing *in vitro* (Halfter and Gallwitz, 1988). Short artificial hairpins inserted to the splicing signals of *RP51A* gene inhibited its splicing both in vivo and in vitro (Goguel et al., 1993).

On the contrary, secondary structures can also aid to splicing by forming binding platforms or displaying weak splicing signals or regulatory elements. A crucial exonic splicing enhancer in the human fibronectin EDA exon is accessible for regulatory SR proteins only when a secondary structure is formed (Muro et al., 1999; Buratti et al., 2004). Alternatively, positive splicing regulation can be achieved by masking of the splicing inhibitory motifs, e.g. splicing silencers.

Many studies reported that pre-mRNA folding enables the interaction of distant splicing signals or regulatory motifs. In the *S. cerevisiae RP51B* gene, the base-

pairing of sequences downstream of the 5´ss decreases a "structural distance" between the 5´ss and BP and facilitates its efficient splicing (Libri et al., 1995; Charpentier and Rosbash, 1996; Rogic et al., 2008). Long-range intramolecular pre-mRNA interactions are involved in the regulation of alternative splicing. In *Drosophila*, mutually exclusive exons of the *Dscam* gene were shown to be selected based on the competitive interaction of a single "docking site" downstream of a constitutive exon with one of a multiple "selector sites" located upstream of each exon variant (Fig. 1.11; Graveley, 2005). A similar mechanism was proposed for the cluster of mutually exclusive exons in 14-3-3ξ pre-mRNA and other genes. In addition, the secondary structures involved in alternative splicing regulation of 14-3-3ξ seem to be evolutionary conserved, suggesting that the mechanism may be broadly applicable in mutually exclusive exon selection (Yang et al., 2011).

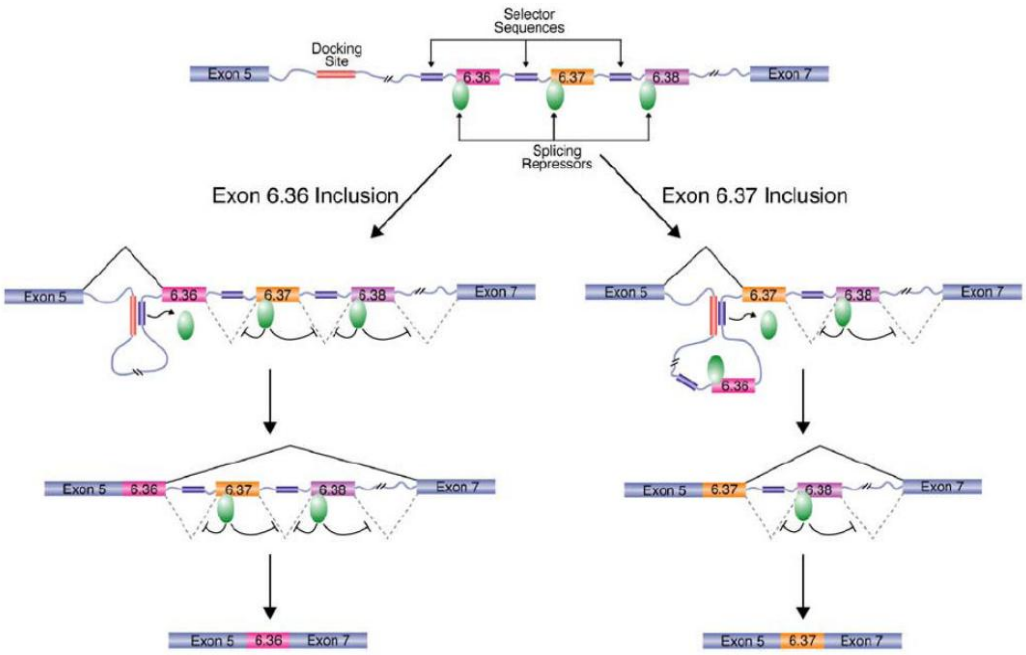


Figure 1.11. Model of pre-mRNA secondary structure based selection of mutually exclusive exons *Drosophila Dscam* gene. Single "docking site" base pairs with one of multiple "selector sequences" located upstream of each of mutually exclusive exons. Formation of this structure prevents from binding of splicing repressor to the involved exon, while others exons are occupied and cannot be recognized by splicing machinery. Adopted from Graveley, 2005.

1.13 Recognition of 3´ splice site

The mechanism of 3´ss recognition is only poorly understood at present. In contrast to the 5´ss and BP sequences, the information content of the 3´ss YAG sequence seems to be insufficient to define the 3´ ends of introns in budding yeast (Lim and

Burge, 2001; Iwata and Gotoh, 2011; Fig. 1.1). The 3´ss is considered to be firstly recognized during the spliceosome assembly prior to the first catalytic step (see section 1.4 Formation of prespliceosome). However, the interaction of BBP/Mud2 (in *S. cerevisiae*) or SF1/U2AF65/U2AF35 (in vertebrates) complexes with intron relies predominantly on the BP or BP/pY-tract sequence, respectively (Wahl et al., 2009). As the 3´ss is not separated from the BP in a vast majority of yeast and human introns by more than 50 nucleotides, the BP/pY-tract/3´ss region is presumably recognized as a single unit. In budding yeast, the lariat formation can even occur on the RNA molecules without any AG dinucleotide downstream of the BP (Rymond et al., 1987; Cheng, 1994). In contrast, the first step occurs in an AG-dependent manner in the typical mammalian intron (although the presence of a strong pY-tract overcomes the requirement of AG). However, the AG encoded signal is, at this stage, required rather only for the assembly of splicing complexes; a definitive 3´ss positioning is achieved only before the second step, as shown by experiments employing the second step AG-containing substrates *in trans* (Anderson and Moore, 2000). Second step specific factors may thus play a role in the determination of the proper acceptor site. *In vitro*, Prp22, Slu7, Prp18, and Prp17, proteins engaged in the second transesterification in *S. cerevisiae*, are required only for the splicing of introns with certain minimal distance between the BP and 3´ss (Brys and Schwer, 1996; Zhang and Schwer, 1997; Schwer and Gross, 1998; Sapra et al., 2004). However, the critical distance is shorter than the BP to 3´ss spacing in the vast majority of annotated yeast introns. Additionally, Slu7 was shown to impact alternative 3´ss choice on artificial substrates (Frank and Guthrie, 1992; Eva Ničová, unpublished results). How exactly the second step splicing factors contribute to the 3´ss recognition remains to be elucidated.

1.14 Scanning model of 3´ss identification

More than 20 years ago, a so-called "scanning model" was proposed for 3´ss recognition in mammals (Smith et al., 1989). According the model, the sequence downstream of BP/pY-tract undergoes linear scanning and the first AG dinucleotide is employed as a 3´ss. In subsequent research, the same group extended and specified their model. They defined the situations in which the BP-proximal AG can be skipped by the scanning machinery. It could be positioned either too close to BP, or

also sequestered by RNA hairpin. In addition, in the case of two closely spaced AG dinucleotides, both can be employed as acceptor sites, suggesting that the mechanism reads through the first AG dinucleotide available and explores also the certain portion of downstream sequence for AG content ("leaky scanning"). In the competition between two close AG dinucleotides, the nucleotide preceding AG is decisive for the strength of the 3´ss in following order: UAG>CAG>AAG>GAG (Smith et al., 1993). In *Saccharomyces cerevisiae*, more than 90 % of introns have UAG or CAG 3´ss and the rest have AAG. Therefore, the 3´ss strength can probably applicable to broad range of organisms.

However, there is substantial inconsistency with the scanning model in *S. cerevisiae*. A relatively large proportion of yeast introns contain the BP-proximal AGs, which are not recognized as 3´ss (Saccharomyces Genome Database). The frequency of AG occurrence upstream of the physiological 3´ss in vertebrate introns is not easy to determine, as the degeneration of BP sequence precludes its reliable identification by computational algorithms. As a consequence, the sequence between BP and 3´ss cannot be explored for AG content. There is also another fundamental problem with the scanning model – its potential molecular mechanism is completely unknown.

In *S. cerevisiae*, the splicing efficiency was shown to decrease with the increasing distance between BP and 3´ss. An artificial extension of the BP to 3´ss distance in *ACT1* intron to > 60 nt completely abolished its splicing (Cellini et al., 1986). However, there are several introns in budding yeast which have the BP to 3´ss distance longer than 100 nt (Spingola et al., 1999; SGD; for distribution of BP to 3´ss distances in *S. cerevisiae* see Fig. 3.25B in Results). Apparently, the spliceosome has to employ an additional mechanism that facilitates 3´ss recognition in these nonconventional introns.

2. MATERIAL AND METHODS

2.1 Microorganisms and cultivations

2.1.1 List of strains

Escherichia coli

Name	Genotype	Source
DH5α	*ϕ80dlacZΔM15 recA gyrA96 thi-1 hsdR17 (r_K^-, m_K^+) supE44 relA1 deoR Δ(lacZYA-argF)U 169*	Stratagene, USA
TOP10	F'[*lacI*^{q} Tn10(tet^R)] *mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZ*ΔM15 *ΔlacX74 deoR nupG recA1 araD139 Δ(ara-leu)7697 galU galK rpsL(Str^R) endA1 λ$^-$*	Invitrogen, USA
XL1 blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′*proAB lacI*qZΔ*M15* Tn*10* (Tetr)]	Stratagene, USA

Saccharomyces cerevisiae

Name	Genotype or description	Source or reference
EGY48	*MATα his3 trp1 ura3 LexA$_{op(x6)}$ – LEU2*	Golemis et al., 1996
KAY02	*prp45(1-169)::kanMX4* knock-in into EGY48	Gahura et al., 2009
BY4741	*MATa his3-Δ1 leu2-Δ met15-Δ ura3-Δ*	Euroscarf
AVY17	*prp45(1-169)::kanMX4* knock-in into BY4741	A. Valentová
BY4999	*MATa dbr1-Δ::KanMX his3-Δ1 leu2-Δ met15-Δ ura3-Δ*	Euroscarf
FPY2D	*prp45(1-169)::kanMX4* knock-in into BY4999	Gahura et al., 2009
46ΔCUP	*MATa ade2 cup1Δ::ura3 his3 leu2 lys2 trp1 ura3, GAL+*	M. Konarska
KGY1522	*MATa cef1-13 ade2-101 leu2-Δ1 lys2-801 trp1-Δ1 ura3-52*	K. Gould
KGY1825	*MATa, prp3-1 ade1 ade2 gal1 his7 lys2 tyr1 ura1*	K. Gould
KGY2818	*MATa prp17/cdc40-Δ his3-Δ1 leu1Δ0 met15Δ ura3Δ0*	K. Gould
KGY2847	*MATa prp22-1 ade2 his3 ura3*	K. Gould
KGY823	*MATa ade2-101 leu2-Δ1 lys2-801 trp1-Δ1 ura3-52*	K. Gould
TSY01	*prp45(1-169)::kanMX4* knock-in into KGY823	T. Šimonová
GRF18	*MATα his3-11 his3-15 leu2-3 leu2-112 canR*	B. Janderová
BJ2168	*MATa leu2 trp1 ura3-52 prb1-1122 pep4-3 prc1-407 gal2*	lab stock
KAY03	*prp45(1-169)::kanMX4* knock-in into BJ2168	K. Abrhámová
#94500	*MAT*a/a *his3- Δ1/his3- Δ1 leu2- Δ0/leu2- Δ0 ura3-Δ0/ura3-Δ0 lys2- Δ0/LYS2 MET15/met15- Δ0 prp45::kanMX4/PRP45*	Invitrogen

2.1.2 Cultivations and manipulations with microorganisms

Escherichia coli

Escherichia coli were cultivated in standard liquid LB media (10 g/l peptone, 5 g/l yeast extract, 5 g/l NaCl) or in Super LB media (20 g/l peptone, 10 g/l yeast extract, 5 g/l NaCl). Nutrient Agar plates were prepared from Živný agar č. 2, Imuna, Slovakia (4 g/l). Ampicillin (100 μg/ml) was added if required.

Transformations were done by electroporation (Gene Pulser Apparatus, Bio-Rad; settings 25 μF, 2.5 kV, 200 Ω) or heat shock (competent cells were incubated with

plasmid on ice for 30 min, shifted to 42°C for 30 s and diluted in LB with 0.5 % glucose). After transformation, cells were pre-cultivated in LB with 0.5 % glucose for 1 hour and spread on agar plates with ampicillin.

Saccharomyces cerevisiae

S. cerevisiae were grown in standard complex YPAD media (20 g/l peptone, 10 g/l yeast extract, 2 % glucose, 100 µg/l adenine) or in minimal SD media (6.7 g/l yeast nitrogen base w/o amino acids supplemented with appropriate Drop-out, Formedia, UK, amino acids, and 2 % glucose).

For agar plates, Agar No. 1 (Oxoid) was added to final concentration 1.5-2 %.

For copper sensitivity assay, YPAD or SD media were supplemented by given amount of $CuSO_4$.

For negative selection of *URA3* expressing cells, media containing 5-fluroorotic acid (5-FOA) were prepared by adding filter-sterilized 5-FOA (Sigma, USA) solution to SD media to the final concentration 1 mg/ml.

Transformation of *Saccharomyces cerevisiae*

S. cerevisiae were transformed by lithium acetate technique (Gietz and Woods, 2002). Briefly, 2-5x10^6 of competent cells were mixed with 100 ng – 1 µg of plasmid DNA, 240 µl of 50 % polyethylenglykol (PEG 3350, Sigma, USA), 36 µl of 1 M LiAc, and 10 µl of denatured ssDNA (10 mg/ml). Mixture was heat shocked in 42°C for 30 – 120 minutes. Cells were spread on appropriate selection media.

Sporulation and tetrad analysis of *Saccharomyces cerevisiae*

Freshly grown single colony of diploid strain was spread on the presporulation agar plate (20 g/l peptone, 10 g/l yeast extract, 10 % glucose, 20 g/l agar) and cultivated overnight at 28°C. Biomass was harvested and resuspended in 50 µl of Ringer solution (7.5 g/l, NaCl, 0.075 g/l KCl, 0.1 g/l $NaHCO_3$, 0.1 g/l $CaCl_2$). The suspension was dropped on the sporulation agar plates (5 g/1 l KAc, 2.3 g/1 KCl, 20 g/1 agar) and incubated at 25°C for 4 to 6 days. The progress of sporulation was checked under microscope. Sporulated cells were diluted in 100 µl of Yeast Lytic Enzyme solution (1 mg/ml; MP Bio, USA) and incubated for 5 to 10 minutes in 25°C. Tetrades were dissected using Singer micromanipulator instrument and cultivated on YPAD media.

2.2 Nucleic acids procedures

2.2.1 List of enzymes

Restriction enzymes (all restriction enzymes were from MBI Fermentas, Lithuania):

*Bam*HI, *Eco*RI, *Not*I, *Sal*I, *Xba*I, *Xho*I

DNA modifying enzymes:

T4 DNA Ligase, MBI Fermentas, Lithuania

Shrimp Alkaline Phosphatase (SAP), MBI Fermentas, Lithuania

T4 polynucleotide kinase, Fermentas, Lithuania

PCR, RT-PCR, qPCR:

Taq polymerase, MBI Fermentas, Lithuania

High Fidelity Enzyme Mix, MBI Fermentas, Lithuania

RevertAid™ M-MuLV Reverse Transcriptase, MBI Fermentas, Lithuania

MESA GREEN qPCR Master Mix Plus for SYBR® Assay No Rox, Eurogentec, Belgium

2.2.2 Kits for nucleic acids purification

Plasmid DNA from *E. coli* was isolated by *Nucleo Spin Plasmid*, Macherey-Nagel, Germany.

DNA fragments from agarose gels were isolated by *Nucleo Spin Extract*, Macherey-Nagel, Germany.

Total RNA from yeast was isolated by *MasterPure™ Yeast RNA Purification Kit*, Epicentre Biotechnologies, USA.

2.2.3 Other kits used in this study

Cloning of PCR amplified DNA molecules was performed by *TOPO TA Cloning®* kit, Invitrogen, USA.

Reverse transcription was performed by *RevertAid™ M-MuLV First Strand cDNA Synthesis Kit*, MBI Fermentas, Lithuania.

5´cDNA ends were identified by *5´ RACE System for Rapid Amplification of cDNA Ends, Version 2.0*, Invitrogen, USA.

Site-directed mutagenesis was performed by *QuikChange® II Site-Directed Mutagenesis Kit*, Stratagene, USA.

2.2.4 DNA molecular weight ladders

λ DNA/PstI, lab stock

GeneRuler™ DNA Ladder Mix, MBI Fermentas, Lithuania

GeneRuler™ 1 kb DNA Ladder, MBI Fermentas, Lithuania

GeneRuler™ 50 bp DNA Ladder, MBI Fermentas, Lithuania

2.2.5 DNA isolation from *S. cerevisiae*

Plasmid or genomic DNA was isolated from *S. cerevisiae* using mechanical disruption by glass beads and phenol:chloroform extraction. 5 to 20 ml of freshly grown cells from stationary phase were harvested, mixed with 0.2 ml of yeast lysis solution (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris, pH 8.0, 1 mM EDTA), 0.2 ml of phenol:chloroform:isoamylalcohol (25:24:1, pH 6.6), and ca 300 mg of glass beads (425-600 μm), and disintegrated by 2 to 4 min vortexing. After centrifugation (16000 g, 5 min, 20°C), the water phase was transferred to the clean microtube, and DNA was ethanol precipitated. Diluted DNA can be treated by 1 μl RNase A (10 mg/ml) to remove contaminant RNA.

Recovered genomic DNA is suitable for PCR amplification. Plasmid DNA can be used for *E. coli* transformation or as a template for PCR, but is not convenient for restriction analysis.

2.2.6 RNA isolation from *S. cerevisiae*

RNA was isolated from *Saccharomyces cerevisiae* using *MasterPure™ Yeast RNA Purification Kit* (Epicentre Biotechnologies, USA). Quality of purification and concentration was determined by absorption spectroscopy (Nanodrop 2000, Thermo Scientific, USA). When required, integrity of RNA was checked by native agarose electrophoresis. Alternatively, RNA was purified by hot acid phenol:chloroform extraction.

2.2.7 PCR

Polymerase chain reactions were run in *Peltier PTC-200* thermal cycler (MJ Research). Enzymes used are listed in section 2.2.1 List of enzymes. Cycling

parameters were adjusted according to the specific application. Primer concentration was typically set to 1 μM. Mg^{2+} concentration was 1.5 mM for PCR from plasmid DNA templates. For colony PCR and genomic DNA templates, 2.5 - 3 mM $MgCl_2$ was used.

2.2.8 qRT-PCR

Reverse transcription was performed by *RevertAid™ M-MuLV Reverse Transcriptase* (MBI Fermentas, Lithuania) on 2 μg of total RNA from random hexanucleotide primers. qPCR was performed in *Light Cycler LC480* (Roche, USA) using *MESA GREEN qPCR Master Mix Plus for SYBR® Assay No Rox* (Eurogentec, Belgium). 0.05 μl of cDNA was used per 12.5 μl volume of single qPCR reaction. Primer concentration was set to 0.3 μM. Primers are listed in section 2.2.12 List of oligonucleotides. Relative quantity of specific cDNA was calculated using the relative standard curve method (Guide to Performing Relative Quantification of Gene Expression Using Real-Time Quantitative PCR, Applied Biosystems).

2.2.9 Primer extension analysis

Primer extension was performed by *RevertAid™ M-MuLV Reverse Transcriptase* (MBI Fermentas, Lithuania) on 3 to 10 μg of total RNA isolated from cells expressing desired reporter construct. Primers were radioactively labeled on 5´ ends using *T4 polynucleotide kinase* (Fermentas, Lithuania) and γ^{32}P-ATP, 4500 Ci/mmol (MP Bio, USA). Products were separated on denaturing polyacrylamide gel electrophoresis (20 cm gel; 8 % acrylamide 19:1, 7 M urea; 1x TBE buffer: 89 mM Tris base, 89 mM boric acid, 1.9 mM EDTA; 300 – 400 V) and visualized by phosphorimaging (exposure screen *Kodak K*; *Molecular Imager PharosFX*, Bio-Rad, USA). Densitometry analysis was performed by *Image Quant™ TL* software (GE Life Sciences).

2.2.10 DNA agarose gel electrophoresis

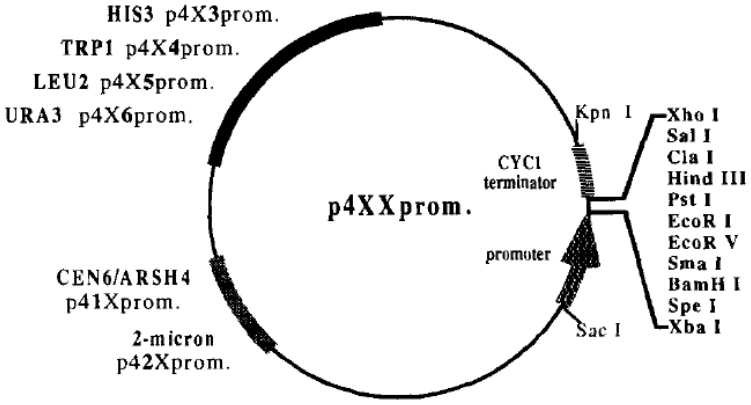
DNA agarose electrophoresis was performed using 1 to 2.5 % agarose (SeaKem LE Agarose, Cambrex, USA) in 1x TAE buffer (40mM Tris, 20mM acetic acid, 1mM EDTA). For fragments up to 150 nt, 4 % mix of SeaKem LE Agarose and NuSieve GTG Agarose (FMC BioProducts, USA) was used in ratio 3:1. Gels were run by

voltage of 3 – 10 V/cm, stained by ethidium bromide (ca 0.5 μg/ml), and photodocumented upon UV illumination.

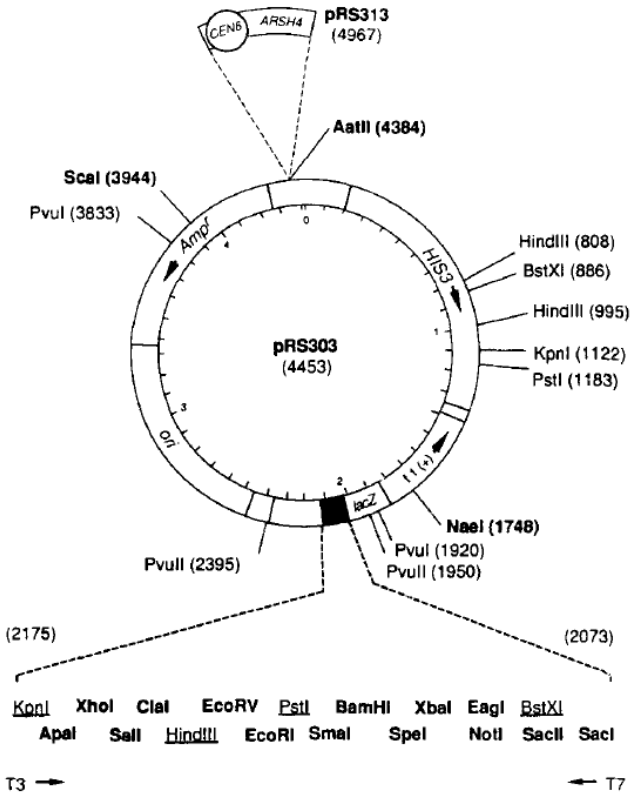
2.2.11 Plasmids

Vectors:

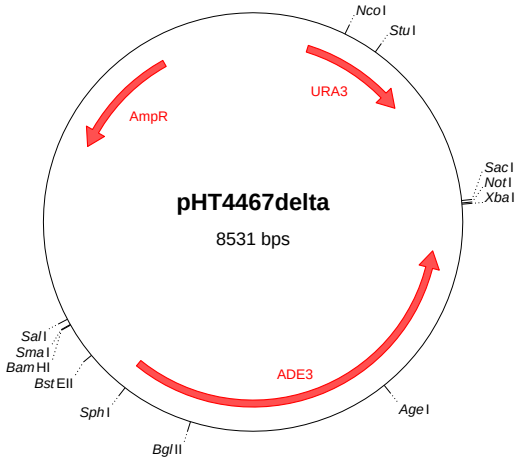
p416ADH, p413ADH, p423GPD – yeast expression vectors containing *ADH* or *GPD* promoter, *CYC1* terminator, and different selection markers (Mumberg et al., 1995).



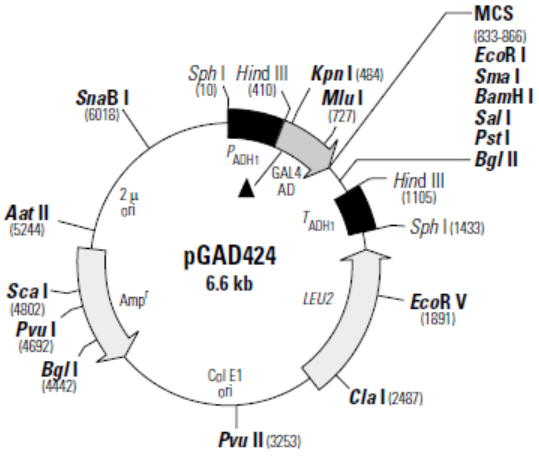
pRS313, pRS414 – yeast centromeric vectors (Sikorski and Hieter, 1989).

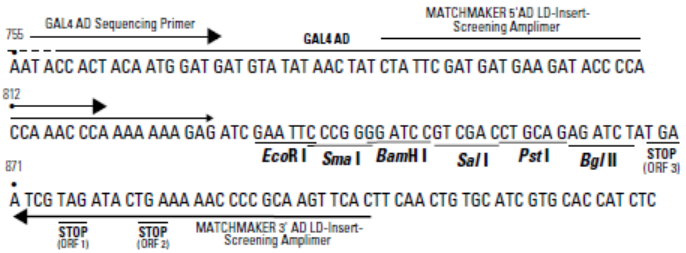


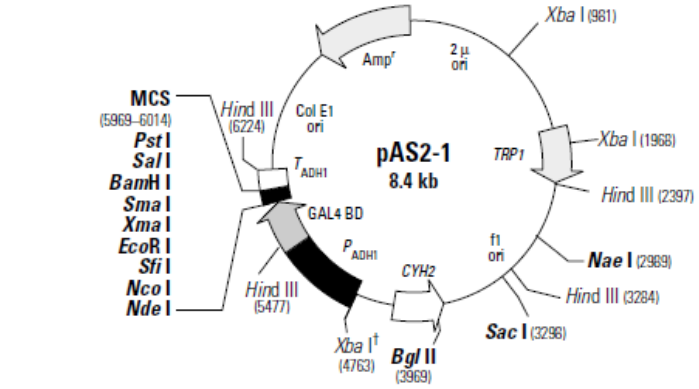
pHT4467Δ – yeast vector with disrupted centromeric sequence resulting in quick loss from cells under non-selective conditions. Convenient for the synthetic lethal screen.

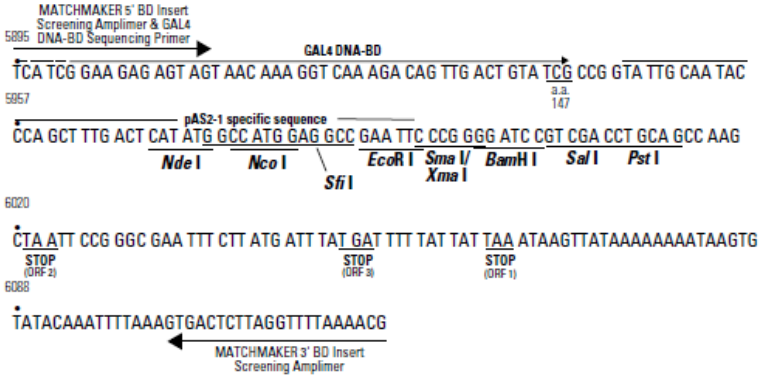


pGAD424, pAS2.1 – vectors for yeast two-hybrid screen. pGAD424 contains GAL4 activation domain (AD), pAS2-1 contains GAL4 DNA-binding domain (BD); MATCHMAKER Two-Hybrid System, Clonetech, USA.

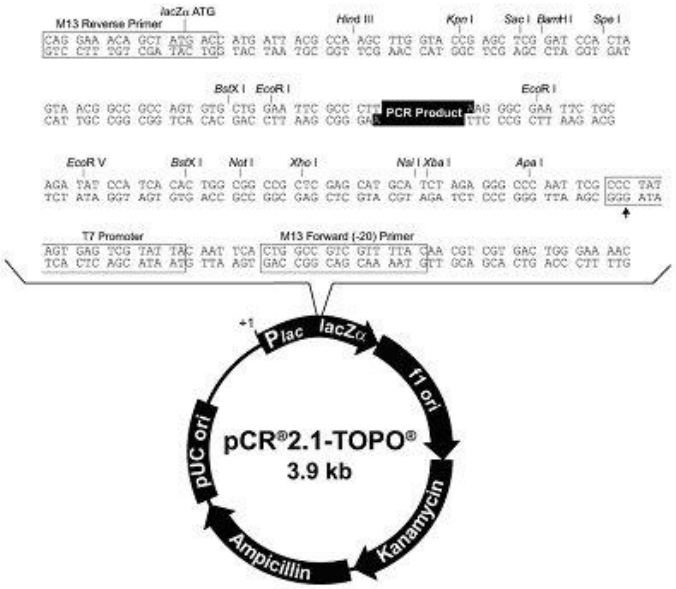








pCR2.1®-TOPO – cloning vector from *TOPO TA Cloning®* kit (Invitrogen, USA)



Plasmids generated in this study

Name	Vector	Description	Cloning technique
pOG02	pHT4467Δ	*PRP45* with own promoter	*Xba*I/*Not*I
pOG08	pRS313	*SLU7* with own promoter	*Bam*HI/*Sal*I
pOG09	pRS313	*SYF3* with own promoter	*Sal*I/*Xba*I
pOG12	p413ADH	*PRP22* overexpression	*Bam*HI/*Sal*I
pOG30	pGAD424	*GAL4* AD-*prp45*(1-169) 2H fusion	*Eco*RI/*Sal*I
pOG32	pGAD424	*GAL4* AD-*prp45*(258-379) 2H fusion	*Eco*RI/*Sal*I
pOG33	pAS2.1	*GAL4* BD-*PRP22* 2H fusion	*Bam*HI/*Sal*I
pOG46	p423GPD	*cof1*(G149A)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG52	p423GPD	*COF1-CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG58	p423GPD	*ACT1-CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG63	p423GPD	*TUB3-CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG64	p423GPD	*PFY1-CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG65	p423GPD	*RPL30-CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG66	p423GPD	*CPT1-CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG67	p413ADH	intronless SNW1 expression in *S. cerevisiae*	*Eco*RI/*Sal*I
pOG68	p413ADH	SNW/SKIP expression in *S. cerevisiae*	*Eco*RI/*Xho*I
pOG69	p413ADH	AtSKIP expression in *S. cerevisiae*	*Bam*HI/*Xho*I
pOG71	p423GPD	*cof1*(C80T)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG72	p423GPD	*cof1*(C80T+G149A)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG73	p423GPD	*cof1*(A148G)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG74	p423GPD	*cof1*(A148T)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG76	p423GPD	*cof1*(A148C)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG80	p423GPD	*cof1*(Δ76-152)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG81	p423GPD	*cof1*(Δ91-140)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG82	p423GPD	*cof1*(Δ91-140+G149A)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG99	p423GPD	*cof1*(Δ91-140+G149A+C80T)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG100	p423GPD	*cof1*(Δ76-176)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG101	p423GPD	*cof1*(hel)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG102	p423GPD	*UBC13-CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG105	p423GPD	*ubc13*(G232A)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG106	p423GPD	*ubc13*(G232A+C148T)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG112	p423GPD	*ubc13*(reordered)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG113	p423GPD	*ubc13*(disordered)-*CUP1* splicing reporter	*Bam*HI/*Sal*I
pOG129	p423GPD	*COF1-HFM1*(100) splicing reporter	*Eco*RI/*Sal*I
pOG130	p423GPD	*cof1*(G149A)-*HFM1*(100) splicing reporter	*Eco*RI/*Sal*I
pOG131	p423GPD	*COF1-HFM1*(200) splicing reporter	*Eco*RI/*Sal*I
pOG132	p423GPD	*cof1*(G149A)-*HFM1*(200) splicing reporter	*Eco*RI/*Sal*I
pOG133	p423GPD	*COF1-HFM1*(400) splicing reporter	*Eco*RI/*Sal*I
pOG134	p423GPD	*cof1*(G149A)-*HFM1*(400) splicing reporter	*Eco*RI/*Sal*I
pOG135	p423GPD	*COF1-HFM1*(600) splicing reporter	*Eco*RI/*Sal*I
pOG136	p423GPD	*cof1*(G149A)-*HFM1*(600) splicing reporter	*Eco*RI/*Sal*I
pOG137	p423GPD	*COF1-HFM1*(1000) splicing reporter	*Eco*RI/*Sal*I
pOG138	p423GPD	*cof1*(G149A)-*HFM1*(1000) splicing reporter	*Eco*RI/*Sal*I
pOG139	p423GPD	*COF1-HFM1*(2000) splicing reporter	*Eco*RI/*Sal*I
pOG140	p423GPD	*cof1*(G149A)-*HFM1*(2000) splicing reporter	*Eco*RI/*Sal*I

Other plasmids used in this study

Name	Vector	Description	Source
p416ADH-*PRP45*	p416ADH	expression of *PRP45*	lab stock
p416ADH-*prp45* (119-379)	p416ADH	expression of *prp45*(119-379)	lab stock
pSE258-*PRP22*	pSE258	*PRP22* with own promoter	B. Schwer
pGAD424-*PRP45*	pGAD424	GAL AD-*PRP45* 2H fusion	lab stock
pRS313-*slu7*(P268L)	pRS313	*slu7*(P268L) with own promoter	lab stock
pRS414-*CEF1*	pRS414	*CEF1* with own promoter	lab stock

Name	Description	Source
pMM1C	*ACT1-CUP1* "recombinant" splicing reporter	M. Konarska
pMM4C	*act1-CUP1* 3´ss gAG "recombinant" splicing reporter	M. Konarska
pMA2C	*act1-CUP1* 5´ss A3C "actin" splicing reporter	M. Konarska
pMA42	*act1-CUP1* BP-C "actin" splicing reporter	M. Konarska
pCQ04	*act1-CUP1* BP-G "actin" splicing reporter	M. Konarska
pC256A	*act1-CUP1* BP-C256A "actin" splicing reporter	M. Konarska
pRH20	*act1-CUP1* 5´ss A3C "recombinant" splicing reporter	M. Konarska
pJP40	*act1-CUP1* 3´ss aAG "actin" splicing reporter	M. Konarska
pJP56	*act1-CUP1* 3´ss cAG "actin" splicing reporter	M. Konarska

Construction of CUP1 fusion reporters

The *CUP1* based splicing reporters were constructed as follows: the fragment of intron containing gene of interest (comprising exon 1 including several nucleotides upstream of ATG, intron, and first 10 to 30 nucleotides of exon 2) and the complete coding sequence of *CUP1* gene were PCR amplified from genomic DNA isolated from BY4741 strain (for primers see section 2.2.12 List of oligonucleotides) and TOPO-TA cloned into pCR®2.1-TOPO® vector (Invitrogen). The *Bam*HI/*Eco*RI fragment of intron containing gene and the *Eco*RI/*Sal*I fragment of *CUP1* were inserted into p423GPD vector, resulting in the *CUP1* fusion gene expressed from *TDH3* promoter.

The *COF1-CUP1* reporters with single-nucleotide substitutions were generated in p423GPD vector by site-directed PCR-based mutagenesis using *QuikChange® II Site-Directed Mutagenesis Kit* (Stratagene). DNA fragments encoding *COF1-CUP1* reporters containing internal intron deletions and all *UBC13-CUP1* reporters were synthesized commercially by GeneArt (Germany) and inserted into *Bam*HI/*Sal*I-digested p423GPD.

Sequences of all genes' fragments inserted in front of the *CUP1* coding sequence and all *cof1-HFM1* sequences are summarized in Appendix A2.

2.2.12 List of oligonucleotides

Name	Sequence	Description
OG01	GAGATCTGATATGTCTGATATATCGAAAC	*PRP22* ORF amplification F
OG04	GGTCGACTTACCTCTTGATACC	*PRP22* ORF amplification R
OG13	GAGATCTACCATATGTCTGATATATCGAAAC	*PRP22* ORF amplification F, *Nde*I site added
OG42	GTCGACTACACAGCTGCAGGTATCTTC	*prp45*(1-169) amplification R
OG44	GGATCCTAACAAAAGAAGATGTCTAGATC	*COF1* intron containing fragment amplification F
OG45	GGAAGTTAATTAATTCGCTGAATTCAGCAACA GCAACACTACAC	*COF1* intron containing fragment amplification R
OG46	GTCGACTATTCGTTTCATTTCCCAG	*CUP1* amplification F; GSP1* for 5´ RACE
OG47	GTGTAGTGTTGCTGTTGCTGAATTCAGCGAAT TAATTAACTTCC	*CUP1* amplification R
OG51	GGATCCTTTTAGATTTTTCACGC	*ACT1* intron containing fragment amplification F
OG52	GAATTCGGCTTTACACATACCAGAAC	*ACT1* intron containing fragment amplification R
OG53	GGATCCAAGCGACTTGAGAC	*TUB3* intron containing fragment amplification F
OG54	GAATTCCCAGCATGCATTAC	*TUB3* intron containing fragment amplification R
OG55	GGATCCAACTACGATCGC	*PFY1* intron containing fragment amplification F
OG56	GAATTCGGTTCCTATTAAGTTATC	*PFY1* intron containing fragment amplification R
OG57	GGATCCTACAGCTTATTCAATTAATC	*RPL30* intron containing fragment amplification F
OG58	GAATTCTTGGGATTTAACTGG	*RPL30* intron containing fragment amplification R
OG59	GGATCCATGGATTCGAAAAC	*CPT1* intron containing fragment amplification F
OG60	GAATTCACTTTGATACCTAAAGTC	*CPT1* intron containing fragment amplification R
OG71	CGACGTTGTCGATGATGAAG	qPCR *TOM22* F
OG72	GCAACGATTCTGTCCAAC	qPCR *TOM22* R
OG75	CTAACAAAAGAAGATGTCTAGATCTGGTG	qPCR *COF1* spliced mRNA F
OG76	GGTTTTAGCATCGTTCAATCCG	qPCR *COF1* spliced mRNA R
OG141	ACAAAAGAAGATGTCTAGATCTGGGTA	qPCR *COF1* pre-mRNA F
OG142	GTTCCATGTTAAAAAGGGAGGAAAGA	qPCR *COF1* pre-mRNA R
OG172	ATGCTGTTTTATCTCGCTCTTTTATGTCCTTTC TTTCCTCC	*cof1* C80T mutagenesis F
OG173	GGAGGAAAGAAAGGACATAAAAGAGCGAGA TAAAACAGCAT	*cof1* C80T mutagenesis R
OG174	AAGGAATGTGGGGGAAAAAATTCATABGAAG CCAGGAAAATG	*cof1* A148B mutagenesis F
OG175	CATTTTCCTGGCTTCVTATGAATTTTTTCCCCC ACATTCCTT	*cof1* A148B mutagenesis R
OG197	AGCAGCATGACTTCTTGGTTTCTT	GSP2* for 5´ RACE
OG200	GAATTCCAGTCTGCTACATTTAAAAG	*HFM1* exon 2 fragment amplification F
OG202	GTCGACTCAAAATGGATATTTCAAAAGTG	*HFM1* exon 2 fragment 2000 nt amplification R
OG203	GTCGACTCAGTCATTTTTGCAATTG	*HFM1* exon 2 fragment 1000 nt amplification R
OG204	GTCGACTCAGTTCTTGTACATCTC	*HFM1* exon 2 fragment 600 nt amplification R
OG205	GTCGACTCAGTTCTCGTTACTC	*HFM1* exon 2 fragment 400 nt amplification R
OG206	GTCGACTCATTTACTAGTTACTTGTG	*HFM1* exon 2 fragment 200 nt amplification R
OG207	GTCGACTCAATCATCCAGTAAACTG	*HFM1* exon 2 fragment 100 nt amplification R
OG208	GTTGGCGGCTATTTTTC	Primer extension *HFM1* constructs
FP20B	GCTCTAGAAGCAAATTGAGATAACTCC	*PRP45* locus amplification F
FP21	GCGGCCGCATCACAGAGCATC	*PRP45* locus amplification R
KM03	GCCGAATTCATGGGAGCAATAGTTAAGC	*prp45*(258-379) amplification F
KM04	GGACTCGAGCTAGGCGCCATAGTTATC	*prp45*(258-379) amplification R
PF20	GCTGAATTCATGTTTAGTAACAGACTACC	*prp45*(1-169) amplification F
U1OG	CGCCGTATGTGTGTGTGACC	Primer extension U1 snRNA
YAC06	GGCACTCATGACCTTC	Primer extension *CUP1* constructs
YU14	ACGATGGGTTCGTAAGCGTACTCCTACCGTGG	Primer extension U14 snoRNA

*GSP – Gene Specific Primer for 5´ RACE

2.3 Protein procedures

2.3.1 Denatured protein extracts from *S. cerevisiae*

Mid-log growing cells were harvested, washed in 1 ml of ddH_2O, and centrifuged. Sediment was incubated for 15 minutes at -20°C. Cells were intensively vortexed (2 minutes) with 100 µl of freshly prepared 1.85 M NaOH with 7% β-merkaptoethanol, and incubated for 10 minutes on ice. After addition of 100 µl of 50% TCA and 5 minutes incubation on ice, lysate was centrifuged (10 minutes, 20000 g, 0°C). Supernantant was quantitatively removed and pellet was resuspended in 500 µl of 1 M Tris and again centrifuged (1 minute, 20000 g, 0°C). Sediment was resuspended in 100 µl of hot 2 × Laemmli buffer including freshly added 1/20 volume of DTT and incubated for 20 minutes at 80°C. Protein concentration in samples was determined using *DC Protein Assay* (Bio-Rad, USA). Samples were stored at -70°C.

2 × Laemmli buffer: 0.112 M Tris-HCl, pH 6.8
3.42 % SDS
12% (v/v) glycerol,
0.002% bromphenol blue

2.3.2 Denaturing proteins electrophoresis

Proteins were separated using SDS polyacrylamide gel electrophoresis in the standard Tris-glycine buffer system. *Mini Protean III* apparatus (Bio-Rad, USA) and 7.5 – 10 % gels were used. Gels were stained by Coomassie Brilliant Blue G-250 (Bio-Rad) according the manufacturer´s instructions.

2.3.3 Western blotting and immunodetection

Proteins were transferred to nitrocellulose membrane using Mini Trans-Blot (Bio-Rad) at 90 – 110 V for 1 – 2 hours. Membranes were blocked by 3 % milk and were incubated with the solution of primary Anti-Prp22 antibody (dilution 1:1000; gift from B. Schwer) for 3 hours at 20°C. Primary antibody was washed and membranes were incubated with the solution of secondary antibody conjugated with alkaline phosphatase (Goat anti-Rabbit AP, Bio-Rad; dilution 1:3000) for 1 hour. After washing, signal was detected by alkaline phosphatase reaction (BCIP and NBT substrates; Bio-Rad).

2.4. *In vivo* assays

2.4.1 Copper sensitivity assay

cup1-Δ cells expressing *CUP1* fusion reporter gene from plasmid were cultivated in appropriate minimal SD media to mid-log phase, harvested, concentrated to OD 0.4, and spotted in the eightfold serial dilutions on set of SD plates containing different amounts of $CuSO_4$ (concentration was set according the requirements of particular experiment usually up to 1 mM $CuSO_4$).

2.4.2 Two hybrid interactions assay

Two hybrid interactions were tested using **β-galactosidase colony lift filter assay**.

Protocol:

- Nylone membrane was placed at the appropriate SD plate. Cells from fresh colonies were spotted on the plate with nylone membrane (1-3 mm spot).
- Plates were incubated at 30°C for 2 – 3 days (colony side down).
- A sterile Whatman paper was presoak at Z-buffer/X-Gal solution.
- The nylone membrane with grown colonies was frozen in liquid nitrogen. After thawing, the permeabilization of cells by freezing was repeated twice again.
- The membrane was placed on the presoaked Whatman, colony side up, avoiding air bubbles under the membrane.
- Membrane was incubated at 30°C and blue coloring of colonies was monitored.

Solutions:

Z-buffer	21.5 g/l $Na_2HPO_4.12H_2O$
	6.2 g/l $NaH_2PO_4.2H_2O$
	0.246 g/l $MgSO_4.7H_2O$
	0.75 g/l KCl
	pH adjusted to 7.0. Sterilized by autoclaving.
X-gal/Z-buffer	66 μl X-gal in DMF (50 mg/ml, Promega, USA) was resolved in 10 ml of Z-buffer with 27 μl β-merkaptoethanol (BIO-RAD)

2.5 Synthetic lethal screen

Synthetic lethal screen was essentially conducted as described in (Barbour and Xiao, 2006). The technique is based on color appearance of colonies with different genetic combination: *ade2 ADE3* cells form red colonies, while *ade2 ade3* colonies are white. Gain or loss of *ADE3* plasmid from a fraction of cells during colony development results in red/white sectoring.

prp45(1-169) cells with appropriate genetic background (*ade2 ade3 leu2 ura3*; white) where transformed by *URA3* plasmid expressing wild-type *PRP45* and *ADE3* genes (red colonies). Thin layer of cell suspension was UV irradiated (254 nm, 20-30 mJ/cm^2) and cells were plated on YPAD. Non-sectoring red colonies (which cannot lose *PRP45* plasmid due to occurrence of SL mutation) were picked and false positives were eliminated by testing of growth on 5-FOA containing plates (5-fluoroorotic acid; negatively selects against *URA3* expressing cells; false positives are able to grow on 5-FOA containing media, as they finally can lose *PRP45 URA3* plasmid). Clones with SL mutations (red) were transformed by YEp13-based genomic library (LEU2; Reed et al., 1989) and plated on SD -Leu. Plasmids were recovered from sectoring colonies (fragment of genome on YEp13 vector support growth without need of *PRP45* on *URA3 ADE3* plasmid). Genomic inserts were sequenced, genes suspect of SL suppression were expressed from plasmid in respective SL clones. The ability of plasmid to support growth on 5-FOA plates confirmed that SL mutation occurred in the gene tested (*PRP45 URA3* plasmid is no more essential for growth due to the presence of gene on plasmid).

2.6 Bioinformatics, data resources, and software tools

2.6.1 Sequence resources

Saccharomyces cerevisiae sequences were downloaded from Saccharomyces Genome Database (SGD; www.yeastgenome.org). Sequence data of other fungal species were obtained from Broad Institute Fungal Genome Initiative website (http://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/fungal-genome-initiative) or from Genolevures database (www.genolevures.org).

Intron data were downloaded from SGD, Ares Lab Yeast Intron Database (http://compbio.soe.ucsc.edu/yeast_introns.html) and from Genosplicing database (http://genome.jouy.inra.fr/genosplicing/).

When appropriate, sequences were retrieved from other public resources, including NCBI databases, EMBL-EBI databases, and others.

2.6.2 RNA secondary structure predicting tools

Following algorithms were used for RNA secondary structure prediction:

Name	Link	Reference
RNAfold	http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi	Gruber et al., 2008
MCfold	http://www.major.iric.ca/MC-Fold/	Parisien and Major, 2008
Mfold	http://www.biology.wustl.edu/gcg/mfold.html	Zuker, 2003
RNA Shapes	http://bibiserv.techfak.uni-bielefeld.de/rnashapes/submission.html	Steffen et al., 2006
Alifold	http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi	Gruber et al., 2008

2.6.3 Other web tools

Multiple sequence alignments were generated by ***ClustalW2*** (Larkin et al., 2007).

Primers for PCR were designed with assistance of ***Primer Quest***^{SM} and ***Oligo Analyzer 3.1*** (IDT DNA Tools).

Similar sequences were searched in databases using ***BLAST*** and derived algortitms.

Proteins´ interaction networks were examined by ***STRING 9.0 Database*** (Szklarczyk et al., 2011).

2.6.4 Software

Clone Manager 9 (Scientific and Educational Software) was used for *in silico* cloning and design of DNA manipulations.

BioEdit 5 (Tom Hall) was employed to analyze multiple types of sequence data.

Image analysis was performed using ***Image Quant***^{TM} ***TL*** (GE Life Sciences).

3. RESULTS

3.1 Prp45 in splicing

(related to publication 1, Gahura et al., 2009)

3.1.1 Prp45(1-169)

In order to map the functional motifs of Prp45, we generated the set of truncated/mutated *prp45* variants (Martinkova et al., 2002; Katerina Abrhamova, PhD. Thesis, 2010; see Fig. 1.10 and section 1.10 Prp45 in Literature Review). In the paper Gahura et al., 2009, we present a characterization of *prp45*(1-169), an allele encoding the protein lacking a C-terminal portion of amino acid sequence, including the conserved SNW motif.

The growth rate of *prp45-Δ* cells overexpressing Prp45(1-168) from a strong *ADH1* promoter steadily declined after the temperature shift to 37°C and became negligible after ~24 hours. To study the cells expressing the temperature sensitive variant of Prp45 at the physiological levels, we prepared a *prp45*(1-169)-3HA::*kanMX6* substitution at the *PRP45* locus, using primer directed homologous recombination in the strain EGY48. The resulting mutant cells (strain KAY02) stop the growth in a liquid culture approximately 6 hours after the transfer to the nonpermissive 37°C temperature (Fig. 3.1B). Although the same results were obtained with the *prp45*(1-169) substitution in the strain W303-1a (KAY01), the level of temperature sensitivity turned out later to be dependent on the genetic background (data not shown).

prp45(1-169) strains exhibit morphological abnormalities suggesting cell division defects. A proportion of cells have deformed, rod-like shaped, prolonged buds (Fig. 3.1A). The testing of diverse stressors showed the hypersensitivity of *prp45*(1-169) cells to microtubule destabilizing drugs (nocodazole and carbendazim) and to cycloheximide (Fig. 3.1C).

All experiments described in this paragraph are included to introduce the *prp45*(1-169) allele and were performed by K. Abrhamová.

3.1.2 Genetic interaction of *prp45*(1-169)

The *prp45*(1-169) allele is lethal only at an elevated temperature, while at permissive temperature (30°C), the *prp45*(1-169) strain grows at the wild-type rate. The moderate phenotype of this mutation makes it suitable to search for negative genetic interactions. We performed a comprehensive screen to identify alleles, which are synthetic lethal with *prp45*(1-169). Procedure of the screen is described and schematically summarized in Fig. 3.2.

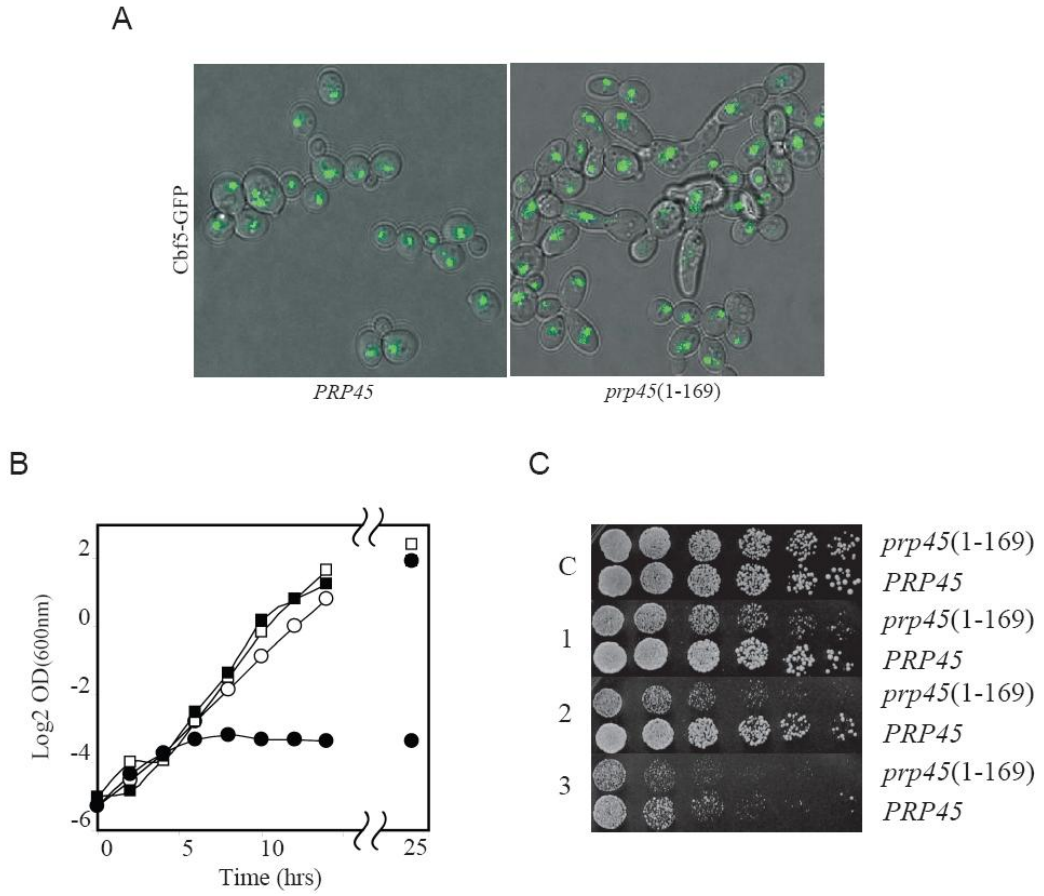


Figure 3.1. *prp45*(1-169) allele exhibits multiple defective phenotypes. (A) Morphological abnormalities of *prp45*(1-169) cells. Indicated strains were grown at 30°C and observed under confocal microscope. Cgf5-GFP was used to visualize nuclei. **(B)** Temperature sensitivity of *prp45*(1-169) cells. Growth curve of wild-type (squares) and *prp45*(1-169) (circles) cells was measured in the SD medium at 30°C (empty symbols) or 37°C (filled symbols). **(C)** *prp45*(1-169) exhibits hypersensitivity to nocodazole and cycloheximide. Exponentially growing cell were harvested, concentrated, and spotted in threefold serial dilution on YPAD plates (C) or YPAD containing 5 μg/ml nocodazole (1), 7.5 μg/ml nocodazole (2), or 0.04 μg/ml cycloheximide (3).

To complement the SL screen results, several previously characterized temperature sensitive mutants of splicing factors were tested for the synthetic lethality with *prp45*(1-169). The *prp45*(1-169) strain KAY02 harboring a plasmid-born *PRP45* (p416ADH-*PRP45*) was crossed with the candidate mutant strains. The diploids obtained were sporulated. The synthetic lethal relationship was then tested by the plasmid shuffling on 5-FOA (Fig. 3.3).

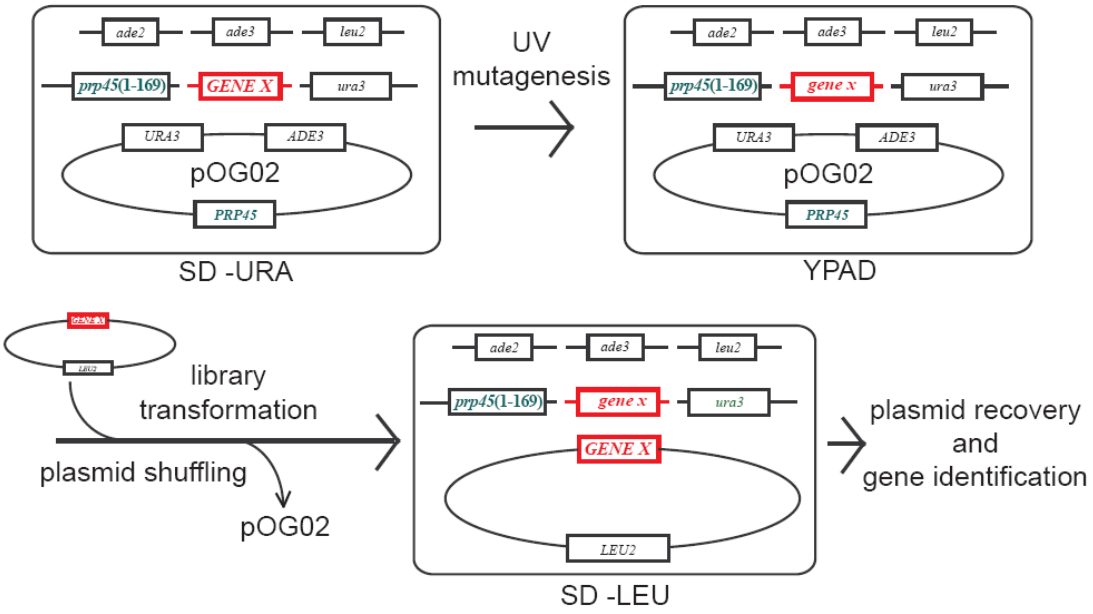


Figure 3.2. Principle of synthetic lethal screen. Strains FPY4B or AVY11 (both with the genotype *prp45*(1-169):*kanMX6 ade2 ade3 leu2 ura3*) expressing wild-type *PRP45* from the plasmid pOG02 (*PRP45 ADE3 URA3*) were randomly mutagenized by UV irradiation. The clones, wherein SL mutations were generated in so far unknown *gene X*, cannot lose the *PRP45* harboring plasmid, which express also the *ADE3* gene. *ADE3* in the combination with the nonfunctional *ade2* allele in the genome leads to the red coloring of colonies. The red non-sectoring colonies were picked and synthetic lethal phenotype was confirmed by inability to grow on the media containing 5-fluoroorotic acid (5-FOA – negative selection against *URA3* expressing cells; *URA3* accompanies *PRP45* and *ADE3* on the pOG02 plasmid; this step is not shown for simplicity). Strains obtained were transformed by the YEp13-based genomic library (Reed et al., 1989). Mutated genes were identified by sequencing of the genomic fragments in plasmids recovered from the white or sectoring colonies, wherein the plasmid shuffling occurred.

Altogether, we identified ten genes with a mutation synthetic lethal with *prp45*(1-169). All interactors are summarized in Table 3.1. Among them, two main groups of splicing factors occur: (i) NTC members (Clf1/Syf3, Syf1, Ntc20, Cef1) and (ii) second step factors (Prp22, Prp18, Slu7 and Prp17). Sequencing of two *slu7* alleles revealed that the single missense P268L or K258E substitutions are responsible for their SL phenotype. Both mutations map to the essential highly conserved central part of the amino acid sequence (Fig. 3.4), to the close proximity of a Prp18 binding motif (Zhang and Schwer, 1997). Only one gene identified (*COF1*) does not encode a splicing factor. Notably, the SL mutation in *COF1* localizes to the intron. All genetic interaction identified are thus in agreement with the role of Prp45 in splicing.

Some of *prp45*(1-169) genetic interaction mentioned here were already published in the diploma theses by O.G., Anna Valentová, or Vanda Munzarová.

Table 3.1. Summary of *prp45*(1-169) synthetic lethal interactions.

Gene	Number of isolates	Identified allele(s)
UV mutagenesis library screen		
PRP22	3	*prp22*(300PPI) *prp22*(-158T) *prp22*(-327A)
SLU7	2	*slu7*(P268L) *slu7*(K252E)
CLF1/SYF3	2	*syf3*(T402I T404F) ND
CEF1	1	ND
PRP18	1	*prp18*(K234N)
SYF1	1	ND
NTC20	1	ND
PRP8	1	ND
COF1	2	G149A in intron
Testing of previously characterized alleles		
PRP22	1	*prp22-1*
PRP17	1	*prp17*-Δ
CEF1	1	*cef1-13*

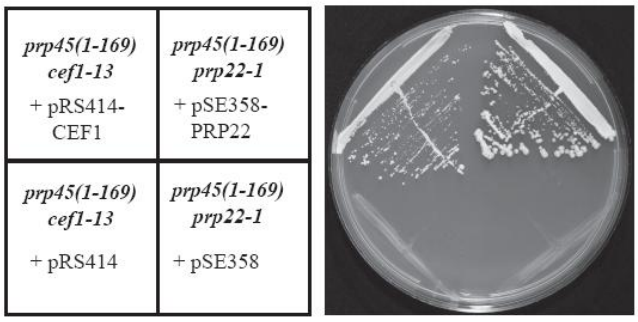


Figure 3.3. Direct testing of synthetic lethal interactions. The double mutant *prp45*(1-169) *cef1-13* strain expressing wild-type *PRP45* from the *URA3* plasmid was transformed by pRS414-*CEF1* or empty vector and tested to growth on the 5-FOA containing media. Analogical testing was performed with the *prp45*(1-169) *prp22-1* strain using pSE358-*PRP22*.

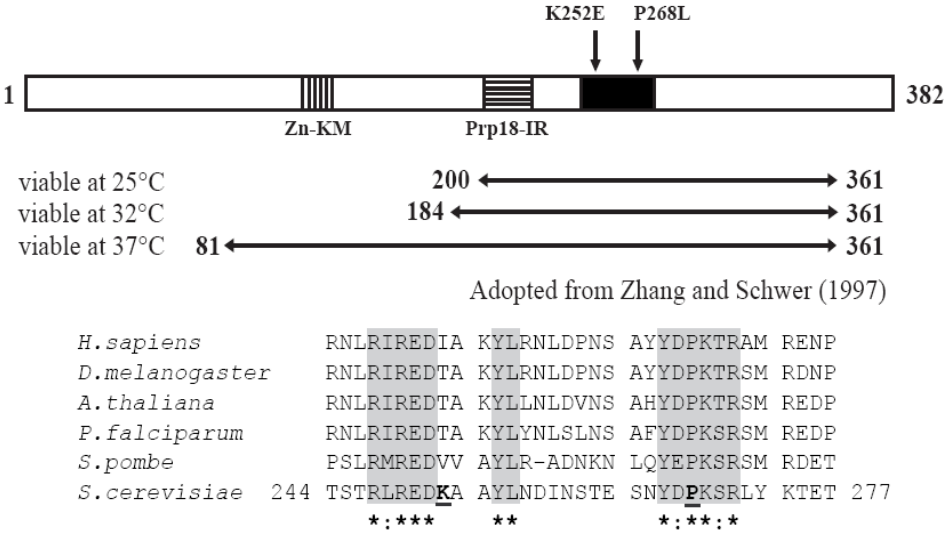


Figure 3.4. *slu7* mutations synthetic lethal with *prp45*(1-169) localize to the conserved C-terminal part. Diagram shows the conserved regions and the amino acid substitutions of Slu7 identified in the *prp45*(1–169) SL screen. Regions required for viability at various temperatures are depicted (Zhang and Schwer, 1997). The alignment of Slu7 proteins from indicated species was constructed using ClustalW. The novel *slu7* mutations are underlined; invariant and conserved residues are marked by stars and colons, respectively.

3.1.3 Functional interaction of Prp45 and Prp22

Using tandem affinity purification (TAP) with a TAP-tagged NTC component Cwc2, we isolated the spliceosomal complexes from wild-type and *prp45*(1-169) cells. Proteins that co-purified with Cwc2 were electrophoretically separated on the gradient SDS polyacrylamide gel. After staining, the bands were annotated based on

the pattern correspondence with the available TAP datasets. In addition, the major bands were cut off and the identities of proteins were confirmed by mass spectrometry. The analysis of complexes revealed significantly decreased partition of the Prp22 protein in the Cwc2-TAP pull-down from *prp45*(1-169) cells in comparison to the wild-type levels. Immunoblot using the anti-Prp22 and anti-TAP specific antibodies confirmed this result. TAP experiments were conducted by Michal Skružný. For results see Fig. 4A in Gahura et al., 2009.

Prp22 is an ATP-dependent RNA helicase from a DExH box family involved in the second transesterification and in the release of spliced mRNA from spliceosome (Schwer and Gross, 1998). Previously, Prp22 was shown to interact in two hybrid system with Prp45. All identified Prp22-interacting Prp45 fragments included the region of amino acids 262-291 (Albers et al., 2003; Fig. 3.5B), which is no more present in the *prp45*(1-169) variant. Indeed, although we confirmed two-hybrid interaction of Prp45 and Prp22, we did not observe Prp22 interaction with the Prp45(1-169) variant. Consistently with results in Albers et al., 2003, we documented the interaction of Prp22 with the C-terminal fragment of Prp45 (amino acids 258 to 379), although much weaker than with the complete protein (Fig. 3.5A).

We hypothesize that Prp45 is one of proteins which are involved in the recruitment or regulation of the Prp22 helicase. We identified four *prp22* alleles synthetic lethal with *prp45*(1-169) (see Table 3.1). Two of them bear a mutation upstream of the coding region (*prp22*(-158T) and *prp22*(-327A)). Western analysis revealed decreased Prp22 amounts in these mutants (Fig. 3.6A). Thus, reduced levels of Prp22 result in the synthetic lethality with *prp45*(1-169). Conversely, a moderate overexpression of Prp22 partly suppressed the temperature sensitive phenotype of *prp45*(1–169) cells (Fig. 3.6B). Other SL allele, *prp22*(300PPI), bears the triple amino acid substitution (R300P, Q301P, L302I) within a region N-terminal to the ATPase domain. This part was documented to be necessary but not sufficient for spliceosome binding (Schneider and Schwer, 2001). Finally, *prp22-1*, the allele which was identified by the direct testing (Fig. 3.3), harbors the G776E substitution in the inter-domain cleft of the ATPase catalytic domain (Arenas and Abelson, 1997). Thus, both the mutations within the C-terminal half of Prp22, containing the ATPase/helicase domain, and within the N-terminal half are synthetic lethal with *prp45*(1–169). Notably, while both halves are required for functional spliceosome

association, they function also when expressed *in trans* (Schneider and Schwer, 2001).

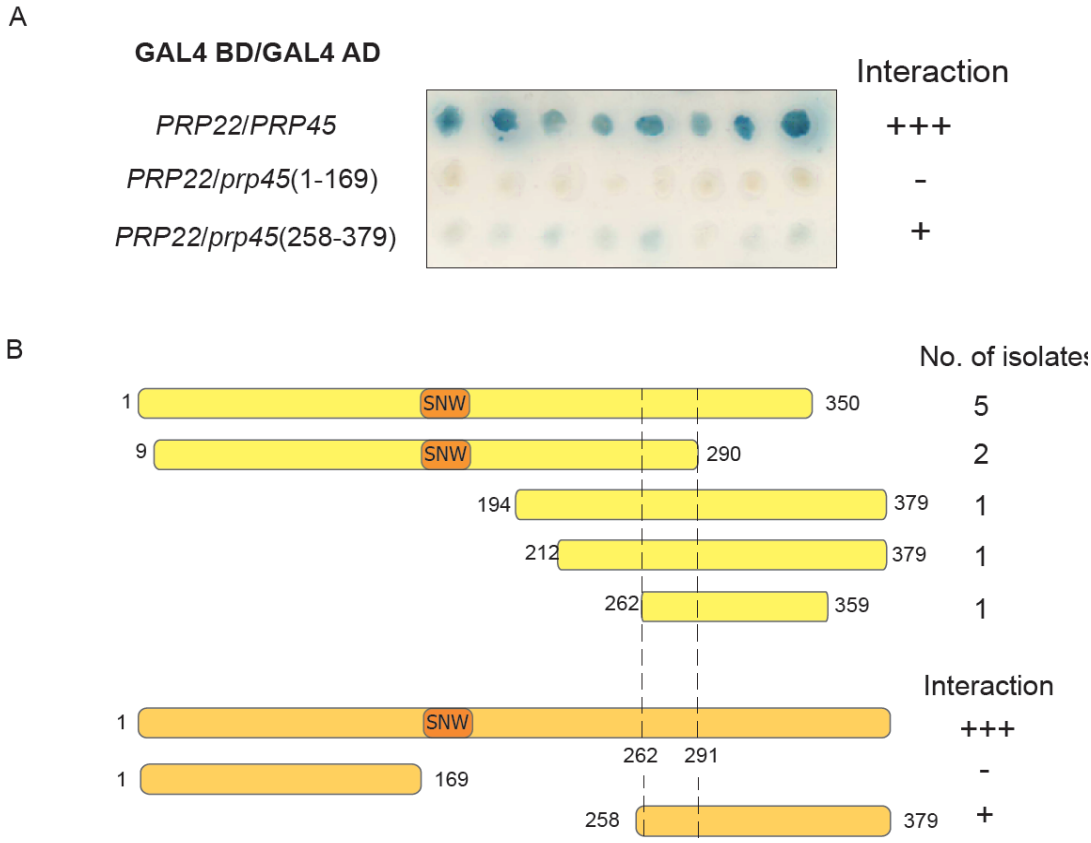


Figure 3.5. Two hybrid interaction of Prp45 with Prp22. (A) The C-terminal part of Prp45 is required for the two hybrid interaction with Prp22. The two hybrid interactions of full length Prp22 with wild-type Prp45 and its truncated variants were tested by β-galactosidase colony lift filter assay. Prp22 was fused to the GAL4 DNA binding domain while Prp45 and its derivatives to the GAL4 activation domain. **(B)** Summary of the Prp45-Prp22 two hybrid interactions. The Prp45 fragments isolated with Prp22 as a bait in a systematic two hybrid screen (Albers et al., 2003) are depicted in yellow. In orange, fragments tested in (A) are schematized.

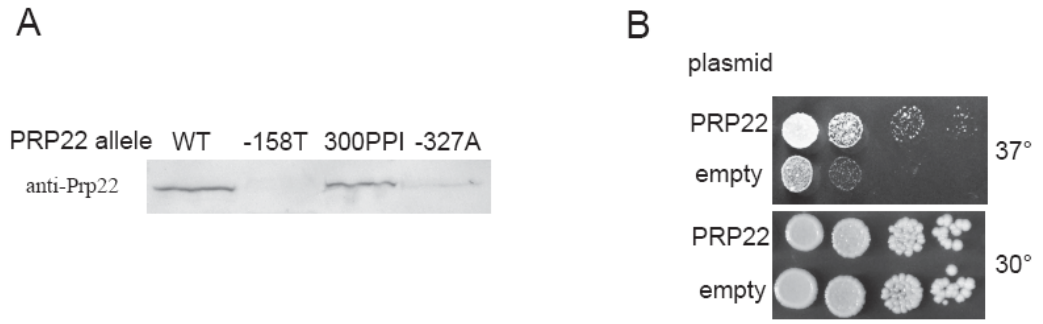


Figure 3.6. Prp22 concentration affects viability of *prp45*(1-169) cells. (A) Two *prp22* mutations which are SL with *prp45*(1-169) result in the decreased amount of Prp22 protein in cells. Prp22 levels in strains with indicated *prp22* alleles were determined by immunodetection with the anti-Prp22 antibody (gift from B. Schwer). **(B)** Prp22 overexpression partially suppresses the temperature sensitivity of *prp45*(1-169) cells. *prp45*(1-169) cells transformed by the 2μ plasmid expressing *PRP22* from the autologous promoter or by an empty vector were cultivated to the mid-log phase, harvested, concentrated, spotted in tenfold serial dilution on YPAD plates, and cultivated for 4 days at indicated temperatures.

3.1.4 Prp45 is required for splicing of substrates with non-canonical splicing signals

We employed a set of reporter genes with *ACT1*-derived introns to assess, whether the *prp45*(1-169) mutation results in a splicing defect and if so, whether the defect can be assign specifically to the first or the second step. For a description of reporter constructs see section 2.2.11 Plasmids. Cells were transformed with *ACT1-CUP1* fusion expressing plasmids, containing either the *ACT1* intron with canonical splicing signals, or the intron with the altered 5´ss, BP, or 3´ss sequence. RNA isolated from cells was subjected to primer extension analysis. Levels of pre-mRNA, splicing lariat-exon 2 intermediate, and spliced mRNA were compared between wild-type and *prp45*(1-169) strains.

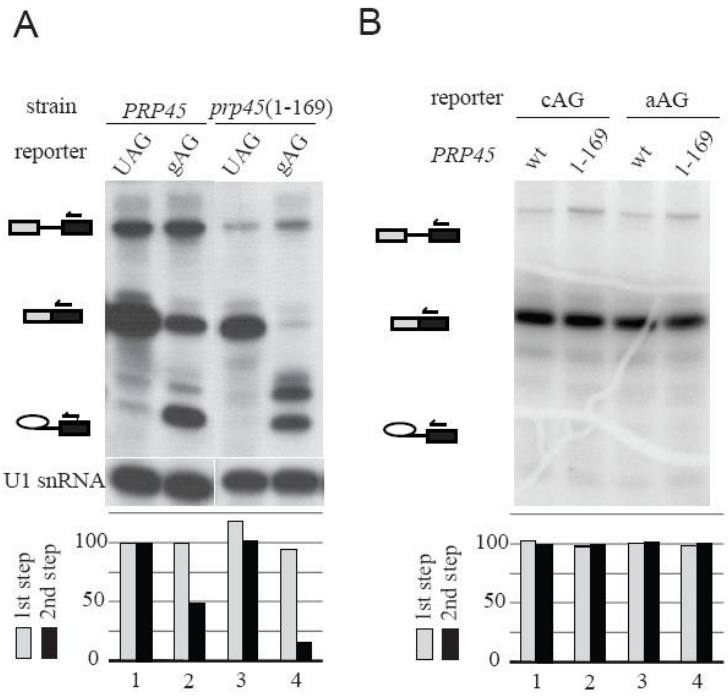


Figure 3.7. Primer extension analysis of splicing efficiency of canonical and 3´ss mutated substrates in wild-type and *prp45*(1-169) cells. Cells expressing indicated variant of *ACT1-CUP1* reporter construct were cultivated at 30°C to mid-log phase. RNA was isolated and subjected to primer extension as described in Materials and Methods. The products corresponding to pre-mRNA (P), mRNA (M), and lariat-exon 2 intermediate (LI) are indicated by their icons. The signal appearing between M and LI bands in A was found to represent primer extension products of the endogenous *CUP1-1/CUP1-2* locus. The U1 snRNA content was assayed as a loading control (A, bottom panel). For the construction of the bar diagrams, we used the phosphorimager signals from several independent experiments to quantify the efficiencies of the 1st (M+LI)/(P+M+LI) and 2nd step M/(M+LI). The values for each construct were normalized to the corresponding efficiencies of the canonical *ACT1* reporter in wild-type strain. **(A)** Splicing defect of the 3´ss mutated gAG substrate was exacerbated in *prp45*(1-169) cells, while splicing of canonical *ACT1* remained unaffected. **(B)** The substrates with cAG or aAG at 3´ss are spliced efficiently in both wild-type and *prp45*(1-169) cells.

The intron with canonical splicing signals were spliced efficiently in both wild-type and *prp45*(1-169) strains (Fig. 3.7A, lanes 1 and 3). Even in *prp45*(1-169) cells shifted to the restrictive temperature 37°C for 4 to 8 hours, when the growth is arrested, its splicing does not seem to be impaired (Fig. 3.8). The U to G substitution three nucleotides upstream of intron/exon junction in the 3´ss (the construct referred to as "gAG") is less efficient substrate for the second step ("second step limiting substrate") and it was previously used to study the second step factors, including Prp22 (Mayas et al., 2006). The second step defect of gAG reporter is highly exacerbated in *prp45*(1-169) cells (Fig. 3.7A, lanes 2 and 4). Splicing in cells cultivated at 37°C yielded the same results (Fig. 3.8). As mentioned above,

prp45(1-169) exhibits various degree of temperature sensitivity, depending on the genetic background. We tested two *prp45*(1-169) temperature resistant strains and both of them displayed the same splicing phenotype with the gAG substrate as the temperature sensitive KAY02 cells (Fig. 3.9). No effects on splicing efficiency were observed for reporters with cAG or aAG sequence at 3´ss (Fig. 3.7B). As CAG and AAG sequences occur in 3´ss of many yeast introns, such substrates can be considered as optimal.

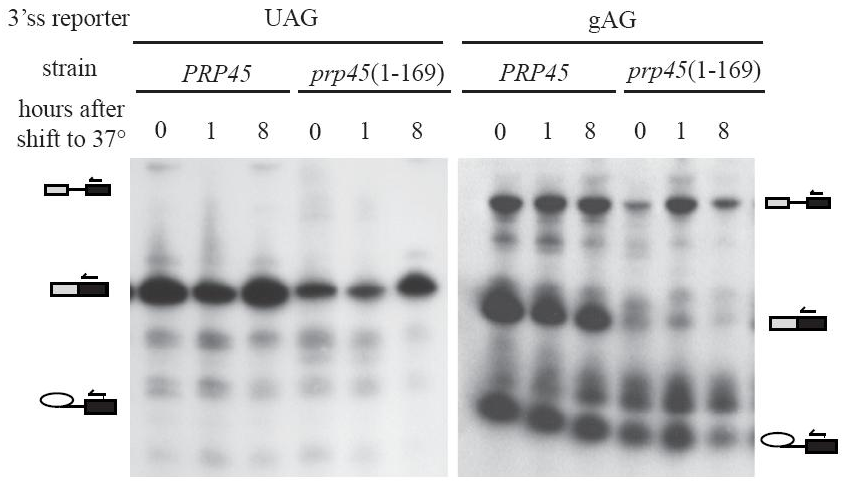


Figure 3.8. Shift to the restrictive temperature does not lead to a profound effect on the splicing of the canonical and 3'ss gAG substrates in *prp45*(1-169) cells. Cells expressing indicated *ACT1-CUP1* fusion constructs were grown at 30° to the mid-log phase, diluted in a preheated medium, and incubated at 37° for the indicated times. RNA was isolated and subjected to primer extension analysis as described in Materials and Methods. The products corresponding to the pre-mRNA (P), mRNA (M), and lariat-exon 2 intermediate (LI) are indicated by their icons.

We used two distinct substrates with the A3C mutation in the 5´ss. The sequence of the first substrate corresponds to the *ACT1* intron (except for the A3C substitution), therefore we call it "actin" substrate. The other A3C reporter sequence differs in the nucleotides flanking UAG in the 3´ss, mainly in the exon. We refer to it as "recombinant" substrate. For details on sequences see Appendix A1. In the wild-type background, both substrates appeared to be limiting for both steps, although the recombinant substrate only mildly (Fig. 3.10A, lanes 1 and 5). However, in the background of *dbr1*-Δ, that harbors the deletion of *DBR1* gene encoding a debranching enzyme, a key player in the degradation pathway of lariat-exon 2 intermediates, the signal from the first step product became more profound and the

second step thus turned out to be predominantly limiting for the A3C substrates (Fig. 3.10A, lanes 3 and 7). The A3C substitution generates an additional nucleotide pair in the helix formed by the U6 snRNA and the 5´ss after the rearrangement possibly facilitated by NTC (Chan and Cheng, 2005). This condition is interpreted as the hyperstabilization of the first step conformation (Konarska et al., 2006). *prp45*(1-169) mutation displayed an exacerbation of the second step defects of both A3C substrates (Fig. 3.10A).

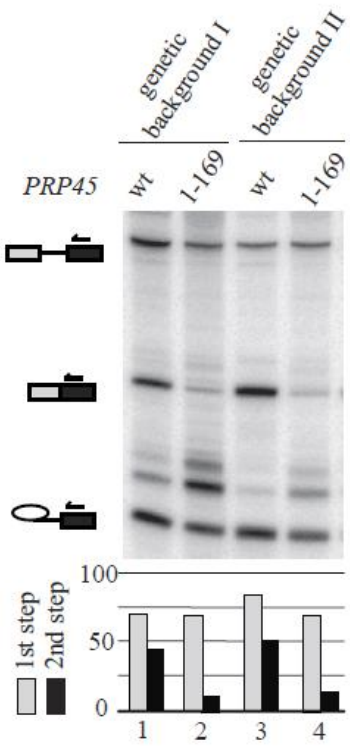


Figure 3.9. Temperature resistant *prp45*(1-169) strains do not exhibit distinct splicing phenotype than temperature sensitive strain KAY02. Primer extension analysis was performed on RNA isolated from cells expressing the gAG *ACT1-CUP1* reporter gene. Quantification of splicing efficiency was done as described in Fig. 3.7. Analysis was performed using two pairs of wild-type and respective *prp45*(1-169) knock-in strains. "Genetic background I" - strain BJ2168. "Genetic background II" - strain KGY823.

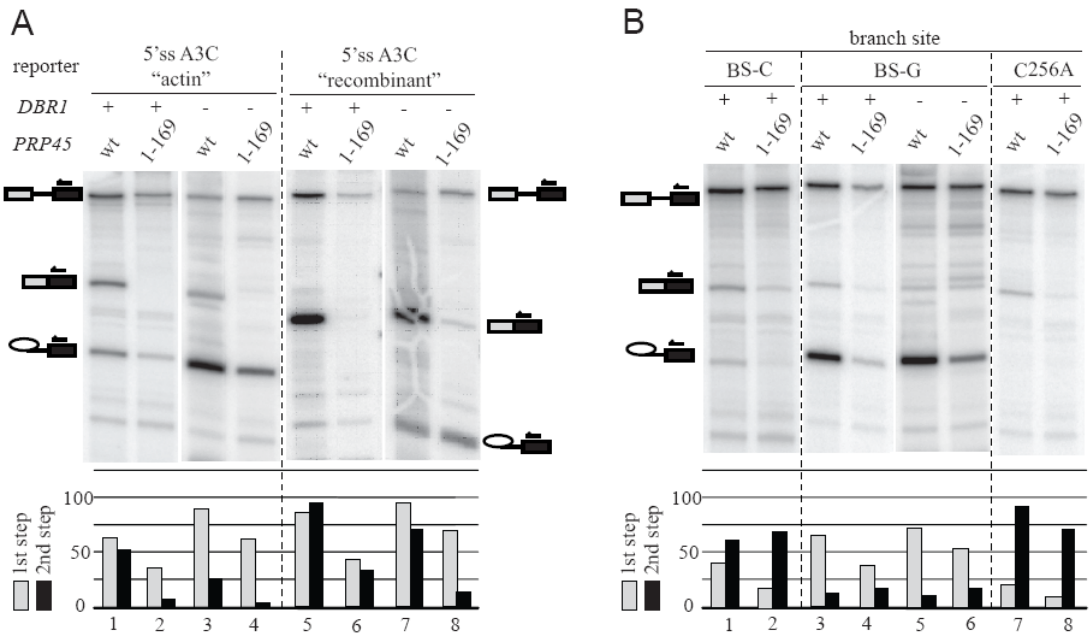


Figure 3.10. Splicing defect of the 5′ss (A) and BP (B) mutated substrates is exacerbated in *prp45*(1-169) cells. Primer extension analysis and quantification was performed as described in Fig. 3.7. The *dbr1-Δ* variant was used to underline the second step defect where indicated.

The splicing of the branch point mutated *ACT1-CUP1* substrates in *prp45*(1-169) cells showed the exacerbation of the first but not the second step defect (Fig. 3.10B). Assays were performed using a construct with the branch site A to G mutation (BS-G), which is predominantly "second step limiting", and with the A to C mutation (BS-C), which forms the lariat-exon 2 product inefficiently and which is limiting for both steps. In the cells lacking the lariat debranching activity (*dbr1-Δ*), the accumulation of lariat-exon 2 intermediate through defective splicing was not attenuated by its removal, thus providing a more direct measure of the second step efficiency (Fig. 3.10B, lanes 5 and 6). In addition, the C256A mutation in the branch point proximity, which destabilizes the first step conformation presumably because of its less efficient base pairing with the U2 snRNA (Smith et al., 2007), was tested. We found that *prp45*(1–169) exacerbated the strong first step defect associated with this substrate (Fig. 3.10B, lanes 7 and 8).

3.1.5 Splicing defects of synthetic lethal *prp45* interactors

In *prp45*(1-169) cells, the second step limiting 3´ss gAG substrate exhibited the most profound splicing defect. Therefore, we decided to analyze its splicing pattern also in the strains with *prp45*(1-169) SL mutations. Interestingly, its splicing efficiency was impaired only by the *slu7*(P268L), *slu7*(K252E), and all identified *prp22* mutants (Fig. 3.11). This observation further supports the relation of Prp45 with the Prp22 functioning in the 3´ss fidelity and in the second transesterification.

The *slu7* alleles did not affect any other substrate tested, including the 5´ss A3C "recombinant" and C256A (data not shown). The *prp22* mutations did not affect splicing of the A3C "recombinant" substrate, while all of them impaired the splicing efficiency of the A3C "actin" substrate. Surprisingly, the defect observed substantially varied among the *prp22* alleles and was not similar to the impairment documented in the *prp45*(1-169) strain (Fig. 3.12). The distinct behavior of A3C substrates might be explained by the differences in the sequence downstream of the 3´ss. It would be in agreement with the finding that Prp22 directly interacts with this region (Schwer, 2008). Another unexpected observation was, that the mutations in *prp22* promoter, which resulted in the decreased Prp22 protein levels in cells (see Fig. 3.6A), affected the splicing efficiency in entirely different manner (Fig. 3.12B, lanes 3 and 5). At present, we do not have any plausible explanation for this observation.

Prp18 is a non-essential second step specific splicing factor directly interacting with Slu7 (James et al., 2002). The *CUP1*-based assays revealed, that *prp18*-ΔCR allele lacking a highly conserved central motif exhibits variable splicing efficiency in dependence on the intron-flanking sequences, while wild-type cells are insensitive to these manipulations (Bacikova and Horowitz, 2005; Crotti et al., 2007). These phenotypes may be at least partially explained by a functional relationship between Prp18 and Prp22. We performed the primer extension analysis of the *prp45*(1-169) synthetic lethal allele *prp18*(K234N) with all *ACT1-CUP1* substrates available, but no significant defect was documented (Fig. 3.13).

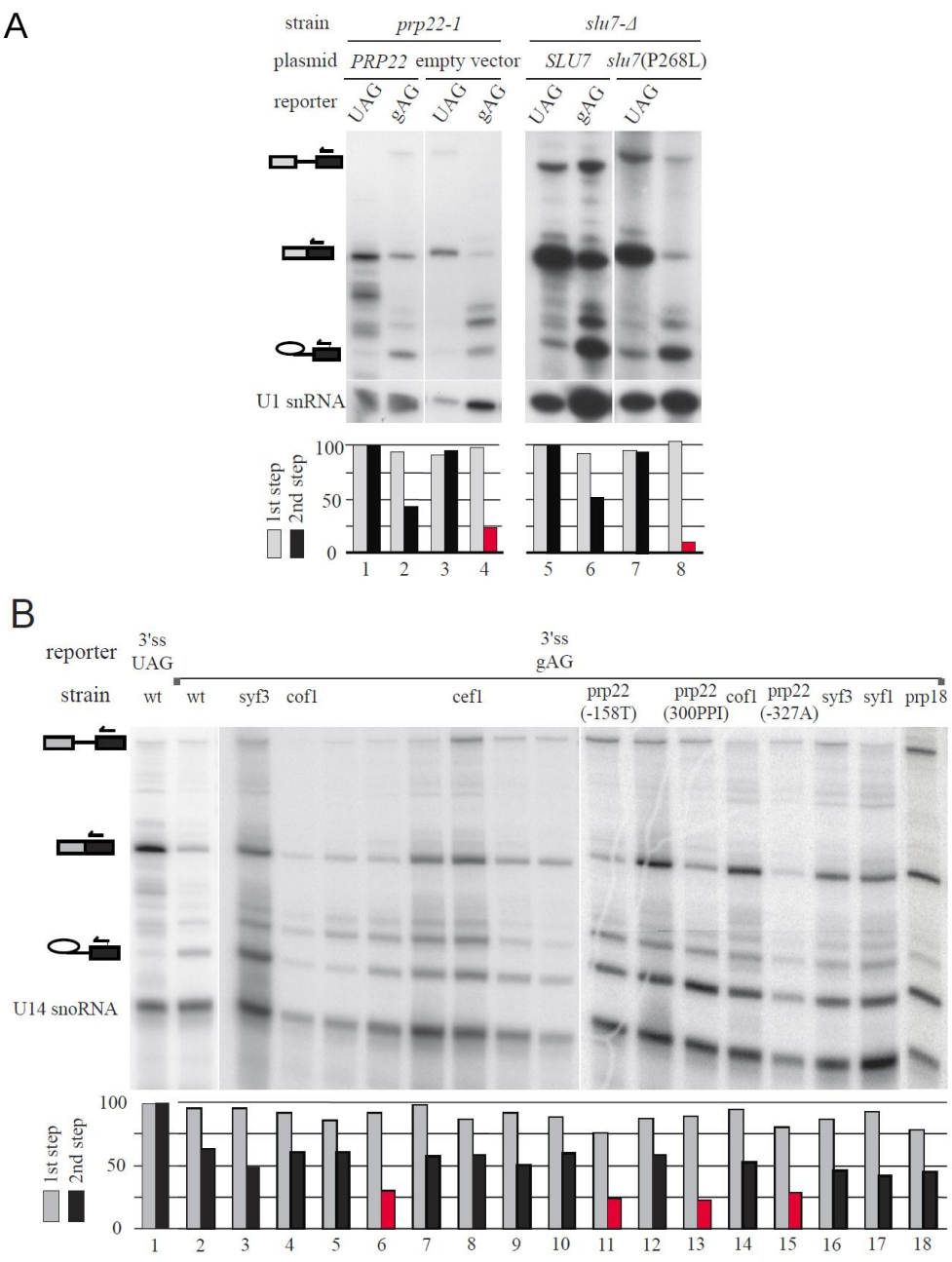


Figure 3.11. Testing of the 3´ss gAG substrate revealed that *prp22* and *slu7* alleles exhibit splicing defect similar to *prp45*(1-169) cells. The canonical (UAG) substrate is involved as a control. Primer extension analysis and quantification was performed as described in Fig. 3.7. Red bars are shown below lanes, where the second step efficiency is decreased in comparison to the wild-type cells, or in comparison to the cells co-expressing the wild-type allele of the respective gene (lanes 1 and 2 in A and B). **(A)** Splicing was assayed in *prp22-1* and *slu7*(P268L) strains transformed by *PRP22* and *SLU7* expressing plasmid, respectively, or by empty vector. **(B)** Splicing of the 3´ss gAG substrate was monitored in all strains obtained in the SL screen with *prp45*(1-169). The SL gene or allele is shown where known.

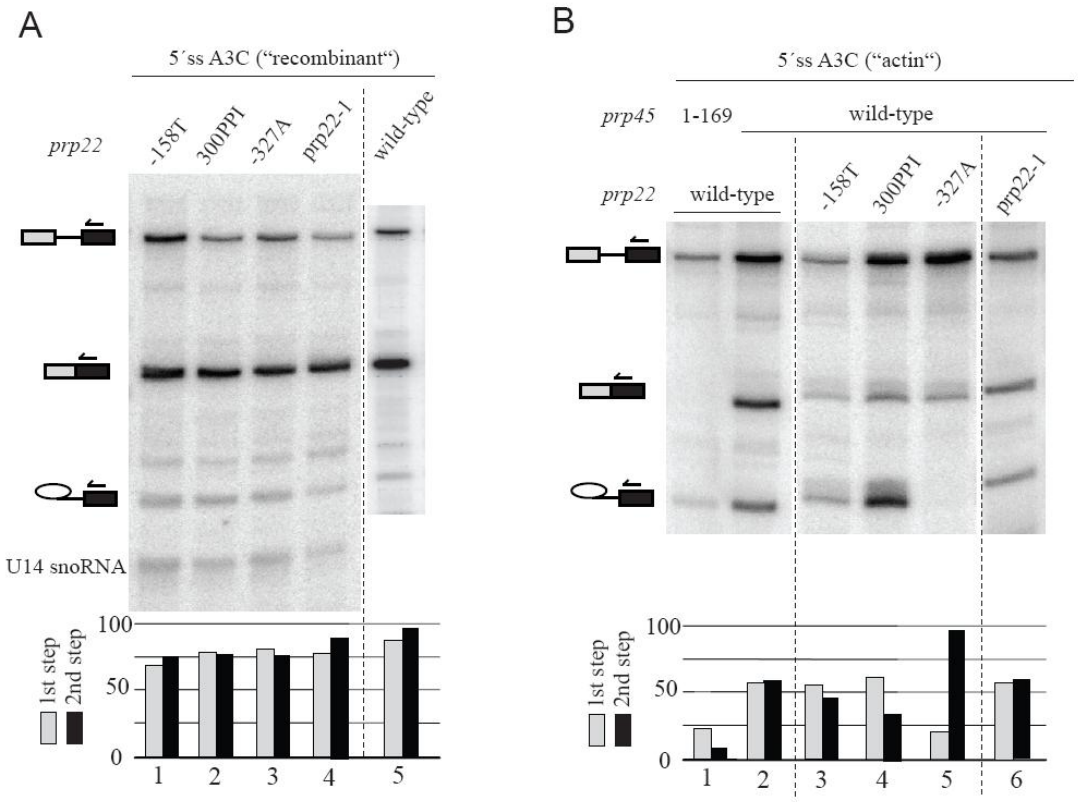


Figure 3.12. The *prp22* alleles SL with *prp45*(1-169) do not affect splicing of the 5´ss A3C "recombinant" substrate (A) and differentially impair splicing of the 5´ss A3C "actin" substrate (B). Primer extension analysis of indicated substrate in indicated genetic background and quantification was performed as described in Fig. 3.7.

3.1.6 Expression of the C-terminal fragment of Prp45 suppresses *prp45*(1-169) phenotypes

As shown by our deletion analysis, Prp45 requires both termini for its essential functioning (Fig. 1.10). A truncation of both ends resulted in the inviable variant *prp45*(53-190). We asked whether the expression of the C-terminal part of Prp45 in *prp45*(1–169) cells *in trans* would affect the *prp45*(1–169) associated phenotypes. Indeed, the expression of Prp45(119–379) eliminated the temperature sensitivity of *prp45*(1–169) cells (Fig. 3.14A).

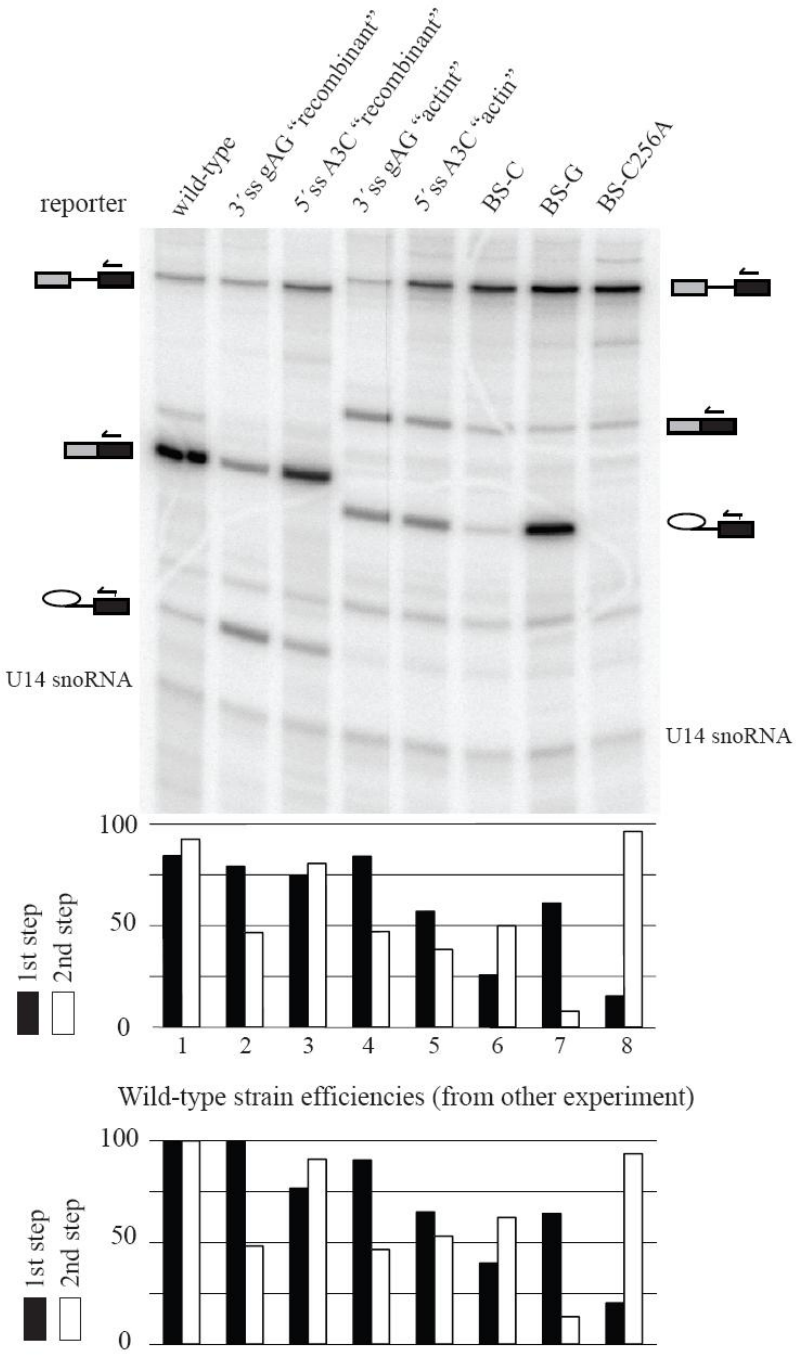


Figure 3.13. The *prp18*(K234N) mutation does not affect splicing of canonical and mutated *ACT1*-derived reporter substrates. Primer extension analysis of indicated substrates in *prp18*(K234N) strain and quantification was performed as described in Fig. 3.7. Lower bar diagram shows splicing efficiencies as quantified in wild-type EGY48 strain in previous experiments.

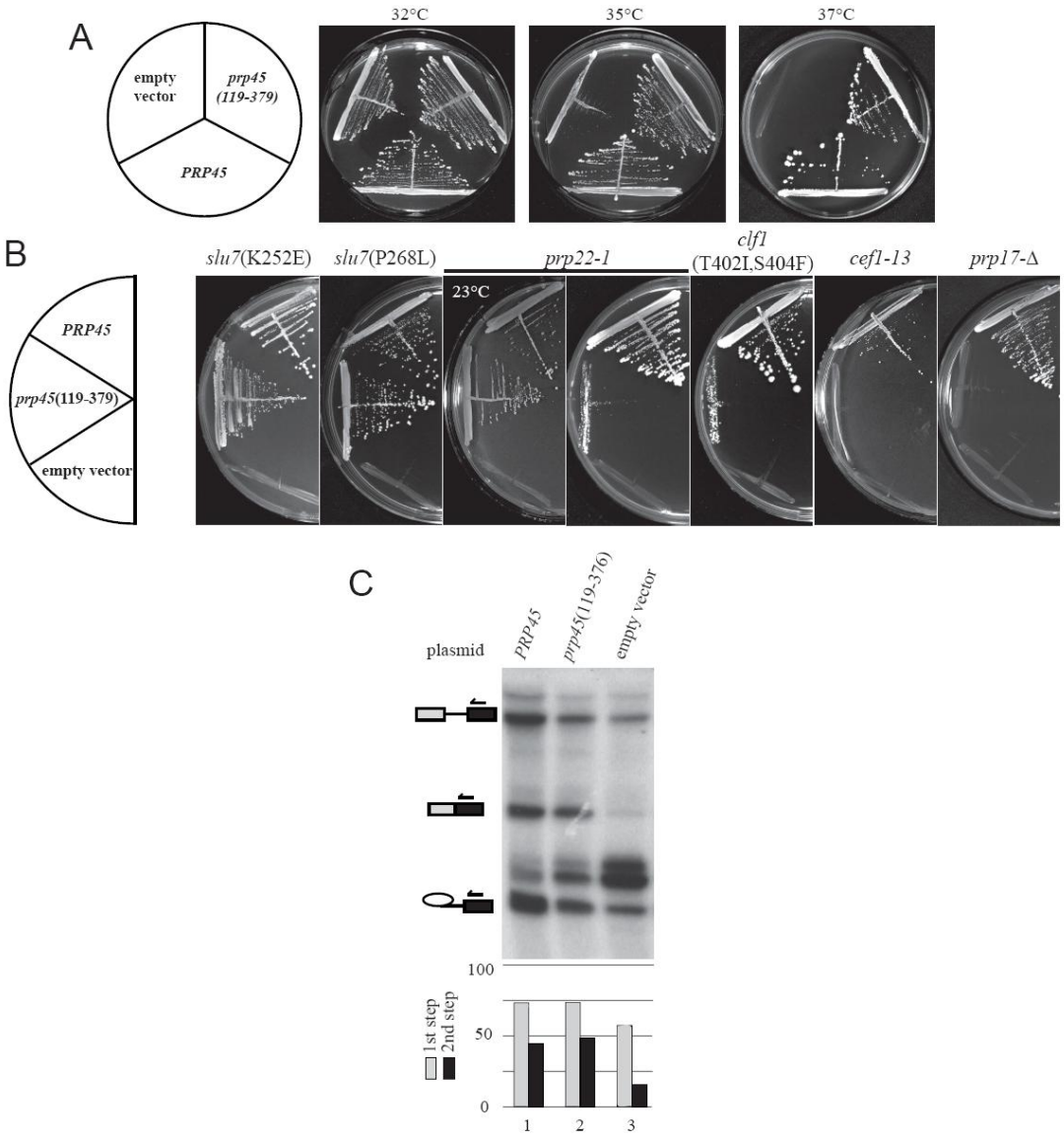


Figure 3.14. Prp45(119-379) complements the *prp45*(1-169) phenotypes *in vivo*. (A) The temperature sensitivity of *prp45*(1-169) strain is suppressed by the expression of Prp45(119-379). *prp45*(1-169) cells transformed by a plasmid expressing Prp45(119-379) were spread on SD plates and cultivated at indicated temperatures. Empty vector and Prp45 expressing plasmid were used as the negative and positive control, respectively. **(B)** Prp45(119–379) suppresses some of the synthetic lethal interactions of *prp45*(1–169). The double mutant strains harboring *prp45*(1–169) and indicated mutation depends on Prp45 expressed from the *URA3* plasmid. Cells were transformed with the plasmid expressing Prp45(119-379) or Prp45; the empty vector was used as a control. Transformants were tested for the growth on 5-FOA plates at 30°C (or at 23°C where indicated). **(C)** Expression of Prp45(119-379) suppresses the splicing defect of the 3´ss gAG substrate in *prp45*(1-169) cells. Splicing was assayed by primer extension analysis as described in Fig. 3.7 on RNA isolated from the *prp45*(1-169) strain expressing the 3´ss gAG reporter gene and Prp45 or Prp45(119-379).

To test whether the synthetic lethal interactions of *prp45*(1-169) would be suppressed by Prp45(119–379), we tested the 5-fluoroorotic acid (5-FOA) sensitivity of the respective double mutant strains expressing *PRP45* from an *URA3* plasmid in the presence of Prp45(119–379). Prp45(119–379) suppressed the synthetic lethal interaction of *prp45*(1–169) with *slu7*(P268L) or *slu7*(K252E) at 30°C as well as with *prp22-1* at 23°C. However, Prp45(119–379) had only a weak suppression capability on *clf1*/*syf3*(T402I,S404F) and had no effect in other SL pairs (*cef1*-13, *prp17-Δ*; Fig. 3.14B).

The Cwc2-TAP purification of spliceosomal complexes from *prp45*(1-169) cells expressing Prp45(119-379) *in trans* revealed the restored partition of Prp22 helicase (Michal Skružný, Fig. 6C in Gahura et al., 2009). The co-expression of the C-terminal fragment of Prp45 therefore can complement the *prp45*(1-169)-related phenotype at the biochemical level. The Prp45 C-terminus is thus required for the proper Prp22 interaction with the spliceosome. To address the question, whether the restored Prp22 partition in spliceosomal species would be reflected by the suppression of the *prp45*(1-169) splicing defects, we assayed the splicing of suboptimal substrates in the *prp45*(1-169) strain co-expressing the C-terminal Prp45(119-379) fragment. Primer extension analysis revealed that under these conditions the splicing efficiencies of all substrates increased and reverted to the wild-type levels (Fig. 3.14C and data not shown)

Taken together, these experiments suggest that the proper Prp45 functioning in splicing can be achieved by the expression of two (overlapping) parts *in trans*.

3.1.7 Overexpression of Prp22 does not suppress splicing defects of *prp45*(1-169)

As *prp45*(1-169) cells exhibit decreased amount of Prp22 in spliceosomal complexes, we asked whether the splicing in such cells would be improved by a high production of Prp22. We overexpressed Prp22 from its autologous promoter on a replicative 2μ plasmid in the *prp45*(1-169) strain (stronger Prp22 overexpression resulted in growth defects) and performed the primer extension analysis with selected reporters. No suppression of splicing defects was observed, as shown on the 5′ss ("actin" A3C) and 3′ss mutated (gAG) substrates (Fig. 3.15). The Prp22 overexpression itself therefore does not suffice to overcome the *prp45* associated splicing defects.

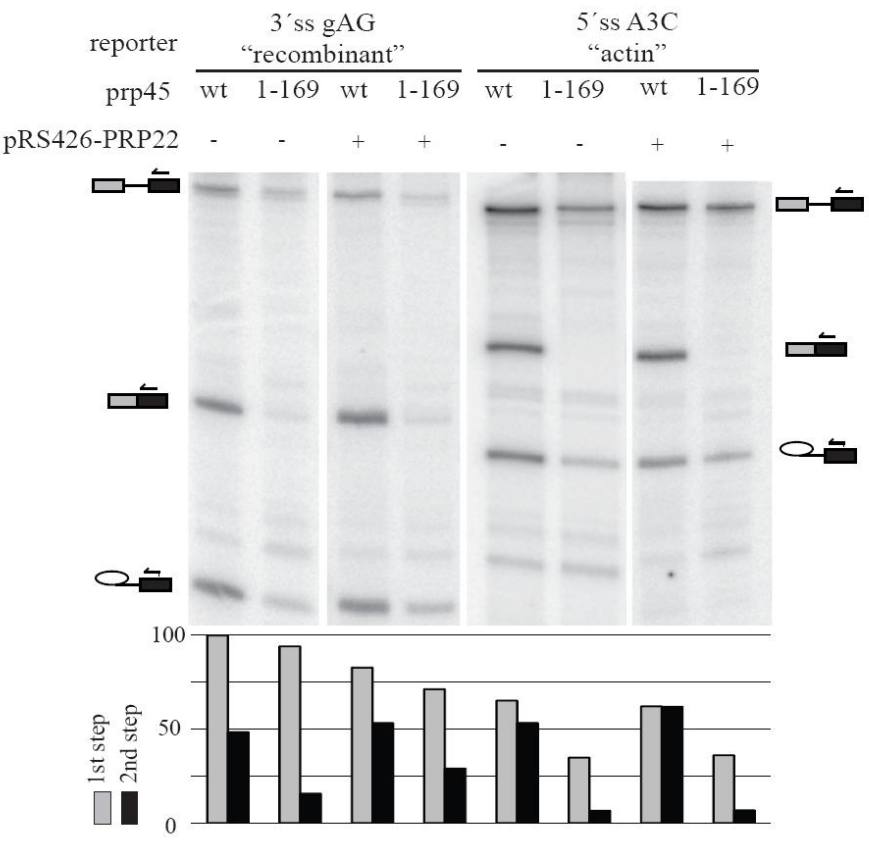


Figure 3.15. Overexpression of *PRP22* does not suppress *prp45*(1-169) splicing defects. Splicing efficiencies of indicated substrates was assayed in wild-type or *prp45*(1-169) strains transformed by a *PRP22* expressing plasmid or empty vector. Primer extension and quantification was performed as described in Fig. 3.7.

3.1.8 Splicing of *COF1* intron depends on Prp45

Our results suggest that Prp45 (or its function related to its C-terminus) may not be required for splicing in general but might affect only a subset of introns. It would be not surprising, as alleles of many splicing factors, including essential core proteins as Prp8, exhibit transcript specific splicing impairments (Clark et al., 2002; Pleiss et al., 2007b). We selected several introns to test their splicing efficiency by primer extension analysis and copper sensitivity assay in *prp45*(1-169) cells. Introns were picked based on the following observations or facts: 1) results of *prp45*(1-169) transcription arrays (K. Abrhamová, PhD Thesis) and splicing specific microarrays, which were done in the collaboration with the Laboratory of Christine Guthrie, UCSF, USA; 2) the finding that Prp45 is required for the splicing of suboptimal substrates; 3) the SL interaction of *prp45*(1-169) with a mutation in *COF1* intron.

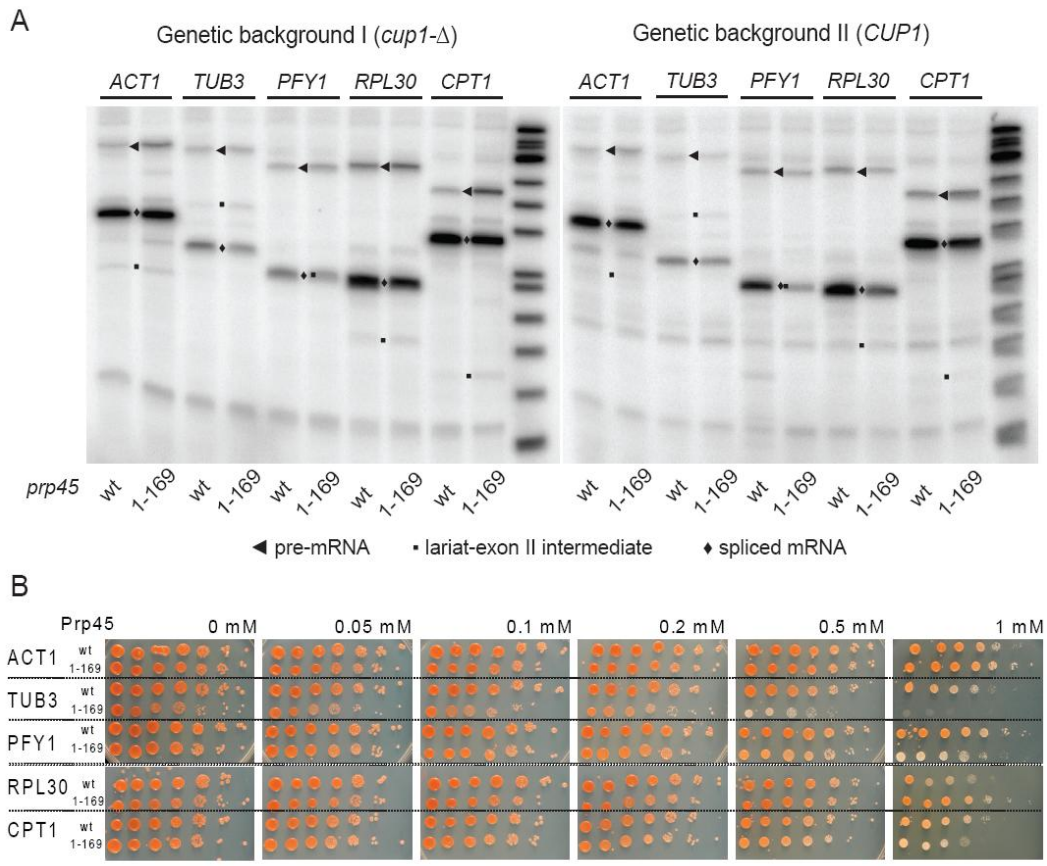


Figure 3.16. *ACT1, TUB3, PFY1, RPL30,* or *CPT1* intron-containing reporter genes are not spliced with decreased efficiency in *prp45*(1-169) cells. (A) Splicing of the reporter genes with indicated introns was assayed by primer extension analysis in wild-type and *prp45*(1-169) cells. Experiment was performed in two genetic backgrounds (left panel: strains 46Δcup vs. MHY04; right panel: strains EGY48 vs. KAY02) **(B)** Wild-type and *prp45*(1-169) cells (strains 46Δcup and MHY04, respectively) expressing *CUP1* fusion splicing reporters with indicated introns were cultivated to mid-log phase, harvested, concentrated to OD 6, spotted in the eightfold serial dilutions on SD minimal media, and cultivated for 3 days at 30°C.

To construct splicing reporters we fused the desired genes' fragments containing the exon 1, intron and 5′ part of exon 2 to the coding sequence of *CUP1* (for details see section 2.2.11 Plasmids). We compared the splicing phenotypes of *prp45*(1-169) with wild-type in two different genetic backgrounds: (i) EGY48 and (ii) *cup1-Δ* strains. Neither primer extension nor spot tests monitoring the resistance to copper revealed any obvious splicing defect of the *ACT1*, *TUB3*, *RPL30*, and *CPT1* introns, although some mild differences between wild-type and *prp45*(1-169) are detectable (Fig. 3.16).

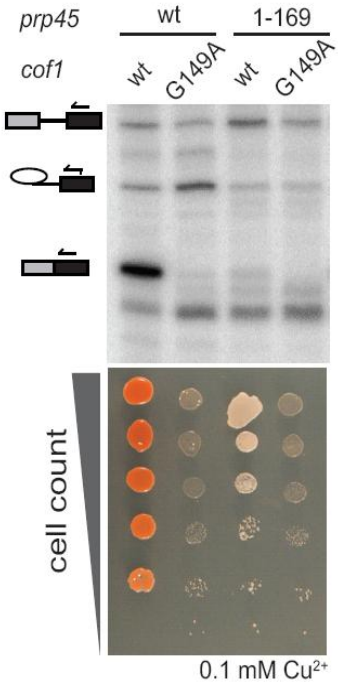


Fig. 3.17. *CUP1*-fusion splicing reporter genes revealed the inefficient *COF1* splicing in *prp45*(1-169) cells. The splicing analysis of genes containing wild-type or G149A *COF1* intron in wild-type or *prp45*(1-169) cells was performed by primer extension as described in Fig. 3.7 (upper panel). Wild-type and prp45(1-169) strains with knock-out *CUP1* gene were transformed by indicated reporter construct, grown to mid-log phase, harvested, concentrated to OD 4, spotted in the 8-fold dilution serie on minimal media with 0.1 mM $CuSO_4$ and cultivated in 30°C (lower panel).

In contrast, the splicing efficiency of *COF1* intron is significantly decreased in *prp45*(1-169) cells. The signal of mRNA is negligible in comparison to the wild-type strain (Fig. 3.17). The mutation G149A in the *COF1* intron which is synthetically lethal with *prp45*(1-169) further exacerbated this defect; *cof1*(G149A) mRNA is undetectable in *prp45*(1-169) cells. As the G149A *COF1* variant is poorly spliced already in wild-type cells (this phenotype will be analyzed in detail in the following part of this work), the synthetic lethality of *prp45*(1-169) and *cof1*(G149A) might be explained as a sum of two independent negative effects. The splicing defect of *COF1* intron in *prp45*(1-169) cells can be suppressed by the co-expression of the C-terminal part Prp45(119-379) *in trans* (Fig. 3.18). Taken together, the experiments with *CUP1* fusion reporters suggest that Prp45 is required for the efficient splicing of a specific subset of genes, which is represented here by *COF1*. The feature of *COF1* which leads to the *prp45* sensitivity is a subject of our present investigation.

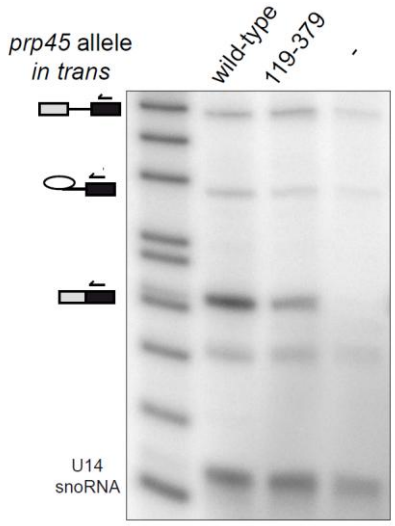


Fig. 3.18. Expression of C-terminal portion of Prp45 facilitates efficient splicing of *COF1* intron in *prp45*(1-169) cells. *prp45*(1-169) cells were transformed by the plasmid expressing *prp45*(119-379) or complete *PRP45* gene, or by an empty vector as a negative control. The splicing analysis of the *COF1-CUP1* reporter genes was performed as described in Fig. 3.7.

3.2 Prp45 – evolutionary conservancy of SNW/SKIP proteins function

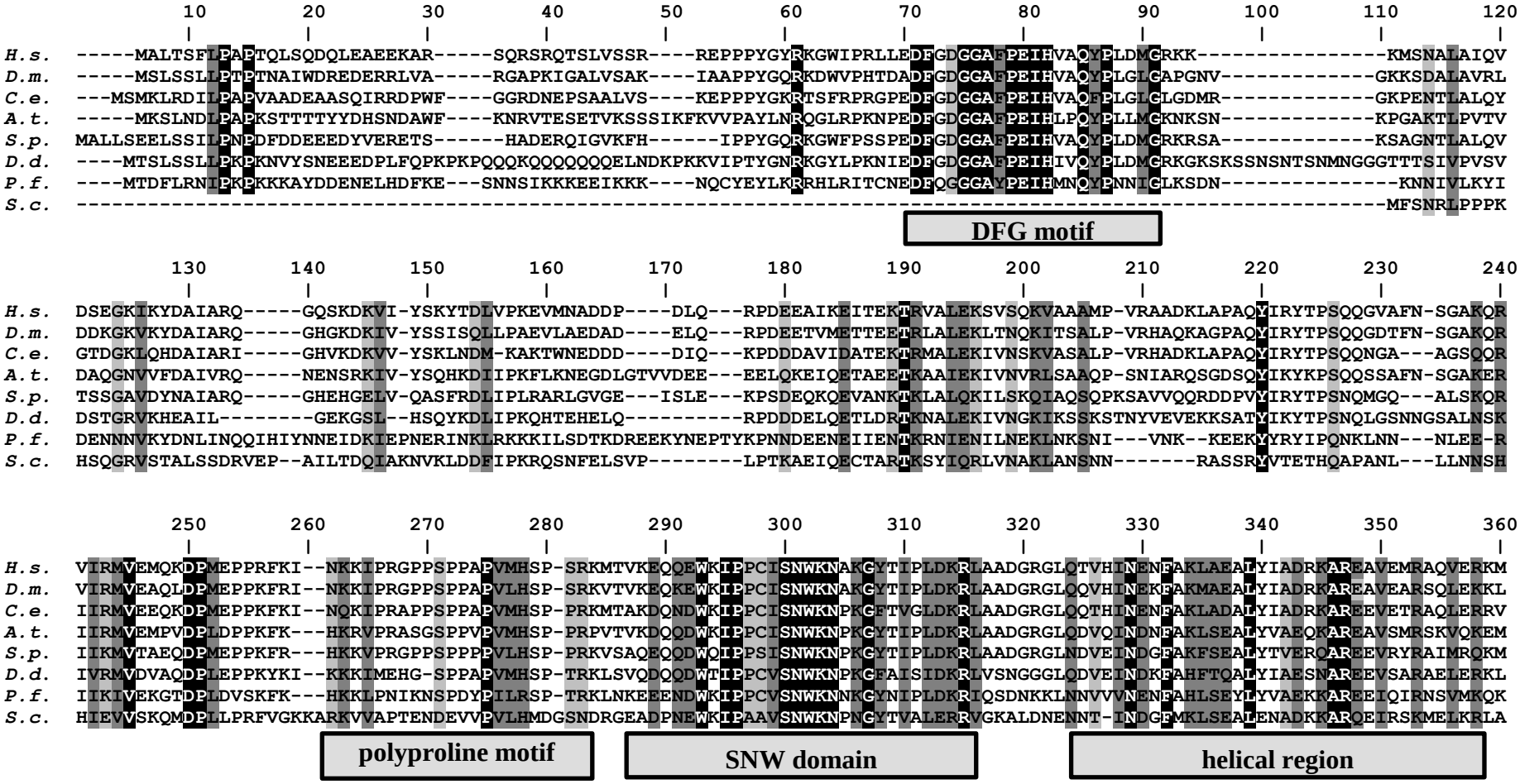
(related to publication 3, Wang et al., submitted for publication)

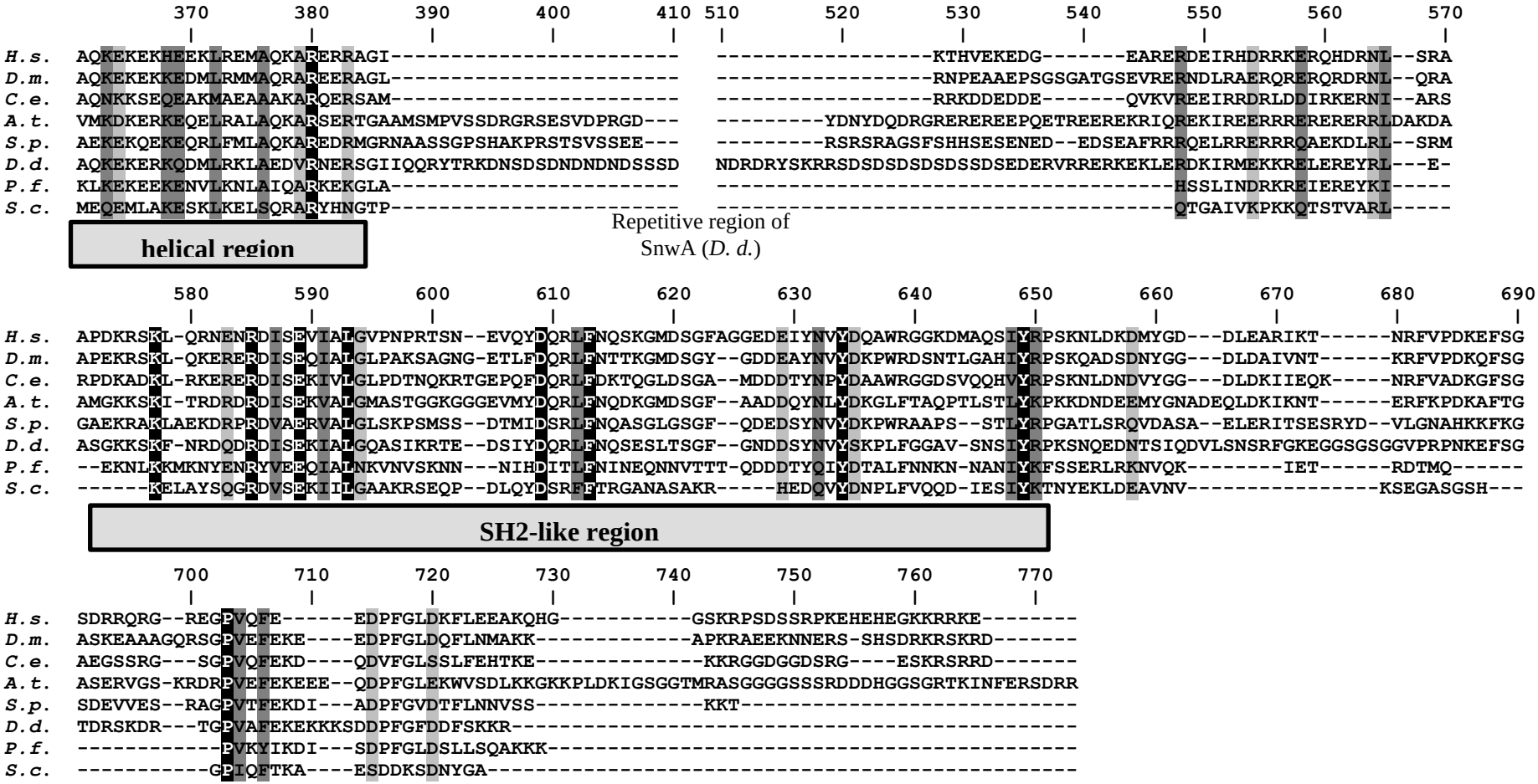
3.2.1 Prp45 homologs do not complement lethal phenotype of *prp45*-Δ cells

The *PRP45* gene is essential in *Saccharomyces cerevisiae*. To assess the evolutionary conservancy of SNW proteins' function, we asked whether the Prp45 orthologs from several species would complement the viability of the strain which does not express Prp45. We tested SNW1 from *Schizosaccharomyces pombe*, human SKIP, and AtSKIP from *Arabidopsis thaliana* (the sequence alignment including also representatives from other species is in Fig. 3.19). The *prp45-Δ* strain expressing Prp45 from an *URA3* plasmid was transformed by the construct encoding SNW1, SKIP, AtSKIP, *PRP45* (as a positive control), or by an empty vector. Cells were spread on the 5-FOA containing media. Only the *PRP45* harboring plasmid supported the growth of cells (Fig. 3.20A). Neither of homologs was thus able to substitute for the Prp45 function in *Saccharomyces cerevisiae*.

In addition, we performed the tetrad analysis of diploid *PRP45/prp45:kanMX4* strain transformed with the AtSKIP expressing plasmid. The analysis confirmed that AtSKIP is not able to support the growth of *prp45-Δ* cells (Fig. 3.20B).

Figure 3.19 (following two pages). Comparison of primary structures of SNW proteins from selected species. *H.s., Homo sapiens* SKIP/NCoA-62; *D.m., Drosophila melanogaster* Bx42; *C.e., Caenorhabditis elegans* CeSKIP; *A.t., Arabidopsis thaliana* AtSKIP; *S.p., Schizosaccharomyces pombe* Snw1; *D.d., Dictyostelium discoideum* SnwA; *P.f., Plasmodium falciparum*; *S.c., Saccharomyces cerevisiae* Prp45. Intensity of shading represents the conservancy of specific amino acid residues. Conserved motifs or expected domains are highlighted by gray boxes. The alignment was generated by ClustalW. Adapted from Skružný, 2003, PhD Thesis.


H.s.
D.m.
C.e.
A.t.
S.p.
D.d.
P.f.
S.c.
DFG motif
polyproline motif
SNW domain
helical region


helical region
Repetitive region of SnwA (D. d.)
SH2-like region

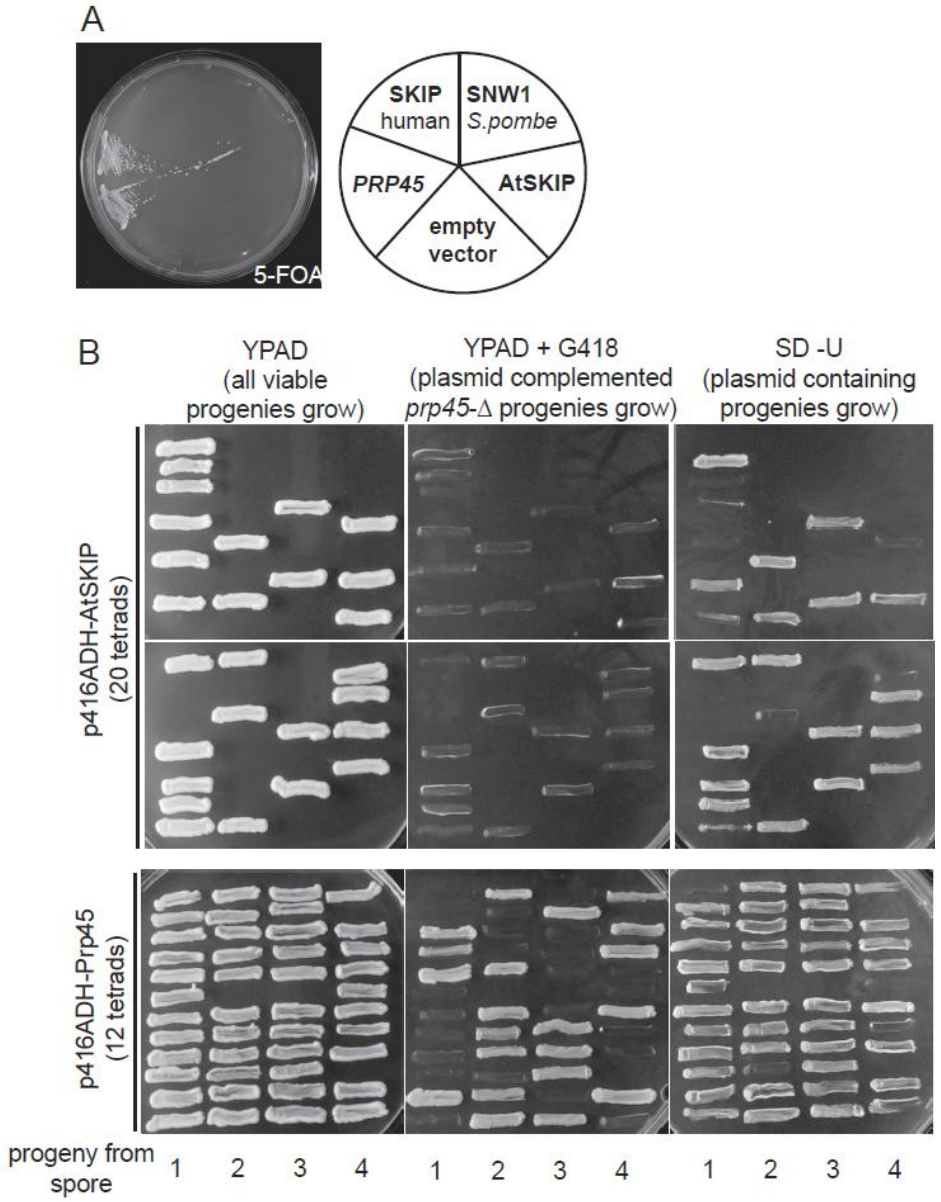


Figure 3.20. The *prp45-Δ* inviability cannot be complemented by the expression of *PRP45* orthologs. (A) *prp45-Δ* cells expressing *PRP45* from the *URA3* plasmid were transformed by a *HIS3* plasmid encoding indicated *PRP45* homolog and were spread on the 5-FOA containing media and cultivated for 2 days at 30°C. Only the positive control *PRP45* expressing *HIS3* plasmid was able to support the growth. **(B)** Tetrade analysis confirmed that AtSKIP does not complement the *prp45-Δ* inviability. Diploid *PRP45*/*prp45-Δ::kanMX4* strain was transformed by the AtSKIP (upper panel) or PRP45 (lower panel) expressing plasmid. After sporulation, spores were dissected and individual viable progenies were tested for the growth on YPAD, YPAD supplemented by G418 (selection of *prp45-Δ::kanMX4* progenies), or on the SD –Ura media (selection of plasmid containing progenies).

3.2.2 Temperature sensitivity of *prp45*(1-169) strain is suppressed by expression of some of its orthologs

We asked whether the expression of *PRP45* orthologs would overcome the temperature sensitivity of *prp45*(1-169) cells. We transformed the *prp45*(1-169) strain KAY02 by plasmids expressing SNW1, SKIP, or AtSKIP and compared their growth on solid media at 30°C and 37°C. Interestingly, AtSKIP from the most phylogenetically distant species *A. thaliana* was completely able to suppress the temperature sensitivity, while human SKIP had only intermediate positive effect, and SNW1 from *S. pombe* supported the growth only negligibly (Fig. 3.21).

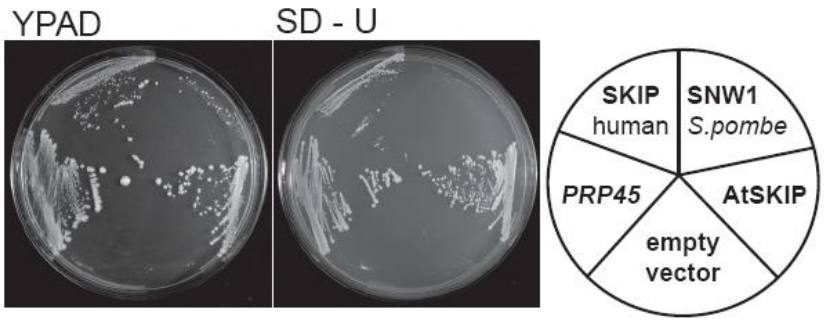


Figure 3.21. Temperature sensitivity of *prp45(1-169)* cells is suppressed by the expression of *PRP45* human (SKIP) and *Arabidospis thaliana* (AtSKIP) homologs. *prp45(1-169)* cells expressing indicated *PRP45* orthologs from the centromeric plasmid under the control of the *ADH1* promoter were spread on YPAD or minimal SD –Ura media and cultivated at 37°C for 5 days.

3.2.3 Splicing defects of *prp45*(1-169) can be partially suppressed by the expression of AtSKIP

AtSKIP and to some extend also human SKIP restored the growth of *prp45*(1-169) cells at an elevated temperature. So far we have not observed any relationship between the sensitivity to the higher temperature and the splicing defects of *prp45*(1-169) strain. We tested whether *PRP45* orthologs would repair the *prp45*(1-169) splicing phenotypes. AtSKIP and human SKIP partially restored the splicing of *prp45* sensitive *COF1* intron (Fig. 3.22). Moreover, the expression of AtSKIP increased also the splicing efficiencies of the suboptimal substrates mutated in the 5′ss, 3′ss and BP (Fig. 5D in publication 3, Wang et al., submitted). In contrast, SNW1 from *S. pombe* did not positively affect the *prp45*(1-169) associated splicing phenotypes. The cells expressing human SKIP exhibited only mild

restoration of splicing. These results suggest that at least in *Arabidopsis thaliana*, function of AtSKIP in pre-mRNA splicing is conserved.

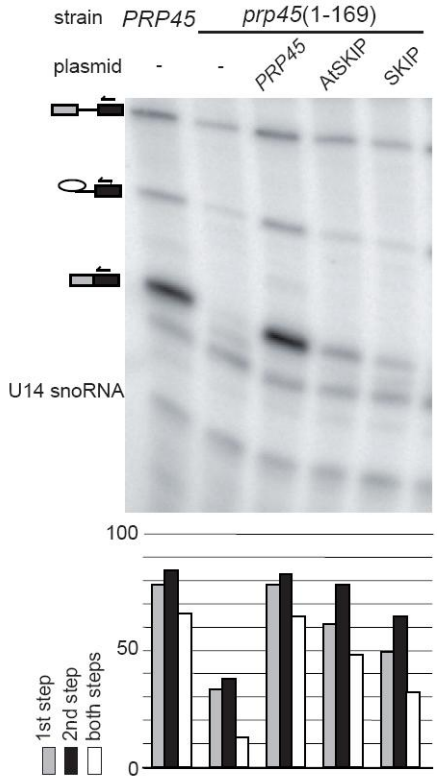


Fig. 3.22. The expression of the *PRP45* homologs partially suppresses the splicing defect of *COF1* intron in *prp45*(1-169) cells. The splicing analysis of the *COF1* intron containing reporter gene in indicated strains expressing indicated *PRP45* orthologs was performed by primer extension as described in Fig. 3.7. For the construction of the bar diagrams, we used the phosphorimager signals to quantify the efficiencies of the first (M+LI)/(P+M+LI), second step M/(M+LI), and overall efficiency M/P+M+LI). M – mRNA; P – pre-mRNA; LI – lariat intron splicing intermediate.

3.3 Role of pre-mRNA secondary structure in 3´splice site recogniton

(related to publication 2, Gahura et al., *in press*)

3.3.1 Single nucleotide G149A substitution in *COF1* intron impairs its splicing

In a screen aimed to search for the synthetic lethal interactions of *prp45*(1-169) (see section 3.1.2 Genetic interactions of *prp45*(1-169)), we identified a single nucleotide G to A substitution in the position 149 of *COF1* intron (referred to as G149A; all intron modifications in this study are numbered according to the nucleotide position in the non-manipulated intron). The G149A mutation affects *COF1* pre-mRNA as well as mRNA levels in cells, as documented by quantitative RT-PCR. The approximately 6-fold increase of pre-mRNA and 2-fold decrease of mRNA as compared to the wild-type cells (Fig. 3.24) suggest a defect in splicing. The G149A substitution occurs 102 nt downstream of the BP and 30 nt upstream of the 3´ss (Fig. 3.25A). *COF1* belongs to the subset of genes in budding yeast with

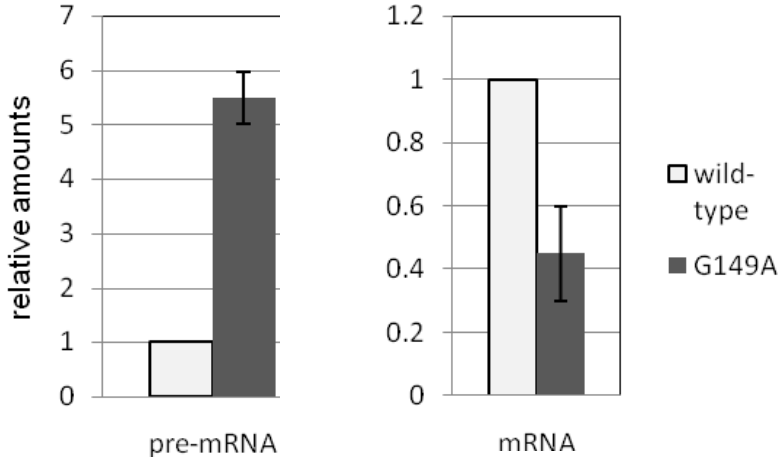


Figure 3.24. The G149A substitution in *COF1* intron results in the accumulation of *COF1* pre-mRNA and the decrease of spliced mRNA levels. Amounts of *COF1* pre-mRNA and mRNA in wild-type and *cof1*(G149A) cells were quantified by qRT-PCR. Data were normalized to the reference gene *TOM22* and related to wild type. Error bars represent standard deviations from two biological replicates.

exceptionally long distance between BP and 3´ss (132 nt; Fig. 3.25B). Notably, the *ACT1* intron with the artificially extended distance between the BP and 3´ss to more than 60 nt is not spliced, supposedly due to difficulties in the 3´ss recognition (Cellini et al., 1986). Thus, *COF1* and other similar introns are expected to employ

so far unknown mechanism facilitating the identification of remote 3´ss and enabling efficient splicing.

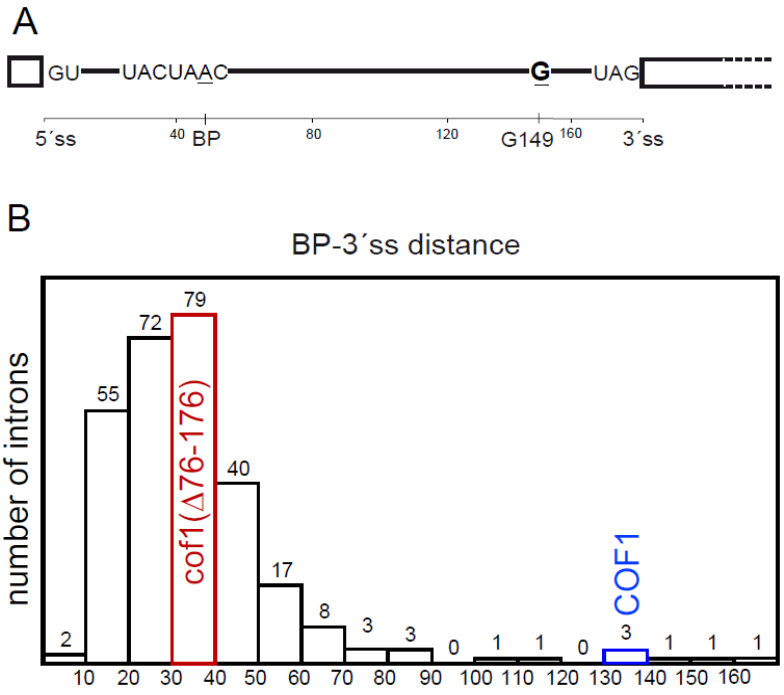


Figure 3.25. The intron of *COF1* gene has an extremely long distance between the BP and 3´ss. (A) The schematic representation of *COF1* intron. The numbering starts at the first intron nucleotide. 5´ss, branch point, 3´ss, and G149 nucleotide are depicted. **(B)** The distribution of BP-3´ss distances in the introns of *Saccharomyces cerevisiae*. The numbers of introns in each category are indicated above respective bars. The categories that include *COF1* and *cof1*(Δ76-176) introns are depicted in blue and red, respectively.

3.3.2 Secondary structure between BP and 3´splice site is required for *COF1* efficient splicing

The secondary structure of *RPS17B* intron, which is above average in length of sequence upstream of BP, was shown to decrease a "structural distance" between the 5´ss and BP and thereby improve its splicing efficiency (Rogic et al., 2008). Therefore, we decided to employ RNA secondary structure predicting tools and compare the models of the wild-type and G149A *COF1* introns. For predictions, we used several available algorithms in parallel, including physics-based types, RNAfold (Hofacker, 2003; Gruber et al., 2008), RNAshapes (Steffen et al., 2006) or Mfold (Zuker, 2003), as well as knowledge-based type (MC-Fold;

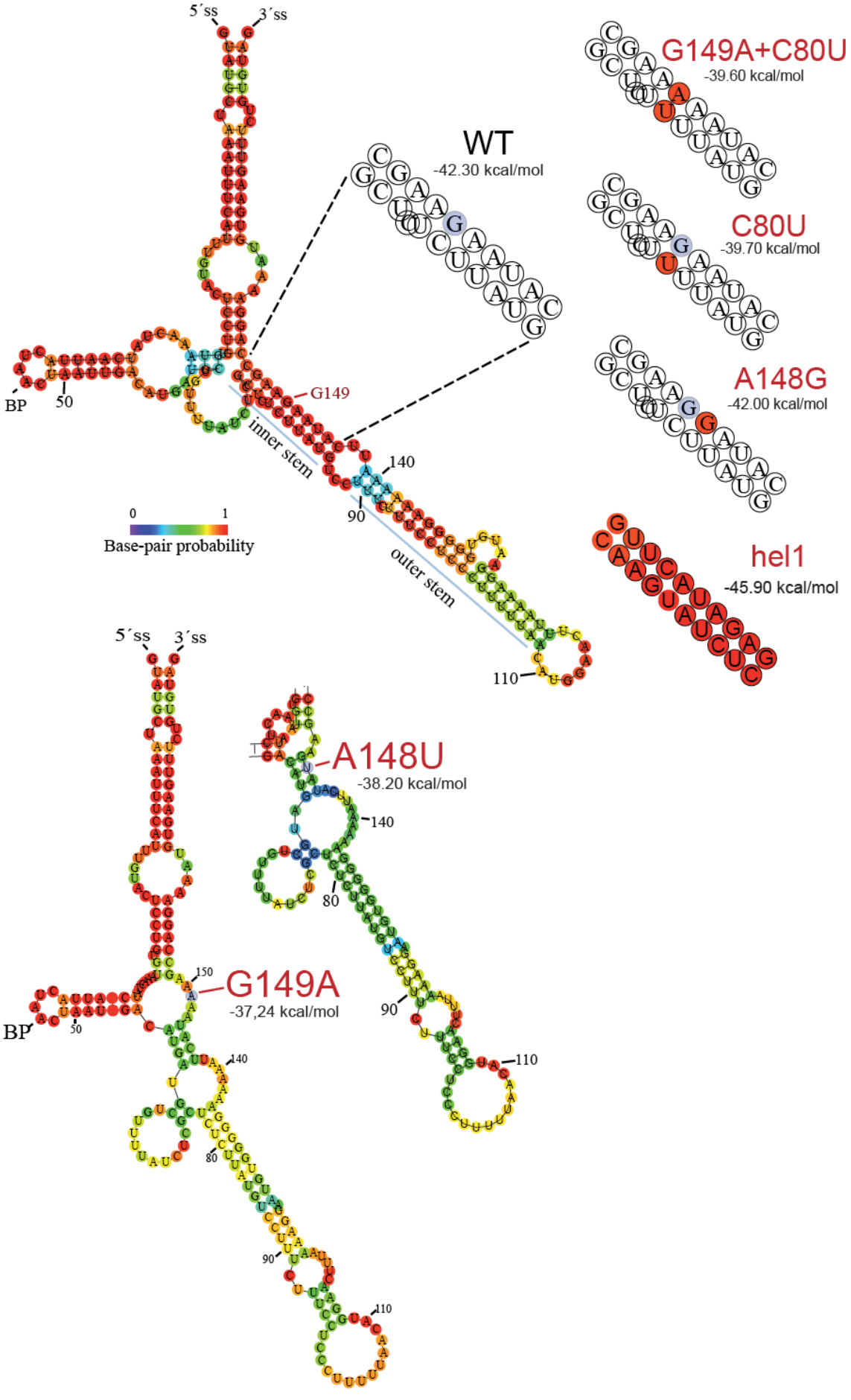

5´ss
3´ss
BP
50
G149
inner stem
outer stem
140
90
110
0
1
Base-pair probability
WT
-42.30 kcal/mol
G149A+C80U
-39.60 kcal/mol
C80U
-39.70 kcal/mol
A148G
-42.00 kcal/mol
hel1
-45.90 kcal/mol
A148U
-38.20 kcal/mol
G149A
-37,24 kcal/mol
150
80

Figure 3.26. (previous page) The secondary structure predictions of *COF1* intron and its mutated variants. The RNAfold algorithm was employed to generate the minimal free energy models of indicated *COF1* intron variants. Complete intron sequences from the 5´ss to 3´ss were used. The values of free energies of respective structures are shown. For variants with the same structure prediction as wild-type, only the inner stems with highlighted nucleotide alterations (red) are depicted. For the A148U variant, only the fragment differing from the G149A model is depicted. The G149 nucleotide is depicted in blue.

Parisien and Major, 2008). All algorithms showed qualitatively very similar or even equivalent results for the vast majority of input sequences we tested.

The minimal free energy model (MFE; the most probable thermodynamic model) generated by RNAfold suggests the formation of a hairpin stem-loop structure between the BP and 3´ss in the wild-type *COF1* intron. The long stem is divided by an internal loop to the "inner" and "outer" parts. Upon the introduction of the G149A substitution into the input sequence, the predicted stem was destabilized and the base pairing was altered (Fig. 3.26). All algorithms employed predicted the same stem loop structure in the wild-type intron. Importantly, any extension of the input sequence on the 5´- and/or 3´-end did not affect the occurrence of this structure. For example, see output of the RNAshapes algorithm when applied on wild-type or G149A intron sequence downstream of the BP (Fig. 3.27). Based on these observations, we decided to study differences in the splicing of wild-type and G149A *COF1* intron in more detail.

We expressed *COF1-CUP1* reporter genes in the wild-type strain EGY48 to assay their splicing by primer extension. The pre-mRNA containing wild-type *COF1* intron was spliced efficiently, whereas the G149A mutation caused severe splicing defect resulting in undetectable quantities of the spliced mRNA (Fig. 3.28, lanes 1 and 2). The G149A mutation is apparently predicted to destabilize the base pairing within the inner stem. We asked whether the negative impact of G149A on the splicing would be suppressed by a potentially compensatory C to U substitution in the position base-pairing with the G149 nucleotide in the wild-type model. Indeed, the predicted structure of the G149A+C80U double mutant of *COF1* intron does not qualitatively differ from the non-mutated variant prediction (Fig. 3.26) and the G149A splicing defect can be suppressed by the C80U mutation (Fig. 3.28A, lane 3). The C80U substitution itself did not affect the prediction (Fig. 3.26) neither the splicing (Fig. 3.28A, lane 4), likely because the predicted G-U pair (substituting for

the wild-type G-C pair) is stable enough to keep the base pairing of the inner stem. In an effort to disrupt the structure by an independent mutation, we manipulated the A148 nucleotide adjacent to G149. The A148U (Figs. 3.26 and 3.28A) and A148C (data not shown) substitutions had negative effect on both predicted structure stability and splicing, similarly to G149A, whereas the A148G resulting in the alternative G-U pairing had no effect (Fig. 3.28A, lane 6). In all cases, the wild-type resembling structure prediction corresponded to the efficient splicing and vice versa. These observations suggest that the folding into the stem structure is required for the efficient splicing.

To definitely exclude the importance of the inner stem sequence for the efficient splicing of *COF1*, we randomized all nucleotides of both inner stem strands in the way that the stability of the modeled structure remained unaffected, generating the variant referred to as *COF1*-hel (Figs. 3.26 and 3.28B). Although the sequence was extensively altered, the splicing of such intron occurred with the wild-type efficiency, as shown by primer extension (Fig. 3.28B left panel). The efficient splicing was demonstrated also by the growth of *cup1-Δ* cells expressing the *COF1*-hel reporter on the Cu^{2+}-containing solid media (Fig. 3.28B right panel), documenting the unchanged resistance as compared to the wild-type *COF1* intron. Taken together, we conclude that secondary structure downstream of BP rather than a sequence specific interaction is required for the efficient splicing of *COF1*, presumably by decreasing the distance between the BP and 3´ss.

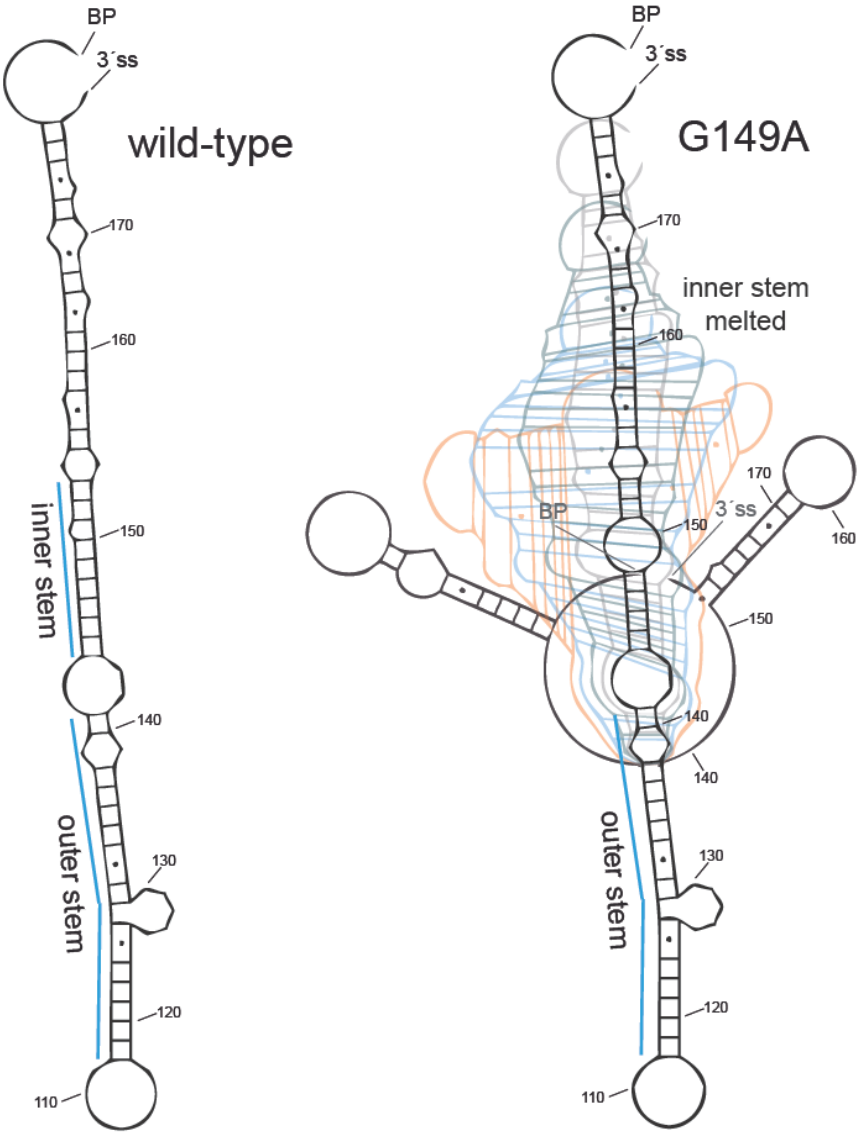


Figure 3.27. The models of wild-type and G149A *COF1* introns generated by RNAShapes. The sequences from the BP adenosine to the 3´ss were used. The dynamic structure of G149A variant is depicted as a set of overlaid still images extracted from an animated output of the program.

3.3.3 RNA probing confirmed *COF1* intron secondary structure

The occurrence of the inner stem in the wild-type *COF1* intron and its destabilization by the G149A substitution was confirmed *in vitro*. Following *COF1* intron variants were subjected to *in vitro* in line probing: wild-type *COF1*, *cof1*(G149A), *cof1*(G149A+C80U; see Fig. 2C in Gahura et al., in press). The experiment was performed by Christian Hammann from Technical University in Darmstadt, Germany.

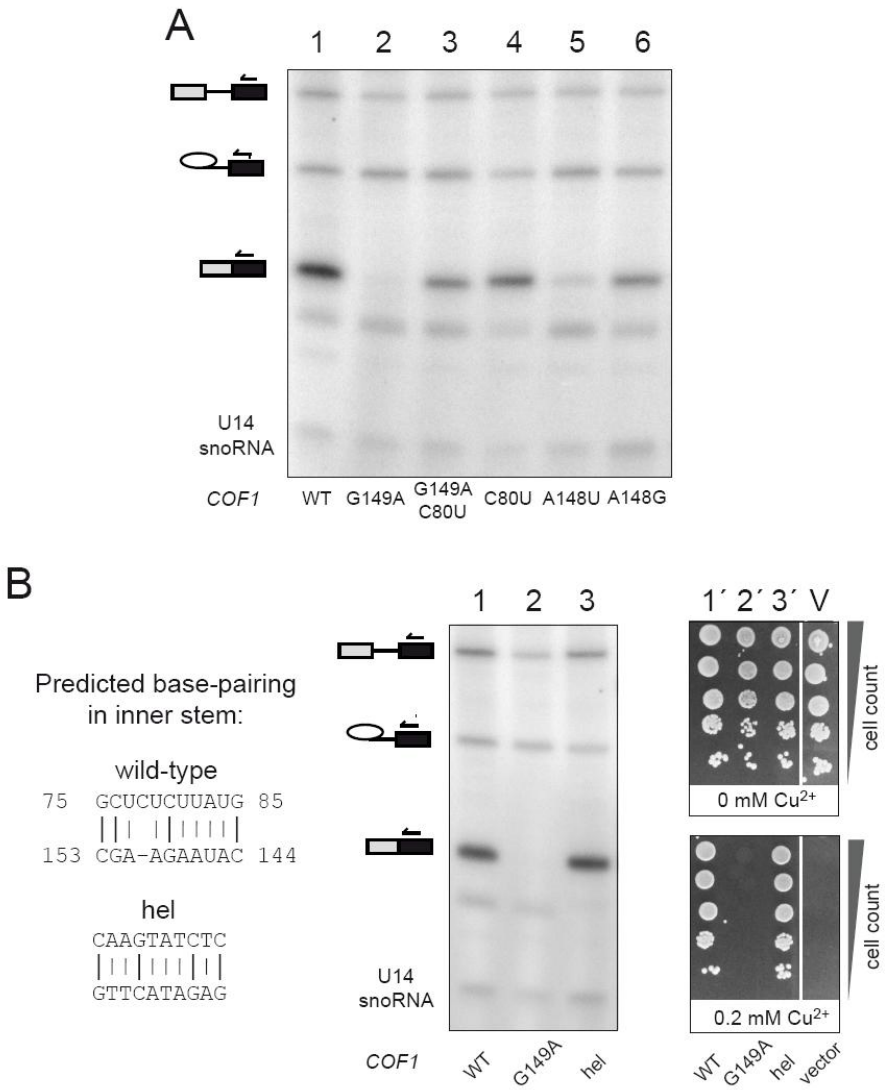


Figure 3.28. Intramolecular structure between BP and 3´ss is required for the efficient splicing of *COF1* intron. (A) RNA from cells expressing indicated *COF1-CUP1* constructs was subjected to primer extension analysis as described in Fig. 3.7. U14 snoRNA was assayed as a loading control. **(B)** Intramolecular stem is indispensable for the efficient splicing of *COF1* intron. We completely altered the sequence of the inner stem of *COF1* intron (left panel) such that the structure's predicted stability remained the same. The "hel" variant was spliced with the same efficiency as wild-type as shown by primer extension (lane 1 and 3) and by Cu^{2+}-resistance of cells expressing the reporter construct (lane 1´ and 3´).

3.3.4 Length of second exon does not affect splicing *COF1* intron

Although pre-mRNA splicing is generally considered to occur cotranscriptionally, in *S. cerevisiae* it has not been reliably proved. Clearly, the spliceosome assembly is started on nascent RNA just upon the synthesis of 5´ss, as shown by chromatin immunoprecipitation with snRNP components (Kotovic et al., 2003; Gornemann et al., 2005; Lacadie and Rosbash, 2005; Tardiff and Rosbash, 2006). The length of

second exon was proposed to be crucial in determining the character of splicing (Tardiff and Rosbash, 2006). Long second exons (> 800 nt) provide splicing machinery enough time to remove intron before transcription is finished. In contrast, later stages of spliceosome assembly and splicing reaction itself occur posttranscriptionally on transcripts with the short second exons. Recently, two independent studies documented that a pausing of the RNA polymerase II (RNApol II) on intron containing genes may facilitate cotranscriptional splicing, independently on the second exon length (Alexander et al., 2010; Oesterreich et al., 2010; commented in Oesterreich et al., 2011).

The *COF1-CUP1* reporter gene contains a short second exon (208 nt + 3´UTR) and is therefore presumably spliced posttranscriptionally. An interaction of spliceosome with the RNApol II complex and chromatin may affect the 3´ss recognition as well as the folding of introns into secondary structure. We tested the impact of the second exon length on the splicing of wild-type and G149A *COF1* intron variants. Substituting *CUP1* sequence in reporter constructs by the fragments of *HFM1* gene exon 2, we generated a set of reporters with the size of exon 2 varying from 100 to 2000 nt; first 108 nt of exon 2 are common for all *HFM1-CUP1* constructs (Fig. 3.29A; for sequences, see Appendix A2). Primer extension analysis revealed that neither the wild-type nor G149A splicing efficiency was affected (Fig. 3.29B). Thus, there is no effect of the second exon length on the splicing efficiency of *COF1* intron with the unstable stem loop downstream of the BP. This suggests that hypothetical switch from post-transcriptional to co-transcriptional splicing does not affect the efficiency of the structure-mediated 3´ss recognition.

3.3.5 Cryptic 3´splice sites in *COF1* intron are masked by secondary structure

To answer the question, whether the sequence folded into the identified stem loop structure in the *COF1* intron is dispensable for its splicing, we generated a set of *COF1* intron truncations (Fig. 3.30A) and tested their splicing efficiency using *CUP1* fusion reporters. Deletion of the nucleotides 91 to 140 (*cof1*-Δ91-140), which are predicted to form the outer part of the stem loop, did not detectably affect the splicing efficiency (Fig. 3.30B, lane 1 and 2). However, the introduction of structure-destabilizing G149A mutation into the sequence ended up with a dramatic decrease of mRNA signal and with appearance of an additional band (Fig. 3.30B, lane 3). Using the 5´-RACE (rapid amplification of cDNA ends) technique we confirmed that

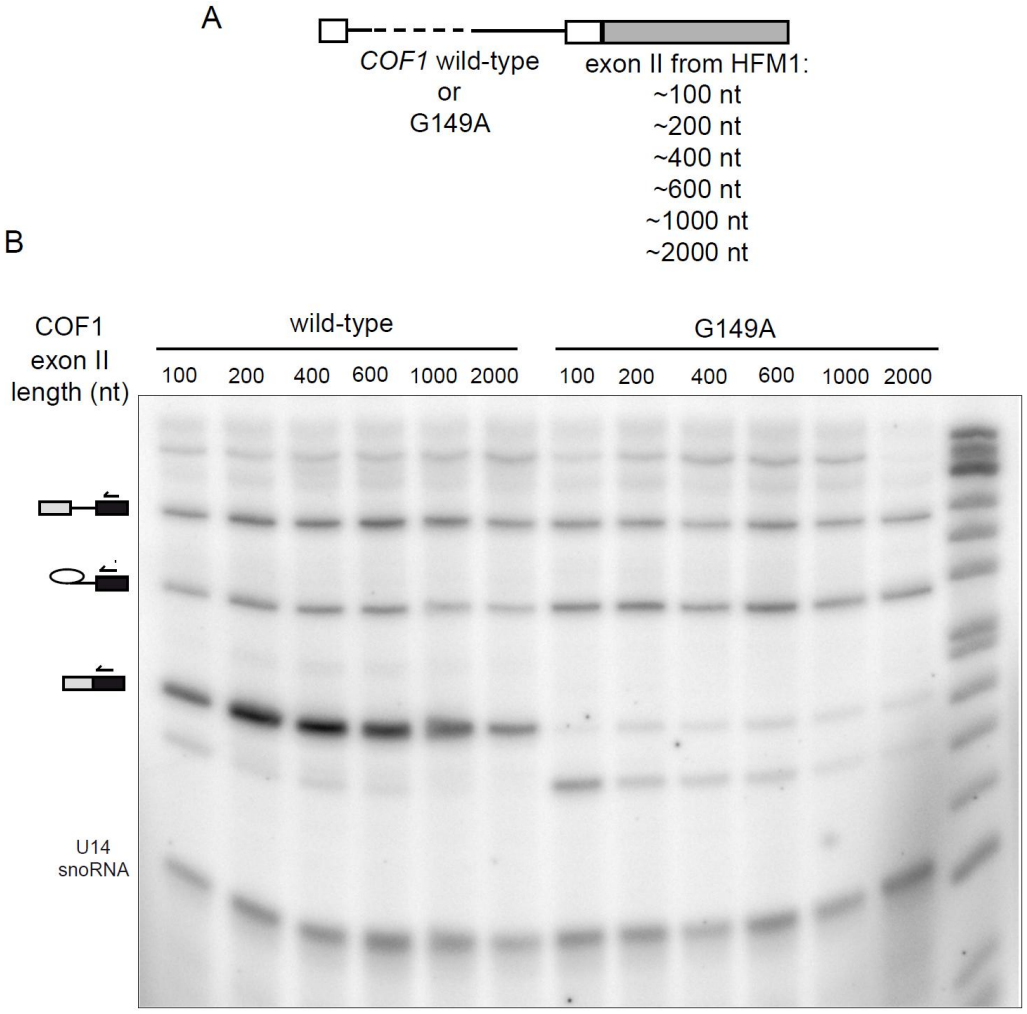


Figure 3.29. The second exon length does not affect the splicing of *COF1* intron. (A) The Schematic representation of the *COF1-HFM1* constructs. The fragment of *COF1* gene including the wild-type or G149A intron was fused to the *HFM1* second exon fragments of indicated lengths. **(B)** The splicing primer extension assay of indicated *COF1-HFM1* reporter genes with indicated *COF1* intron variant was performed as described in Fig. 3.7.

the band corresponds to an RNA molecule spliced to a cryptic 3´ss (AAG-152) located 27 nt upstream of the annotated 3´ss. Thus, the BP-proximal NAG trinucleotide was recognized as the 3´ss in the truncated and destabilized *COF1* intron. As expected, the C80U mutation stabilizing the G149-containing stem partially restored the use of the canonical 3´ss (Fig. 3.30B, lane 4). We then deleted additional nucleotides to obtain the truncated variant lacking the complete stem loop structure having 56 nt distance between the BP and 3´ss (*cof1*-Δ76-152). In accordance with the models, which do not predict the formation of any stable structure, this variant is spliced to the CAG-156, the first potential 3´ss downstream of the BP. (Fig. 3.30B, lane 5). The variant *cof1*-Δ76-176 with the BP-3´ss distance

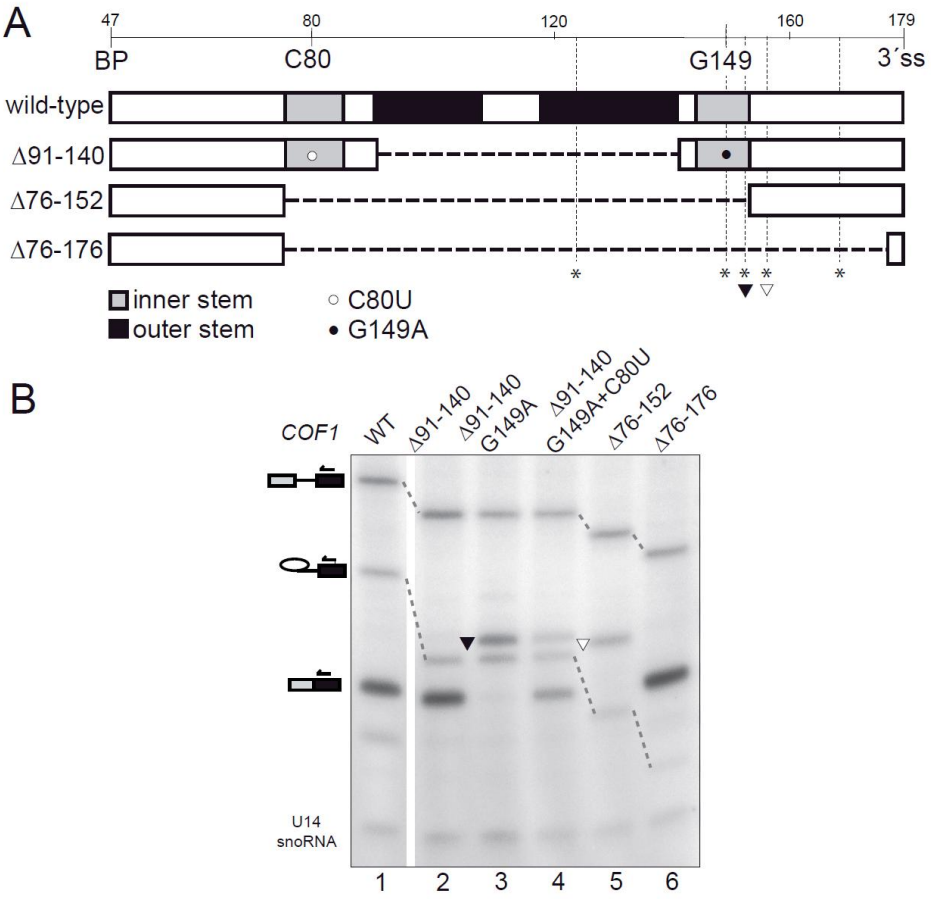


Figure 3.30. The destabilization of intramolecular stem within the *COF1* intron unmasks potential 3´ splice sites. (A) The summary of truncated *COF1* intron constructs. The sequences forming the inner and outer stem are shaded; deleted regions are represented by dashed lines. The positions of G149 and C80 are marked by empty and filled circle, respectively. Potential 3´ss (all A/C/UAG trinucleotides) are marked by asterisks; acceptor sites used in the mutants, AAG152 and CAG156, are indicated by filled and empty arrowhead, respectively. **(B)** The destabilization or removal of stem within the *COF1* intron proves the structure's role in presenting the appropriate portion of the pre-mRNA molecule to spliceosome for 3´ss recognition. Filled and empty arrowheads indicate bands corresponding to the cryptic splice sites marked in (A). Notably, the cryptic sites of the *COF1* intron mutants were only used when they were not blocked within the secondary structure or when they were brought closer to BP through deletion.

of 31 nt (corresponding approximately to *S. cerevisiae* intron median; see Fig. 3.25B) is spliced as efficiently as the non-truncated *COF1* intron, generating the mRNA of regular size (Fig. 3.30B, lane 6). We conclude that the sequence forming the stem-loop structure per se is not required for the *COF1* splicing. Instead, the secondary structure formed between the BP and 3´ss of *COF1* intron masks the sequences, which might otherwise serve as the acceptor sites for the second transesterification, and thereby enables the proper 3´ss recognition.

3.3.6 Yeast introns with distant 3´ss are predicted to fold into similar structure as *COF1*

COF1 is not the unique intron in *S. cerevisiae* in terms of the BP-downstream region length (Fig. 3.25B). We modeled the structures of all yeast introns with the distance between BP and 3´ss longer than 50 nt. The vast majority of these introns are predicted to fold into a structure resembling the long stem-loop in the *COF1* intron. We categorized the structures according the proportions of base paired nucleotides and corrected categorization upon the visual inspections (Table 3.2; examples of structures and summary is shown in Fig. 3.31). This analysis suggested that most of *S. cerevisiae* introns with remote 3´ss might employ the same mechanism for 3´ss recognition which was identified for *COF1*.

A

BP -3´ss distance	≥ 40 nt		≥ 50 nt	
Category	Number	%	Number	%
A	48	56	28	72
B	20	24	8	20
C	12	14	3	8
D	5	6	0	0
Total	85	100	39	100

B

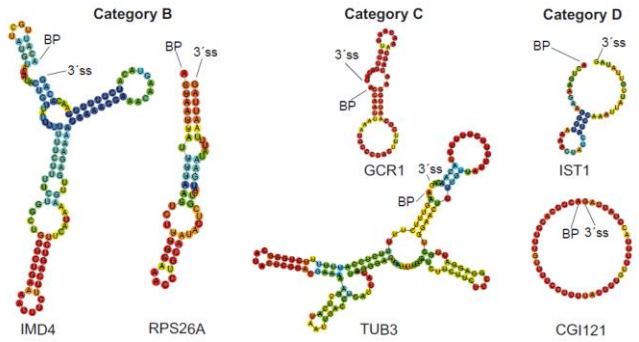


Figure 3.31. The majority of dBP introns of *S. cerevisiae* are predicted to form a secondary structure between BP and 3´ss. (A) The sequences between BP and 3´ss were analyzed by RNAfold. Predicted structures were sorted into four categories. A - stems with high probability of base pairing; none to small bulge/internal loops. B - stems with medium-to-high probability of base pairing; bulge/internal loops. C - other type of secondary structure. D - unstable structures or unstructured. The relative proportions of the categories are shown. Full list of introns is provided in Tab. 3.2. **(B)** Two examples of predicted structures in each category are depicted in the right panel. Category A is represented by *COF1* and *UBC13* (Fig. 3.26 and 3.32).

Table 3.2 (next page). List of introns categorized according to the type of the secondary structure predicted between BP and 3´ss. Introns within a category are sorted by the branch point 3´ss distance (column „BP-3´ss"). A – structures with extensive and stable stems; none to small bulge/internal loops. B - structures with less extensive but still stable stems; bulge/internal loops. C - other type of secondary structure. D - unstable structures or unstructured. We grouped the RNAfold-predicted structures into four arbitrary categories based on the stem_length/BP_to_3'ss_distance ratios (column „stem/length") and visual inspection. We put the structures in category A if the ratio =/> 0,125 (at least 1/4 of the BP-3'ss region was base paired; *RPL20A* and *TEF4* were included in

categories B and A, respectively, based on visual inspection). We counted base pairs as stems when the base pairing probability was above 0.5 for more than four consecutive bases and their position was >10bp from either BP or 3'ss.

A

ORF name	Gene	BP-3´ss	stem/ length
YML017W	PSP2	166	0,211
YDR092W	UBC13	155	0,213
YDR381C-A		145	0,228
YLL050C	COF1	132	0,303
YHR039C-A	VMA10	105	0,162
YGL251C	HFM1	83	0,193
YJL205C-A	NCE101	82	0,232
YJR145C	RPS4A	82	0,232
YOR122C	PFY1	72	0,208
YLR306W	UBC12	70	0,229
YOR293W	RPS10A	69	0,130
YGR148C	RPL24B	65	0,154
YKL006C-A	SFT1	64	0,219
YNL138W-A	YSF3	64	0,203
YAL030W	SNC1	62	0,177
YPR132W	RPS23B	60	0,250
YNL147W	LSM7	58	0,224
YDL130W	RPP1B	58	0,155
YML024W	RPS17A	58	0,155
YIL133C	RPL16A	57	0,140
YIL069C	RPS24B	57	0,246
YFL034C-B	MOB2	56	0,214
YLR061W	RPL22A	55	0,145
YJL189W	RPL39	55	0,182
YHR021C	RPS27B	55	0,127
YER056C-A	RPL34A	53	0,132
YDR064W	RPS13	53	0,170
YHR077C	NMD2	49	0,143
YCR031C	RPS14A	49	0,163
YKL081W	TEF4	49	0,122
YBR191W	RPL21A	48	0,208
YKL006W	RPL14A	48	0,188
YDL191W	RPL35A	48	0,146
YHL001W	RPL14B	47	0,213
YGL226C-A	OST5	46	0,174
YLR185W	RPL37A	46	0,152
YHR141C	RPL42B	46	0,174
YEL012W	UBC8	46	0,130
YER044C-A	MEI4	44	0,205
YGL031C	RPL24A	44	0,136
YGL103W	RPL28	43	0,140
YNL112W	DBP2	43	0,140
YOL127W	RPL25	41	0,195
YDL083C	RPS16B	41	0,146
YHR203C	RPS4B	41	0,171
YNL301C	RPL18B	40	0,150
YML026C	RPS18B	40	0,150
YOR182C	RPS30B	40	0,125

B

ORF name	Gene	BP-3´ss	stem/ length
YER014C-A	BUD25	131	0,053
YML056C	IMD4	113	0,071
YGL137W	SEC27	71	0,085
YKL186C	MTR2	68	0,088
YPL079W	RPL21B	62	0,113
YKL157W	APE2	55	0,109
YGL189C	RPS26A	55	0,091
YHR097C		54	0,074
YER117W	RPL23B	52	0,096
YMR242C	RPL20A	47	0,149
YLR329W	REC102	46	0,109
YPR153W		45	0,089
YBL087C	RPL23A	44	0,114
YFL039C	ACT1	43	0,093
YDL115C	IWR1	42	0,095
YER074W	RPS24A	42	0,071
YMR125W	STO1	42	0,071
YPL031C	PHO85	41	0,098
YBL027W	RPL19B	41	0,098
YGL030W	RPL30	40	0,075

C

ORF name	Gene	BP-3´ss
YML124C	TUB3	139
YKR094C	RPL40B	79
YPL075W	GCR1	51
YOR234C	RPL33B	49
YDR450W	RPS18A	45
YPR028W	YOP1	45
YER074W-A	YOS1	43
YJR079W		43
YGL187C	COX4	42
YGL087C	MMS2	42
YML085C	TUB1	41
YHR016C	YSC84	40

D

ORF name	Gene	BP-3´ss
YFR024C-A	LSB3	47
YJL041W	NSP1	43
YML036W	CGI121	42
YNL265C	IST1	40
YPL109C		40

3.3.7 *UBC13* pre-mRNA folds into a structure facilitating its splicing

To further support the evidence that splicing of *S. cerevisiae* introns with distant 3´ss depend on a secondary structure between BP and 3´ss, we designed several *CUP1*-fused splicing reporters containing the intron of *UBC13* or its derivatives. The *UBC13* intron has the second longest sequence between the BP and 3´ss (155 nt) in *S. cerevisiae*. We introduced several mutations into the BP-3´ss region, which dramatically disordered the predicted secondary structure ("disordered" variant, Fig. 3.32). The disordered *UBC13* intron was spliced to the cryptic CAG 3´ss located 38 nt downstream of the BP (5´-RACE confirmed); this site is apparently masked by the secondary structure in the wild-type intron (Fig. 3.32B, lane 1 and 2). When the base pairing (but not the original sequence) in the predicted stem was restored through the set of complementary mutations (Fig 3.32A), wild-type splicing pattern was restored (Fig 3.32B, lane 3).

The point mutation G232A, which caused only mild decrease in the stability of the predicted structure at 30°C (data not shown), resulted in the splicing to the multiple 3´ss, including the cryptic as well as the canonical one (Fig 3.33 lane 3). However, at 39°C the cryptic 3´ss were preferentially employed (Fig. 3.33, lane 4). This finding is in agreement with the fact that the melting of folded RNA increases with the temperature, supporting the hypothesis that the stem structure is responsible for the proper 3'ss selection. The compensating A148T mutation suppressed aberrant 3´ss selection also in the temperature-dependent manner (Fig. 3.33, lanes 5 and 6). We reason that *UBC13* and likely other *S. cerevisiae* introns fold into secondary structure in order to overcome the constraints given by the long BP-3´ss distance and to mask sequences, which could be incorrectly interpreted as alternative aberrant 3´splice sites.

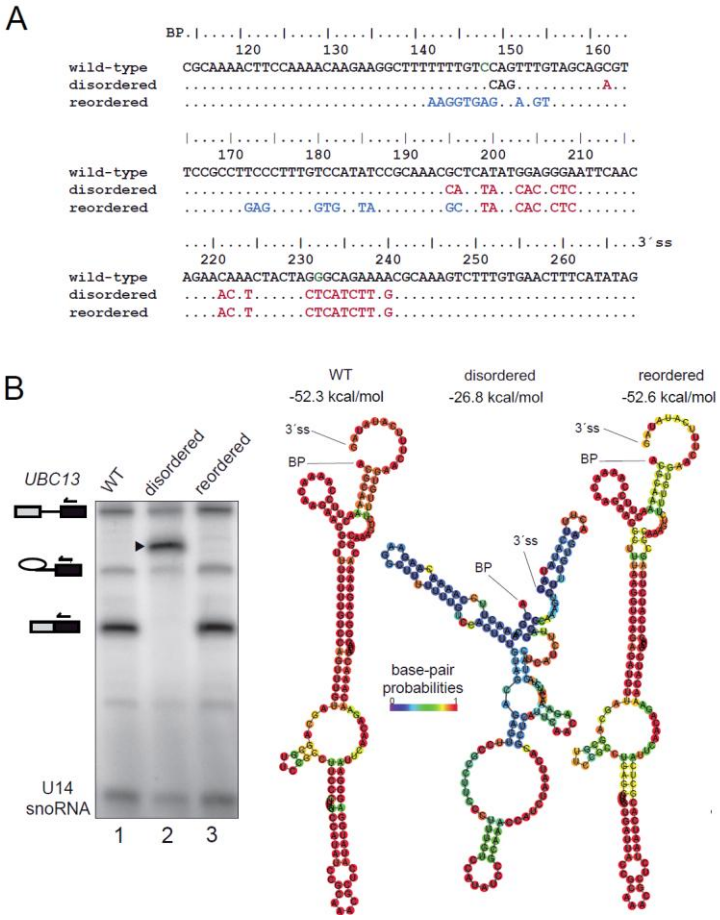


Figure 3.32. Secondary structure of *UBC13* intron is responsible for proper 3'ss selection. (A) A summary of wild-type and manipulated UBC13 sequences. Intron sequences are shown from the BP adenosine. Mutations introduced to generate "disordered" variant are shown in red font; mutations introduced into "disordered" variant to obtain "reordered" variant are in blue. G232 and C148 are in green. **(B)** Mutations of one arm of the predicted stem in *UBC13* intron, which destabilized the structure, resulted in splicing to the cryptic CAG located 38 nt downstream of the BP (lane 2; band is marked by arrowhead). Introducing mutations to the complementary arm, such that the pairing energy but not the original sequence was restored, repaired the defect (lane 3). RNAfold based models of BP to 3´ss regions in wild-type and manipulated *UBC13* intron variants are shown in the right panel. The values of free energy are indicated for each structure.

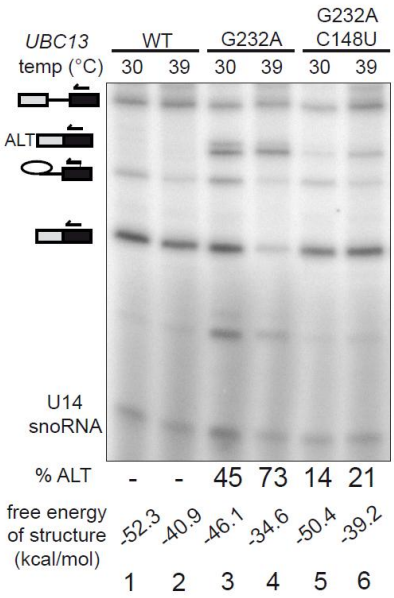


Figure 3.33. Use of cryptic 3´ss in a G232A *UBC13* point mutant with partially destabilized structure is enhanced at higher temperature. G232A substitution in *UBC13* intron mildly destabilized the secondary structure and led to the use of several 3´ss (lane 3; bands corresponding to mRNA spliced to cryptic 3´ss are marked by ALT + mRNA icon). The defect was exacerbated at 39°C (lane 4). Compensating substitution partially suppressed the defect in temperature dependent manner (G232A+C148U; lanes 5 and 6). Densitometric quantification of alternatively spliced RNA (expressed as percent of total spliced RNA; % ALT) and free energy of RNAfold predicted MFE structure calculated for a given temperature are indicated below each lane.

3.3.8 *COF1* and *UBC13* intron structures are conserved in *Saccharomyces sensu stricto* genus

To assess whether the structure based mechanism facilitating distant 3´ss recognition can be conserved in evolution we modeled the structures of *COF1* and other introns in several yeast species. RNAfold predicted very similar secondary structures downstream of the BP in *COF1* introns from six species representing *Saccharomyces* "*sensu stricto*" genus (*S. cerevisiae, S. paradoxus, S. kudriavzevii, S. mikatae, S. bayanus,* and *S. pastorianus*; Fig 3.34). Output from RNAalifold (Gruber et al., 2008), an algorithm predicting consensus secondary structures from the alignments of several related RNA sequences, confirmed conservancy in the base-paired regions of predicted models (Fig. 3.35). Dissimilarities between species in double-strand forming sequences in most cases do not impact base-pairings. Changes in one strand either preserve the base-pairing with the other strand (e.g. A-U pair is altered by G-U pair), or are accompanied by the co-evolution in opposite strand (e.g. G-C pair is

replaced by A-U). Thus, there seems to be a selection against the mutations destabilizing the secondary structure. The analysis of primary and predicted secondary structure between the BP and 3´ss of *UBC13* intron shows also the conservancy among all *Saccharomyces* species listed above (data not shown).

Neither the *COF1* and *UBC13* sequences nor their structures seem to be conserved in other *Saccharomycotina* yeasts. However, all introns with remote 3´ss from this taxon tested are predicted to fold into a stable structure downstream of BP. This is in contrast with the human dBP introns in HTR4, PTB1, GABBRI, or tropomyosin genes, which are predicted to be completely unstructured in this region. Evolutionary aspects of 3´ss recognition and intron architecture will be further focused in Discussion.

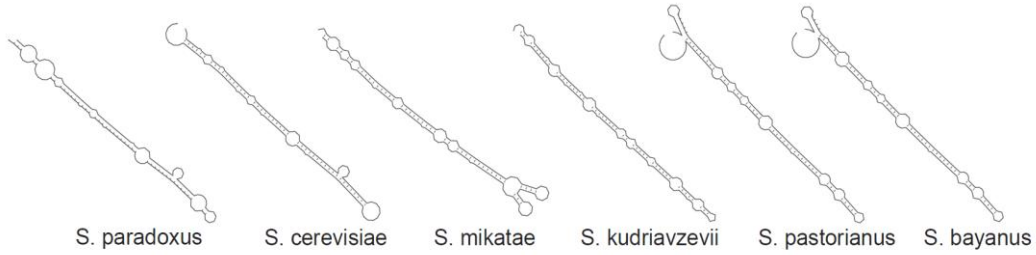


Figure 3.34. Models of secondary structures of *COF1* introns from *Saccharomyces sensu stricto* species. RNAfold was employed to generate *COF1* intron secondary structure predictions downstream of BP for indicated species.

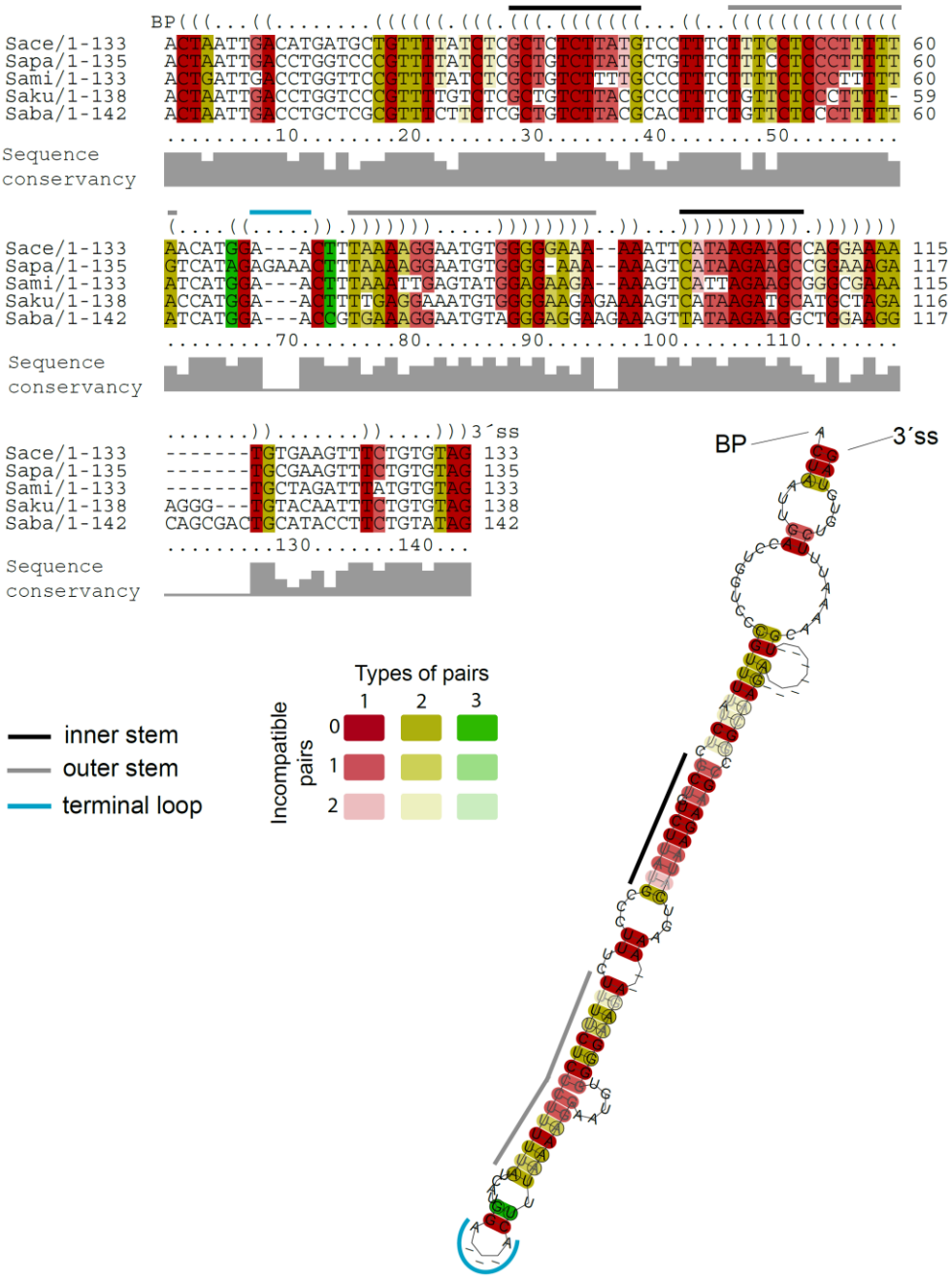


Figure 3.35. The structure of BP-3´ss region of *COF1* intron is conserved in *Saccharomyces* „sensu stricto“ genus. The structure annotated alignment and consensus secondary structure of *COF1* intron BP-3´ss region from five *Saccharomyces* species (Sace – *Saccharomyces cerevisiae*, Sapa – *S. paradoxus*, Sami – *S. mikatae*, Saku – *S. kudriavzevii*, Saba – *S. bayanus*) were generated by the RNAalifold algorithm. Red, yellowish and green types of pairs indicate that one, two, or three types of nucleotide pairs, respectively, are found at a given position of the aligned sequences. Incompatible pairs refer to the number of sequences in which nucleotide does not base pair at a given position (Gruber et al., 2008).

DISCUSSION

4.1 Role of Prp45 in splicing

Prp45 was implicated in splicing based on: (i) its co-purification with spliceosomal complexes (Ohi et al., 2002), (ii) its two hybrid interaction with splicing factors Prp46, Prp22 and Syf3, (iii) the observation that the cells with downregulated *PRP45* expression accumulate unspliced precursors (Albers et al., 2003), and (iv) its homology with human SNW/SKIP, which is a component of active spliceosome (Bessonov et al., 2008b) and it affects alternative splicing of specific transcripts (Bres et al., 2005). For details see section 1.10 Prp45.

We took advantage of the hypomorphic allele *prp45*(1-169) and performed an exhaustive search for its synthetic lethal interactors. Most of the genes wherein we had identified an SL mutation can be functionally divided into the two following groups. The first group encodes subunits of NTC complex (Clf1/Syf3, Syf1, Ntc20, Cef1; for details see section 1.11 NTC complex in Literature Review) while the other group encodes the second step splicing factors (Prp22, Prp18, Slu7 and Prp17). In addition, we identified an allele of *PRP8* and a single nucleotide mutation in the intron of *COF1* gene. All documented SL interactions thus strongly support the engagement of Prp45 in pre-mRNA splicing. Therefore, we employed *prp45*(1-169) in a series of experiments with the goal to elucidate its role in splicing in more detail. Results obtained will be discussed below.

4.1.1 Prp45 and NTC complex

As *PRP45* encodes one of the NTC-related proteins, its synthetic lethality with NTC members is not surprising. There is a network of both genetic and physical interactions among NTC components and NTC-related genes/proteins (Fig. 4.1; (Chen et al., 1998; Chen et al., 1999; Ben Yehuda et al., 2000a; Ohi and Gould, 2002). The Cef1 protein, which is a core component of the NTC, ensures the interaction of Prp19 tetramer with the rest of the complex through its contacts with almost all other subunits (Chen et al., 2001b; Chen et al., 2002). Syf1 and Ntc20 together with Isy1 and Syf2 form a subcomplex, which is compact even under conditions, when the NTC itself is destabilized. Clf1/Syf3 contains fifteen tetratricopeptid (TPR) repeats. TPRs serve as a scaffold for the binding of other

proteins. Although Clf1/Syf3 interacts with many NTC members, the importance of particular TPR repeats for specific interactions was not studied. The allele *clf1*(T402I,S404F), which is SL with *prp45*(1-169), contains two changed amino acids in TPR11. It localizes to the essential part of protein; the deletion of TPRs 10-15 is lethal (Vincent et al., 2003). This region was predicted to fold into a suprahelical structure serving as a platform for protein-protein interactions (Ben Yehuda et al., 2000a).

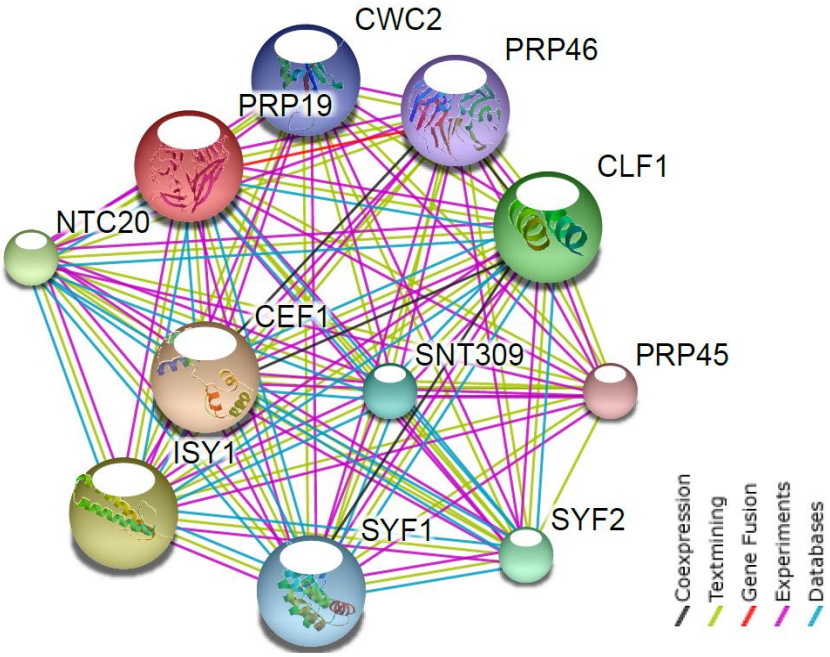


Fig. 4.1 Interaction network of NTC components and NTC related proteins. Each edge represent an interaction identified according the legend. Generated by STRING 9.0 Database (Szklarczyk et al., 2011).

Prp45 interacts in two hybrid screen with Clf1/Syf3 and Prp46, an essential NTC-related splicing factor with an unknown function. The Prp45-Clf1/Syf3 interaction seems to be mediated via an N-terminal part of Clf1/Syf3 (TPR1-8; (Albers et al., 2003). The mechanistic explanation of *prp45* genetic interactions with NTC members is not clear. As the environment of the NTC involves multiple interactions, it is possible that disruption of couple of them due to two mutations might be deleterious, possibly because of the overall destabilization of the complex. The NTC may facilitate Prp45 recruitment to the spliceosome, which could be compromised by the synthetic effect of a pair of mutations.

Our documentation of *prp45* genetic interactions with the second step factors and the previously published Prp45-Prp22 two hybrid interaction (Albers et al., 2003) suggest that Prp45 may be involved in later stages of splicing. Prp22, Prp17, Prp18 and Slu7 are recruited after the first step is finished and their association with spliceosome is facilitated by multiple mutual interactions (Jones et al., 1995; van Nues and Beggs, 2001; James et al., 2002). However, it is not clear which spliceosomal components mediate the first contact of the second step factors with the machinery. Hypothetically, subunits of NTC are possible candidates, as many genetic and physical interactions between NTC and second step factors were documented. Among others, genes *CLF1/SYF3, SYF1* and *SYF2* were initially identified as SL interactors of *PRP17/CDC40* deletion (*syf* - synthetic lethal with cdc forty; (Ben Yehuda et al., 2000b).

4.1.2 Prp45 aids to the regulation of Prp22 helicase

Using the *prp45*(1-169) allele, we documented the functional interaction of Prp45 with the Prp22 helicase. Biochemical analysis revealed significantly decreased levels of Prp22 in spliceosomal complexes isolated from *prp45*(1-169), while the stoichiometry of other proteins does not seem to be affected. The most straightforward explanation of this observation is obvious: the C-terminal part of Prp45 is required for the Prp22 recruitment to spliceosome. This hypothesis is supported by additional observations. Firstly, in a two hybrid system, the C-terminus of Prp45 is required for the Prp45-Prp22 interaction (Fig. 3.5). Secondly, two of *prp22* alleles SL with *prp45*(1-169) result in decreased amounts of Prp22 in cells.

Alternatively, the reduced Prp22 signal in the Cwc2-TAP pull-down from *prp45*(1-169) cells may reflect a shortened period for which Prp22 persists in contact with the splicing machinery. It is known that Prp22 associates with spliceosome only transiently and, after finishing its task, it is released (Schwer and Gross, 1998; Schneider and Schwer, 2001). *prp45*(1-169) might limit the duration of Prp22 residence in spliceosome, either by weakening mutual contacts required for a proper interaction, or by disregulating Prp22 function. Indeed, the *prp45*(1-169) dependent acceleration of Prp22 performance would be in agreement with the decreased splicing efficiency of suboptimal substrates in *prp45*(1-169) cells. Such substrates are spliced slowly and as a consequence are preferentially discarded by helicase based proofreading mechanisms (Mayas et al., 2006; Konarska et al., 2006). A higher

rate of the Prp22 mediated rearrangements would result in the increased rejection of such suboptimal slow substrates.

Prp22 overexpression does not suffice to suppress the *prp45*(1-169) associated splicing defects. We did not test Prp22 content in the Cwc2 pull downs under these conditions. However, the experiments with the co-expression of Prp45 termini *in trans* suggest the correlation between presence of Prp22 in the spliceosome and effective splicing. We thus reason that Prp22 overexpression would not result in its restored stoichiometry in the spliceosome. This would confirm the strict requirement of intact Prp45 for the proper Prp22 recruitment and/or functioning in the splicing machinery.

4.1.3 Prp45 is required for efficient splicing of specific transcripts

The results of the primer extension splicing assay using *ACT1-CUP1* reporter system with suboptimal splicing signals suggest that splicing might not be impaired in general in *prp45*(1-169) cells. Rather, splicing defects may be restricted to a subset of specific transcripts. Indeed, our experiments showed that only the *COF1-CUP1* fusion reporter gene is spliced poorly in *prp45*(1-169) cells, while other introns tested were excised with the wild-type efficiency, as shown by primer extension and copper sensitivity assay (Figs. 3.16 and 3.17). Intriguingly, the *RPL30* intron containing mutated 5´ss (A3C substitution) was spliced efficiently in the *prp45*(1-169) background, documenting that suboptimal splicing signals as such are not responsible for Prp45 sensitivity.

Data published here and also our additional preliminary observations clearly show differential behavior of the intron containing transcripts in *prp45*(1-169) cells. We are currently trying to identify the pre-mRNA hallmark(s) that are responsible for splicing defects resulting in decreased amounts of respective spliced mRNA in the *prp45*(1-169) mutant. We already documented that a truncation of extremely long BP – 3´ss distance in *COF1* intron did not result in the restoration of *COF1-CUP1* splicing efficiency (data not shown). Thus, the demand of Prp45 for *COF1* splicing is not caused by the occurrence of remote 3´ss. We plan to test additional features of *COF1* intron architecture to resolve which determinant is the cause of Prp45 dependency.

4.1.4 Prp45 – only second step splicing factor?

The functional interaction of Prp45 with Prp22 and the splicing defects of the 3´ss mutated substrates in *prp45*(1-169) cells suggest an engagement of Prp45 in the second step of splicing. However, in contrast to the second step factors (Prp22, Slu7, Prp18, Prp17) which contact the spliceosome only after the first step is finished (Jones et al., 1995; James et al., 2002), Prp45 enters into the splicing process before the first transesterification. Impairment of substrates with suboptimal 5´ss and BP in *prp45*(1-169) also suggest the first step related role of Prp45. However, as splicing reaction and the reorganization of catalytic core could be reversible (Tseng and Cheng, 2008), it is possible that the observed first step defects are only a consequence of the improper functioning of the Prp45 in the second step. On the other hand, the genetic interaction of *PRP45* with *PRP8* encoding one of the core splicing factors (Grainger and Beggs, 2005) might suggest an employment of Prp45 in both steps. Indeed, our recent unpublished results documenting the accumulation of unspliced pre-mRNA molecules in *prp45*(1-169) cells indicate that Prp45 also affects the first splicing reaction.

4.2 Function of SNW proteins – *Saccharomyces cerevisiae* versus higher eukaryotes

Human SNW/SKIP protein acts in multiple events in gene expression (for details see section 1.9 SNW proteins). It may be employed either autonomously in distinct nuclear processes, or it may serve to couple transcription initiation or early elongation with posttranscriptional processing.

In *Saccharomyces cerevisiae*, the function of Prp45 in splicing is clearly documented. In contrast, there are no reports on the participation of Prp45 in transcription. However, the genome-wide expression analysis in *prp45*(1-169) cells indicates that Prp45 can be involved also in the regulation of intronless genes (K. Abrhamova, PhD Thesis). Indeed, we and other labs have recently obtained data indirectly supporting the Prp45 role in transcription elongation. On the contrary, some functions of human SNW/SKIP cannot be applicable to Prp45, as *S. cerevisiae* lacks several mechanisms in which SNW/SKIP was implicated. It refers mainly to the SNW/SKIP co-regulated transcriptional responses to the signaling pathways that are specific for multicellular organisms. Differences in SNW proteins' amino acid sequences may also reflect functional divergency. Prp45 lacks approximately 80 N-terminal amino acid residues, in comparison to human SNW/SKIP or other homologs. In addition, some sequential motifs present in higher eukaryotes are not preserved in Prp45. For example, the polyproline motif, which contains six prolines in higher eukaryotes, has only two prolines in *S. cerevisiae* (Fig. 3.19).

Previously, SNW/SKIP was reported to complement the deletion of *PRP45* in *Saccharomyces cerevisiae* (Figueroa and Hayman, 2004). However, our results are in strong contradiction with this observation, as we repeatedly documented that SNW/SKIP does not support the growth of *prp45*-Δ cells. Moreover, neither SNW1 from *Schizosaccharomyces pombe* nor AtSKIP from *Arabidopsis thaliana*, nor SNWA from *Dictyostelium discoideum* were able to complement Prp45 essential function (Fig. 3.20 and data not shown). The inability to complement the *PRP45* deletion might come out from incompatible sequential differences or from non-conserved functioning. The inability to complement is likely not due to the low abundance of the protein(s) as they have been expressed from high copy number plasmids and from strong promoters. Sufficient levels of SNWA protein from

D. discoideum expressed in *S. cerevisiae* were confirmed by a specific antibody (František Půta; data not shown).

Interestingly, the temperature sensitive phenotype of *prp45*(1-169) cells can be suppressed by the expression of AtSKIP (*A. thaliana*) or human SNW/SKIP, while SNW1 (*S. pombe*) or SNWA from *Dictyostelium discoideum* did not support growth at 37°C (Fig. 3.21). Our experiments monitoring splicing efficiency of reporter substrates or multiple genomic transcripts suggest that *prp45*(1-169) related splicing defects are not exacerbated at a higher temperature. Therefore, we hypothesize, that *prp45*(1-169) temperature sensitivity is rather a consequence of a defect in a different process.

The co-expression of the Prp45 C-terminal part clearly restores the defective splicing of reporter genes in *prp45*(1-169) cells to wild-type levels (Figs 3.14C and 3.18). Inefficient splicing of *CUP1* fusion substrates is partially improved by the expression of AtSKIP, and to some extent also by human SNW/SKIP and SNW1 from *S. pombe* (Fig. 3.22 and Fig. 5D in Wang et al., submitted). This suggests that the functioning of SNW proteins in splicing is conserved across eukaryotes, although sequential differences do not support a complete rescue of the phenotype. The report of Wang et al. further documents the negative impact of AtSKIP alleles on splice site selection, which leads via disregulation of clock genes' splicing to an impairment of circadian biorhythms in *A. thaliana*.

4.3 Intron secondary structure-mediated 3´ss recognition

We have shown that the formation of a secondary structure between the BS and 3´ss is indispensable for proper 3´ss recognition and as a consequence is required for efficient splicing of *COF1* and *UBC13* introns in *Saccharomyces cerevisiae*. The folding of introns into the stem-loop structure decreases the distance between the two splicing signals and the stable base pairing masks the potential cryptic 3´splice sites. Outputs from the RNA secondary structure predicting tools strongly suggest that similar structures can facilitate 3´ss identification in a vast majority of *S. cerevisiae* introns with remote 3´ss. Importantly, at the same time when our study was published (Gahura et al., September 2011), the J. Vilardell group from Barcelona reported exactly the same phenomenon (Meyer et al., 2011). They used the same approaches (primer extension analysis of *CUP1* fusion reporter substrates accompanied by intron mutagenesis) to document the importance of secondary structures for the 3´ss recognition in four different dBP genes (*RPS23B*, *VMA10*, *APE2*, *REC102*). Thus, we can very reliably conclude that the structure-mediated mechanism of 3´ss recognition is common for considerably large subset of yeast introns. Nevertheless, there are several interesting questions which emerged with our findings and which will be discussed below.

4.3.1 How does a decreased distance between BP and 3´ss contribute to the 3´splice site recognition?

The most straightforward explanation for the need of secondary structure downstream of BP in introns with distant 3´ss is related to the spatial range of interactions during 3´ss recognition. Regardless the splicing stage, in which the 3´ss is definitely localized, the spliceosome or its part responsible for contacts with 3´ss has certain limits in physical distance which can be reached. Distant 3´ss may lie beyond these limits and the secondary structure brings the 3´ss sequence closer in order to make it available to the spliceosome. Although this explanation may appear plausible, it must be considered, that the intron structure is not rigid. Rather, the stem represents the most probable state of gradually changing conformations. In addition, it can be further affected by interactions with other associated molecules.

An alternative explanation is based on the kinetics of reaction. Remote 3´ss might require more time to be found and dBP introns might therefore represent

suboptimal "slow" substrates. Such substrates are discarded from spliceosomes by mechanisms ensured by RNA helicases. This kinetic proofreading occurs in all stages of splicing (Query and Konarska, 2006; Mayas et al., 2006; Mayas et al., 2010; Koodathingal et al., 2010).

The second presented hypothesis is in agreement with the observation that the splicing efficiency of the *ACT1* intron gradually decreases with the progressive extension of sequence downstream of the BP (Cellini et al., 1986). A kinetic model can be tested using the mutants of helicases (e.g. *prp16*-K379A, *prp22*-K512A or others) which are defective in the proofreading activity and which were shown to enable the splicing of slow or suboptimal substrates that are rejected under wild-type conditions (Mayas et al., 2006; Koodathingal et al., 2010).

4.3.2 May secondary structures potentiate alternative splicing in *S. cerevisiae*?

We showed that the sequence forming the stem loop structure in the *COF1* intron can be deleted without any detectable impact on the splicing of *COF1-CUP1* reporter gene. The same observation was documented for intron *VMA10* (Meyer et al., 2011). This begs the question: why were these structures with no apparent benefits for splicing not eliminated during evolution? The analyses suggest a conservancy of secondary structures in most dBP introns in *Saccharomyces sensu stricto* genus (this work and Meyer et al., 2011), which represents 15 – 20 million years of evolution (Scannell et al., 2007). It is tempting to speculate that the destabilization of structures might contribute to the regulation of gene expression at an elevated temperature via decreasing splicing efficiency of dBP introns. In an extreme situation, the unmasking of a cryptic 3´ss due to the melting of the structure might lead to an alternative splicing pattern. Indeed, the experiments of Vilardell and his coworkers revealed that the proportion of *APE2* mRNA spliced to the cryptic 3´ss is higher when cells are cultivated at 37°C (Meyer et al., 2011). They proposed an intron secondary structure-based thermosensor regulating gene expression. However, they provided neither biological significance nor proved any physiological role for *APE2* alternative splicing. Although the *APE2* expression might be regulated via the temperature dependent alteration of the 3´ss selection, we do not think the same mechanism would be applicable for other genes having the secondary structure based 3´ss recognition. There are at least two indirect hints contradicting this hypothesis. Firstly, although the secondary structures between the BP and 3´ss seem to be

conserved across the *Saccharomyces* genus, the positions of branch point proximal HAG trinucleotides, which are expected to be employed as acceptor sites upon secondary structure melting, are diverse. For most dBP introns, splicing to the cryptic 3´ss would give rise to a protein coding mRNA in some species, while in others it would result in the generation of premature stop codons. Secondly, the RNA-seq analysis of yeast transcriptome did not reveal an altered usage of 3´ss upon heat shock, with the exception of *APE2* gene (Yassour et al., 2009). Moreover, dBP-intron containing genes do not exhibit clustering to any category according to Gene Ontology annotations, which also rather opposes the hypothesis of common mechanism of their regulation. Taken together, we reason that the secondary structure is not exploited by the spliceosome to directly regulate the quality of splicing outcomes. However, we cannot exclude a role in a fine regulation of splicing efficiency which can turn out to be crucial only under some specific conditions. Thus, the secondary structures between BP and 3´ss may resemble yeast introns as such, because a systematic deletion approach revealed, that vast majority of introns in *S. cerevisiae* are dispensable for growth under normal conditions. Only upon specific stress do certain genes require introns for precise regulation of their mRNA levels (Parenteau et al., 2008).

4.3.3 Evolutionary conservancy of structure mediated 3´ss recognition

As secondary structures mediate 3´ss recognition in the majority of dBP introns in *S. cerevisiae*, it would be interesting to know whether this mechanism is conserved in other species. Modeling algorithms proposed that dBP introns from *S. bayanus, S. mikatae, S. kudriavzevii,* and *S. paradoxus* fold into similar structures downstream of BP as in *S. cerevisiae* (see Figs. 3.34 and 3.35). Moreover, other intron features, including their lengths or positions in coding sequences, are also conserved among all *Saccharomyces* species, supporting the high conservancy of splicing processes in these organisms. Therefore, it proves quite convincing that all species from *Saccharomyces sensu stricto* genus employ the same mechanism for the recognition of remote 3´ss.

The situation becomes more complicated when taking in account other yeasts from the *Saccharomycotina* group (Hemiascomycetes; Suh et al., 2006). Although all species from this taxon belong to “intron poor” organisms, the frequency and the distribution of introns in genomes are variable. For example, *UBC13* gene is

intronless in most *Saccharomycotina* species. Interestingly, a published bioinformatic analysis revealed that distributions of BP-3´ss distances are also very divergent within the group (Irimia and Roy, 2008). Some species, including *S. cerevisiae*, have "intronome" with large variability of BP-3´ss distances ("unconstrained" species), while other species possess only introns with 3´ss located closely to the BP ("constrained" species; Fig. 4.2). The available data do not allow us to decide which pattern is original and which is evolutionary derived as the analysis was performed only for a limited number of species. Unconstrained species rather appear to be interspersed over the whole *Saccharomycotina* phylogenetic tree. Regardless the phylogenetic relationships, some species are able to exploit the general ability of RNA to fold into secondary structures to localize distant 3´ss, while the others from some reason do not take the advantage of this RNA feature. Hypothetically, an unknown protein factor and its "chaperon" activity, which is not present in all species, might contribute to the secondary structure folding or stabilization. However, at present we have no data indicating an involvement of any splicing factor in the structure-mediated 3´ss recognition.

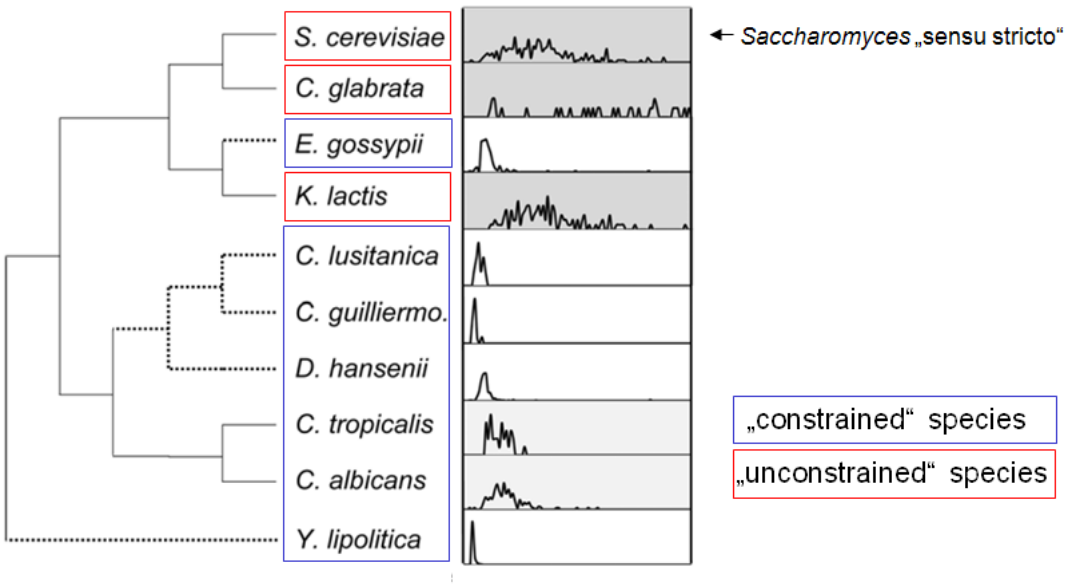


Fig. 4.2 Distribution of BP – 3´ss sequences in selected *Saccharomycotina* species. *C. glabrata – Candida glabrata; E. gosypii – Eremothecium gossypii; K. lactis – Kluyveromyces lactis; C. lusitanica – Candida lusitanica; Candida guilliermo. – Candida guilliermondi; D. hansenii – Debaryomyces hansenii; C. tropicalis – Candida tropicalis; C. albicans – Candida albicans; Y. lipolytica – Yarrowia lipolytica*. Adapted from Irimia and Roy, 2008.

The extreme representative of constrained species is *Yarrowia lipolytica*. Its BP and 3´ss splicing signals form one combined sequence with no more than five nucleotides between the branch point adenosine and the intron/exon boundary (Schwartz et al., 2008; Neuveglise et al., 2011). It is in sharp contrast to *S. cerevisiae*, which is not able to employ AG dinucleotides as 3´ss when located less than ten nucleotides from BP (Meyer et al., 2011). The diverse architecture of BP-3´ss regions also suggests different mechanisms for 3´ss recognition. Remarkably, the juxtaposition of BP and 3´ss was observed also in distant eukaryotes *Trichomonas vaginalis* and *Giardia lamblia* (Vanacova et al., 2005), indicating the possible convergence in the evolution of splicing signal recognition. Notably, all the dBP introns from unconstrained *Saccharomycotina* species (including *Candida glabrata* and *Kluyveromyces lactis*) we tested are predicted to form stable stem-like structures downstream of BP, indicating that the structure based mechanism might be plausible in multiple yeasts.

In higher eukaryotes, it is more difficult to determine the distribution of BP-3´ss distances. The main obstacle lies in the variability of BP sequences. This variability prevents credible identification of BP positions, which is the crucial prerequisite for calculating BP-3´ss distances. In contrast to exon/intron boundaries, which can be determined genome-widely from large scale transcriptome datasets, every single predicted BP position must be confirmed experimentally. There was an attempt to identify BPs of large number of verified human exons. The algorithm used in the study was based on the assumption that the sequences between BP and 3´ss are devoid of AG dinucleotides (AG exclusion zone; AGEZ). The estimated proportion of dBP human introns is between 1/160 and 1/400 (Gooding et al., 2006). We used RNAfold to model the secondary structure in several human dBP introns with experimentally confirmed BP, including four introns from the serotonin receptor 4 gene (HTR4; Hallegger et al., 2010). Surprisingly, in all introns tested, the region downstream of the BP is predicted to be completely unstructured. In contrast, random sequences with equivalent GC content always exhibit folding to at least some extent. Thus, this strongly suggests that secondary structure based mechanism for remote 3´ss recognition is not generally applicable in humans. Rather, this observation is in agreement with the previously proposed scanning model, wherein the spliceosome linearly explores the sequence downstream of BP and employs the first AG dinucleotide as an acceptor site (Smith et al., 1989; Smith et al., 1993).

Taken together, the variability in distributions of distances between BP and 3´ss together with the dissimilar (predicted) ability of dBP introns from unrelated species to fold into a stable secondary structure suggest the existence of multiple mechanisms for 3´ss recognition that are in general understood only very partially. More detailed understanding will be required to appreciate the master scheme for the 3´ss recognition from the various evolutionary adaptations that exist in eukaryotes. We believe that the identified "RNA based" mechanism of 3´ss recognition is not a part of this ancestral "master scheme". Rather, we expect that it has arisen in evolution of (yeast) genomes as a consequence of changes in intron architecture.

.

CONCLUSIONS

Prp45 in pre-mRNA splicing

- In the genetic screen we identified ten genes (*CLF1/SYF3, SYF1, NTC20, CEF1, PRP22, PRP17, PRP18, SLU7, PRP8,* and *COF1*) with mutation(s) synthetic lethal with *prp45*(1-169).

- In two hybrid assay we showed that the C-terminus of Prp45 is indispensable for the interaction of Prp45 with the Prp22 RNA helicase.

- We showed that the levels of Prp22 protein affect the viability of *prp45*(1-169) cells.

- We documented that the *prp45*(1-169) allele exacerbates the splicing defects of *ACT1-CUP1* substrates with suboptimal 5´ splice site, branch point, or 3´ splice site sequences, while the canonical *ACT1* intron is spliced with the wild-type efficiency.

- We mapped splicing defects of strains with the alleles synthetic lethal with *prp45*(1-169). All *prp22* mutants identified exhibit a similar splicing defect on the 3´ splice site mutated substrate.

- We showed that Prp45 differentially affects splicing of diverse introns. A severe splicing defect in *prp45*(1-169) cells was documented for the *COF1* intron.

- Using ectopic expression of the C-terminal portion of Prp45, we suppressed most *prp45*(1-169) related defect phenotypes (including temperature sensitivity, splicing defects, decreased Prp22 partition in spliceosomal complexes, and some synthetic lethal interactions). Prp45 can function in splicing when expressed *in trans*.

- We showed that the *prp45*(1-169) splicing defects can be partially suppressed by the expression of AtSKIP and SNW/SKIP, the *A. thaliana* and human Prp45 orthologs, respectively, suggesting an evolutionary conserved function of Prp45 in splicing.

Intron secondary structures in 3´ splice site recognition

- We showed the formation of a secondary structure between the branch point and the 3´ splice site in *COF1* and *UBC13* introns.
- Identified structures decrease an "effective" distance between the branch point and the 3´ splice site and mask potential cryptic 3´ss. Both mechanisms are required for proper 3´ss recognition and for efficient splicing.
- The resembling secondary structures are predicted in most *Saccharomyces cerevisiae* introns with remote 3´splice sites.
- The secondary structures in *COF1* and *UBC13* introns are conserved in all species from the *Saccharomyces sensu stricto* genus.

REFERENCES

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

Albers,M., Diment,A., Muraru,M., Russell,C.S., and Beggs,J.D. (2003). Identification and characterization of Prp45p and Prp46p, essential pre-mRNA splicing factors. RNA *9*, 138-150.

Alexander,R.D., Innocente,S.A., Barrass,J.D., and Beggs,J.D. (2010). Splicing-dependent RNA polymerase pausing in yeast. Mol. Cell *40*, 582-593.

Ambrozkova,M., Puta,F., Fukova,I., Skruzny,M., Brabek,J., and Folk,P. (2001). The fission yeast ortholog of the coregulator SKIP interacts with the small subunit of U2AF. Biochem. Biophys. Res. Commun. *284*, 1148-1154.

Anderson,K. and Moore,M.J. (2000). Bimolecular exon ligation by the human spliceosome bypasses early 3' splice site AG recognition and requires NTP hydrolysis. RNA. *6*, 16-25.

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.

Bacikova,D. and Horowitz,D.S. (2005). Genetic and functional interaction of evolutionarily conserved regions of the Prp18 protein and the U5 snRNA. Mol. Cell Biol. *25*, 2107-2116.

Barbour,L. and Xiao,W. (2006). Synthetic lethal screen. Methods Mol. Biol. *313*, 161-169.

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.

Ben Yehuda,S., Dix,I., Russell,C.S., McGarvey,M., Beggs,J.D., and Kupiec,M. (2000a). Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in Saccharomyces cerevisiae. Genetics *156*, 1503-1517.

Ben Yehuda,S., Russell,C.S., Dix,I., Beggs,J.D., and Kupiec,M. (2000b). Extensive genetic interactions between PRP8 and PRP17/CDC40, two yeast genes involved in pre-mRNA splicing and cell cycle progression. Genetics *154*, 61-71.

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., Krasauskas,A., Golas,M.M., Sander,B., Will,C.L., Urlaub,H., Stark,H., and Luhrmann,R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. RNA *16*, 2384-2403.

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

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

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

Bonnal,S., Martinez,C., Forch,P., Bachi,A., Wilm,M., and Valcarcel,J. (2008). RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. Mol. Cell *32*, 81-95.

Bragulat,M., Meyer,M., Macias,S., Camats,M., Labrador,M., and Vilardell,J. (2010). RPL30 regulation of splicing reveals distinct roles for Cbp80 in U1 and U2 snRNP cotranscriptional recruitment. RNA *16*, 2033-2041.

Bres,V., Gomes,N., Pickle,L., and Jones,K.A. (2005). A human splicing factor, SKIP, associates with P-TEFb and enhances transcription elongation by HIV-1 Tat. Genes Dev. *19*, 1211-1226.

Bres,V., Yoshida,T., Pickle,L., and Jones,K.A. (2009). SKIP interacts with c-Myc and Menin to promote HIV-1 Tat transactivation. Mol. Cell *36*, 75-87.

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

Brys,A. and Schwer,B. (1996). Requirement for SLU7 in yeast pre-mRNA splicing is dictated by the distance between the branchpoint and the 3' splice site. RNA *2*, 707-717.

Buratti,E. and Baralle,F.E. (2004). Influence of RNA secondary structure on the pre-mRNA splicing process. Mol. Cell Biol. *24*, 10505-10514.

Buratti,E., Muro,A.F., Giombi,M., Gherbassi,D., Iaconcig,A., and Baralle,F.E. (2004). RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. Mol. Cell Biol. *24*, 1387-1400.

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.

Cartegni,L., Chew,S.L., and Krainer,A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat. Rev. Genet. *3*, 285-298.

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.

Chan,S.P. and Cheng,S.C. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. J. Biol. Chem. *280*, 31190-31199.

Chan,S.P., Kao,D.I., Tsai,W.Y., and Cheng,S.C. (2003). The Prp19p-associated complex in spliceosome activation. Science *302*, 279-282.

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,C.H., Tsai,W.Y., Chen,H.R., Wang,C.H., and Cheng,S.C. (2001b). Identification and characterization of two novel components of the Prp19p-associated complex, Ntc30p and Ntc20p. J. Biol. Chem. *276*, 488-494.

Chen,C.H., Tsai,W.Y., Chen,H.R., Wang,C.H., and Cheng,S.C. (2001a). Identification and characterization of two novel components of the Prp19p-associated complex, Ntc30p and Ntc20p. J. Biol. Chem. *276*, 488-494.

Chen,C.H., Yu,W.C., Tsao,T.Y., Wang,L.Y., Chen,H.R., Lin,J.Y., Tsai,W.Y., and Cheng,S.C. (2002). Functional and physical interactions between components of the Prp19p-associated complex. Nucleic Acids Res. *30*, 1029-1037.

Chen,H.R., Jan,S.P., Tsao,T.Y., Sheu,Y.J., Banroques,J., and Cheng,S.C. (1998). Snt309p, a component of the Prp19p-associated complex that interacts with Prp19p and associates with the spliceosome simultaneously with or immediately after dissociation of U4 in the same manner as Prp19p. Mol. Cell Biol. *18*, 2196-2204.

Chen,H.R., Tsao,T.Y., Chen,C.H., Tsai,W.Y., Her,L.S., Hsu,M.M., and Cheng,S.C. (1999). Snt309p modulates interactions of Prp19p with its associated components to stabilize the Prp19p-associated complex essential for pre-mRNA splicing. Proc. Natl. Acad. Sci. U. S. A *96*, 5406-5411.

Chen,J.Y., Stands,L., Staley,J.P., Jackups,R.R., Jr., Latus,L.J., and Chang,T.H. (2001c). 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.

Chen,Y., Zhang,L., and Jones,K.A. (2011). SKIP counteracts p53-mediated apoptosis via selective regulation of p21Cip1 mRNA splicing. Genes Dev. *25*, 701-716.

Cheng,S.C. (1994). Formation of the yeast splicing complex A1 and association of the splicing factor PRP19 with the pre-mRNA are independent of the 3' region of the intron. Nucleic Acids Res. *22*, 1548-1554.

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.

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.

Crotti,L.B., Bacikova,D., and Horowitz,D.S. (2007). The Prp18 protein stabilizes the interaction of both exons with the U5 snRNA during the second step of pre-mRNA splicing. Genes Dev. *21*, 1204-1216.

Dahl,R., Wani,B., and Hayman,M.J. (1998). The Ski oncoprotein interacts with Skip, the human homolog of Drosophila Bx42. Oncogene *16*, 1579-1586.

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

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

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.

Figueroa,J.D. and Hayman,M.J. (2004). The human Ski-interacting protein functionally substitutes for the yeast PRP45 gene. Biochem. Biophys. Res. Commun. *319*, 1105-1109.

Folk,P., Puta,F., Krpejsova,L., Blahuskova,A., Markos,A., Rabino,M., and Dottin,R.P. (1996). The homolog of chromatin binding protein Bx42 identified in Dictyostelium. Gene *181*, 229-231.

Folk,P., Puta,F., and Skruzny,M. (2004). Transcriptional coregulator SNW/SKIP: the concealed tie of dissimilar pathways. Cell Mol. Life Sci. *61*, 629-640.

Fox-Walsh,K.L., Dou,Y., Lam,B.J., Hung,S.P., Baldi,P.F., and Hertel,K.J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc. Natl. Acad. Sci. U. S. A *102*, 16176-16181.

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

Gietz,R.D. and Woods,R.A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol. *350*, 87-96.

Goguel,V., Wang,Y., and Rosbash,M. (1993). Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing. Mol. Cell Biol. *13*, 6841-6848.

Golemis,E.A., Gyuris,J. and Brent,R. (1996) Interaction trap/two-hybrid system to identify interacting proteins. In Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.D., Smith,J.A., Struhl,K. (ed.), Current protocols in molecular biology, Unit 20.1.

Gooding,C., Clark,F., Wollerton,M.C., Grellscheid,S.N., Groom,H., and Smith,C.W. (2006). A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. Genome Biol. *7*, R1.

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.

Gozani,O., Feld,R., and Reed,R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes Dev. *10*, 233-243.

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.

Gruber,A.R., Lorenz,R., Bernhart,S.H., Neubock,R., and Hofacker,I.L. (2008). The Vienna RNA websuite. Nucleic Acids Res. *36*, W70-W74.

Guthrie,C. and Patterson,B. (1988). Spliceosomal snRNAs. Annu. Rev. Genet. *22*, 387-419.

Halfter,H. and Gallwitz,D. (1988). Impairment of yeast pre-mRNA splicing by potential secondary structure-forming sequences near the conserved branchpoint sequence. Nucleic Acids Res. *16*, 10413-10423.

Hallegger,M., Sobala,A., and Smith,C.W. (2010). Four exons of the serotonin receptor 4 gene are associated with multiple distant branch points. RNA. *16*, 839-851.

Harris,S.D., Cheng,J., Pugh,T.A., and Pringle,J.R. (1992). Molecular analysis of Saccharomyces cerevisiae chromosome I. On the number of genes and the identification of essential genes using temperature-sensitive-lethal mutations. J. Mol. Biol. *225*, 53-65.

Hofacker,I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res. *31*, 3429-3431.

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. *15*, 1515-1527.

Hossain,M.A., Rodriguez,C.M., and Johnson,T.L. (2011). Key features of the two-intron Saccharomyces cerevisiae gene SUS1 contribute to its alternative splicing. Nucleic Acids Res. *39*, 8612-8627.

Iwata,H. and Gotoh,O. (2011). Comparative analysis of information contents relevant to recognition of introns in many species. BMC. Genomics *12*, 45.

James,S.A., Turner,W., and Schwer,B. (2002). How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. RNA *8*, 1068-1077.

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.

Jurica,M.S., Licklider,L.J., Gygi,S.R., Grigorieff,N., and Moore,M.J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. RNA *8*, 426-439.

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

Kaida,D., Berg,M.G., Younis,I., Kasim,M., Singh,L.N., Wan,L., and Dreyfuss,G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature *468*, 664-668.

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.

Kim,Y.J., Noguchi,S., Hayashi,Y.K., Tsukahara,T., Shimizu,T., and Arahata,K. (2001). The product of an oculopharyngeal muscular dystrophy gene, poly(A)-binding protein 2, interacts with SKIP and stimulates muscle-specific gene expression. Hum. Mol. Genet. *10*, 1129-1139.

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. (1999). Site-specific derivatization of RNA with photocrosslinkable groups. Methods *18*, 22-28.

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.

Koodathingal,P., Novak,T., Piccirilli,J.A., and Staley,J.P. (2010). The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing. Mol. Cell *39*, 385-395.

Kotovic,K.M., Lockshon,D., Boric,L., and Neugebauer,K.M. (2003). Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. Mol. Cell Biol. *23*, 5768-5779.

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.

Lambowitz,A.M. and Zimmerly,S. (2011). Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb. Perspect. Biol. *3*, a003616.

Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R., Thompson,J.D., Gibson,T.J., and Higgins,D.G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics. *23*, 2947-2948.

Lehmann,K. and Schmidt,U. (2003). Group II introns: structure and catalytic versatility of large natural ribozymes. Crit Rev. Biochem. Mol. Biol. *38*, 249-303.

Leong,G.M., Subramaniam,N., Figueroa,J., Flanagan,J.L., Hayman,M.J., Eisman,J.A., and Kouzmenko,A.P. (2001). Ski-interacting protein interacts with Smad proteins to augment transforming growth factor-beta-dependent transcription. J. Biol. Chem. *276*, 18243-18248.

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.

Lim,L.P. and Burge,C.B. (2001). A computational analysis of sequence features involved in recognition of short introns. Proc. Natl. Acad. Sci. U. S. A *98*, 11193-11198.

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.

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.

Madhani,H.D. and Guthrie,C. (1994). Dynamic RNA-RNA interactions in the spliceosome. Annu. Rev. Genet. *28*, 1-26.

Makarov,E.M., Makarova,O.V., Urlaub,H., Gentzel,M., Will,C.L., Wilm,M., and Luhrmann,R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science *298*, 2205-2208.

Makarova,O.V., Makarov,E.M., Urlaub,H., Will,C.L., Gentzel,M., Wilm,M., and Luhrmann,R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. EMBO J. *23*, 2381-2391.

Martin,A., Schneider,S., and Schwer,B. (2002). Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. J. Biol. Chem. *277*, 17743-17750.

Martinkova,K., Lebduska,P., Skruzny,M., Folk,P., and Puta,F. (2002). Functional mapping of Saccharomyces cerevisiae Prp45 identifies the SNW domain as essential for viability. J. Biochem. *132*, 557-563.

Mayas,R.M., Maita,H., Semlow,D.R., and Staley,J.P. (2010). Spliceosome discards intermediates via the DEAH box ATPase Prp43p. Proc. Natl. Acad. Sci. U. S. A *107*, 10020-10025.

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.

McPheeters,D.S. and Muhlenkamp,P. (2003). Spatial organization of protein-RNA interactions in the branch site-3' splice site region during pre-mRNA splicing in yeast. Mol. Cell Biol. *23*, 4174-4186.

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.

Meyer,M., Plass,M., Perez-Valle,J., Eyras,E., and Vilardell,J. (2011). Deciphering 3'ss selection in the yeast genome reveals an RNA thermosensor that mediates alternative splicing. Mol. Cell *43*, 1033-1039.

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.

Mishra,S.K., Ammon,T., Popowicz,G.M., Krajewski,M., Nagel,R.J., Ares,M., Jr., Holak,T.A., and Jentsch,S. (2011). Role of the ubiquitin-like protein Hub1 in splice-site usage and alternative splicing. Nature *474*, 173-178.

Mitrovich,Q.M. and Guthrie,C. (2007). Evolution of small nuclear RNAs in S. cerevisiae, C. albicans, and other hemiascomycetous yeasts. RNA *13*, 2066-2080.

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.

Mumberg,D., Muller,R., and Funk,M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene *156*, 119-122.

Muro,A.F., Caputi,M., Pariyarath,R., Pagani,F., Buratti,E., and Baralle,F.E. (1999). Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. Mol. Cell Biol. *19*, 2657-2671.

Nandabalan,K. and Roeder,G.S. (1995). Binding of a cell-type-specific RNA splicing factor to its target regulatory sequence. Mol. Cell Biol. *15*, 1953-1960.

Neuveglise,C., Marck,C., and Gaillardin,C. (2011). The intronome of budding yeasts. C. R. Biol. *334*, 662-670.

Newman,A.J. (1997). The role of U5 snRNP in pre-mRNA splicing. EMBO J. *16*, 5797-5800.

Newman,A.J. (2008). RNA interactions in mRNA splicing. In Encyclopedia of Life Sciences, John Wiley & Sons, New York.

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

Nielsen,H. and Johansen,S.D. (2009). Group I introns: Moving in new directions. RNA Biol. *6*, 375-383.

Oesterreich,F.C., Preibisch,S., and Neugebauer,K.M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Mol. Cell *40*, 571-581.

Oesterreich,F.C., Bieberstein,N., and Neugebauer,K.M. (2011). Pause locally, splice globally. Trends Cell Biol. *21*, 328-335.

Ohi,M.D. and Gould,K.L. (2002). Characterization of interactions among the Cef1p-Prp19p-associated splicing complex. RNA *8*, 798-815.

Ohi,M.D., Link,A.J., Ren,L., Jennings,J.L., McDonald,W.H., and Gould,K.L. (2002). Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. Mol. Cell Biol. *22*, 2011-2024.

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

Parenteau,J., Durand,M., Veronneau,S., Lacombe,A.A., Morin,G., Guerin,V., Cecez,B., Gervais-Bird,J., Koh,C.S., Brunelle,D., Wellinger,R.J., Chabot,B., and Abou,E.S. (2008). Deletion of many yeast introns reveals a minority of genes that require splicing for function. Mol. Biol. Cell *19*, 1932-1941.

Parisien,M. and Major,F. (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature *452*, 51-55.

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.

Patel,A.A. and Steitz,J.A. (2003). Splicing double: insights from the second spliceosome. Nat. Rev. Mol. Cell Biol. *4*, 960-970.

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.

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

Pleiss,J.A., Whitworth,G.B., Bergkessel,M., and Guthrie,C. (2007b). Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. PLoS. Biol. *5*, e90.

Prathapam,T., Kuhne,C., and Banks,L. (2002). Skip interacts with the retinoblastoma tumor suppressor and inhibits its transcriptional repression activity. Nucleic Acids Res. *30*, 5261-5268.

Pyle,A.M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. Annu. Rev. Biophys. *37*, 317-336.

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. (1998a). A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. Science *279*, 857-860.

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

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

Reed,S.I., Hadwiger,J.A., Richardson,H.E., and Wittenberg,C. (1989). Analysis of the Cdc28 protein kinase complex by dosage suppression. J. Cell Sci. Suppl *12*, 29-37.

Rogic,S., Montpetit,B., Hoos,H.H., Mackworth,A.K., Ouellette,B.F., and Hieter,P. (2008). Correlation between the secondary structure of pre-mRNA introns and the efficiency of splicing in Saccharomyces cerevisiae. BMC. Genomics *9*, 355.

Rosbash,M. and Seraphin,B. (1991). Who's on first? The U1 snRNP-5' splice site interaction and splicing. Trends Biochem. Sci. *16*, 187-190.

Rymond,B.C. (2010). The branchpoint binding protein: in and out of the spliceosome cycle. Adv. Exp. Med. Biol. *693*, 123-141.

Rymond,B.C., Torrey,D.D., and Rosbash,M. (1987). A novel role for the 3' region of introns in pre-mRNA splicing of Saccharomyces cerevisiae. Genes Dev. *1*, 238-246.

Sapra,A.K., Arava,Y., Khandelia,P., and Vijayraghavan,U. (2004). Genome-wide analysis of pre-mRNA splicing: intron features govern the requirement for the

second-step factor, Prp17 in Saccharomyces cerevisiae and Schizosaccharomyces pombe. J. Biol. Chem. *279*, 52437-52446.

Scannell,D.R., Butler,G., and Wolfe,K.H. (2007). Yeast genome evolution--the origin of the species. Yeast *24*, 929-942.

Scherrer,F.W., Jr. and Spingola,M. (2006). A subset of Mer1p-dependent introns requires Bud13p for splicing activation and nuclear retention. RNA *12*, 1361-1372.

Schneider,S. and Schwer,B. (2001). Functional domains of the yeast splicing factor Prp22p. J. Biol. Chem. *276*, 21184-21191.

Schwartz,S.H., Silva,J., Burstein,D., Pupko,T., Eyras,E., and Ast,G. (2008). Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. Genome Res. *18*, 88-103.

Schwer,B. (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. Mol. Cell *30*, 743-754.

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. (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.

Sharma,S., Kohlstaedt,L.A., Damianov,A., Rio,D.C., and Black,D.L. (2008). Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. Nat. Struct. Mol. Biol. *15*, 183-191.

Sikorski,R.S. and Hieter,P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19-27.

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.

Singh,N.N., Singh,R.N., and Androphy,E.J. (2007). Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. Nucleic Acids Res. *35*, 371-389.

Singh,R., Banerjee,H., and Green,M.R. (2000). Differential recognition of the polypyrimidine-tract by the general splicing factor U2AF65 and the splicing repressor sex-lethal. RNA *6*, 901-911.

Singh,R. and Valcarcel,J. (2005). Building specificity with nonspecific RNA-binding proteins. Nat. Struct. Mol. Biol. *12*, 645-653.

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., Konarska,M.M., and Query,C.C. (2009). Insights into branch nucleophile positioning and activation from an orthogonal pre-mRNA splicing system in yeast. Mol. Cell *34*, 333-343.

Smith,D.J., Query,C.C., and Konarska,M.M. (2007). trans-splicing to spliceosomal U2 snRNA suggests disruption of branch site-U2 pairing during pre-mRNA splicing. Mol. Cell *26*, 883-890.

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. and Ares,M., Jr. (2000). A yeast intronic splicing enhancer and Nam8p are required for Mer1p-activated splicing. Mol. Cell *6*, 329-338.

Spingola,M., Armisen,J., and Ares,M., Jr. (2004). Mer1p is a modular splicing factor whose function depends on the conserved U2 snRNP protein Snu17p. Nucleic Acids Res. *32*, 1242-1250.

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.

Steffen,P., Voss,B., Rehmsmeier,M., Reeder,J., and Giegerich,R. (2006). RNAshapes: an integrated RNA analysis package based on abstract shapes. Bioinformatics. *22*, 500-503.

Suh,S.O., Blackwell,M., Kurtzman,C.P., and Lachance,M.A. (2006). Phylogenetics of Saccharomycetales, the ascomycete yeasts. Mycologia. *98*, 1006-1017.

Sun,J.S. and Manley,J.L. (1995). A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. Genes Dev. *9*, 843-854.

Szklarczyk,D., Franceschini,A., Kuhn,M., Simonovic,M., Roth,A., Minguez,P., Doerks,T., Stark,M., Muller,J., Bork,P., Jensen,L.J., and von Mering,C. (2011). The

STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. *39*, D561-D568.

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

Tarn,W.Y., Hsu,C.H., Huang,K.T., Chen,H.R., Kao,H.Y., Lee,K.R., and Cheng,S.C. (1994). Functional association of essential splicing factor(s) with PRP19 in a protein complex. EMBO J. *13*, 2421-2431.

Toor,N., Keating,K.S., and Pyle,A.M. (2009). Structural insights into RNA splicing. Curr. Opin. Struct. Biol. *19*, 260-266.

Tsai,W.Y., Chow,Y.T., Chen,H.R., Huang,K.T., Hong,R.I., Jan,S.P., Kuo,N.Y., Tsao,T.Y., Chen,C.H., and Cheng,S.C. (1999). Cef1p is a component of the Prp19p-associated complex and essential for pre-mRNA splicing. J. Biol. Chem. *274*, 9455-9462.

Tseng,C.K. and Cheng,S.C. (2008). Both catalytic steps of nuclear pre-mRNA splicing are reversible. Science *320*, 1782-1784.

Valadkhan,S. (2010). Role of the snRNAs in spliceosomal active site. RNA Biol. *7*, 345-353.

Valadkhan,S. and Jaladat,Y. (2010). The spliceosomal proteome: at the heart of the largest cellular ribonucleoprotein machine. Proteomics. *10*, 4128-4141.

van Nues,R.W. and Beggs,J.D. (2001). Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of Saccharomyces cerevisiae. Genetics *157*, 1451-1467.

Vanacova,S., Yan,W., Carlton,J.M., and Johnson,P.J. (2005). Spliceosomal introns in the deep-branching eukaryote Trichomonas vaginalis. Proc. Natl. Acad. Sci. U. S. A *102*, 4430-4435.

Vincent,K., Wang,Q., Jay,S., Hobbs,K., and Rymond,B.C. (2003). Genetic interactions with CLF1 identify additional pre-mRNA splicing factors and a link between activators of yeast vesicular transport and splicing. Genetics *164*, 895-907.

Wagner,J.D., Jankowsky,E., Company,M., Pyle,A.M., and Abelson,J.N. (1998). The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. EMBO J. *17*, 2926-2937.

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

Wang,Q., Hobbs,K., Lynn,B., and Rymond,B.C. (2003). The Clf1p splicing factor promotes spliceosome assembly through N-terminal tetratricopeptide repeat contacts. J. Biol. Chem. *278*, 7875-7883.

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

Wang,Y., Fu,Y., Gao,L., Zhu,G., Liang,J., Gao,C., Huang,B., Fenger,U., Niehrs,C., Chen,Y.G., and Wu,W. (2010). Xenopus skip modulates Wnt/beta-catenin signaling and functions in neural crest induction. J. Biol. Chem. *285*, 10890-10901.

Wang,Y. and Guthrie,C. (1998). PRP16, a DEAH-box RNA helicase, is recruited to the spliceosome primarily via its nonconserved N-terminal domain. RNA *4*, 1216-1229.

Warf,M.B. and Berglund,J.A. (2010). Role of RNA structure in regulating pre-mRNA splicing. Trends Biochem. Sci. *35*, 169-178.

Wieland,C., Mann,S., von Besser,H., and Saumweber,H. (1992). The Drosophila nuclear protein Bx42, which is found in many puffs on polytene chromosomes, is highly charged. Chromosoma *101*, 517-525.

Will,C.L. and Luhrmann,R. (1997). Protein functions in pre-mRNA splicing. Curr. Opin. Cell Biol. *9*, 320-328.

Will,C.L. and Luhrmann,R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. Curr. Opin. Cell Biol. *13*, 290-301.

Will,C.L. and Luhrmann,R. (2005). Splicing of a rare class of introns by the U12-dependent spliceosome. Biol. Chem. *386*, 713-724.

Will,C.L. and Luhrmann,R. (2011). Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. *3*.

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

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.

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,Y., Zhan,L., Zhang,W., Sun,F., Wang,W., Tian,N., Bi,J., Wang,H., Shi,D., Jiang,Y., Zhang,Y., and Jin,Y. (2011). RNA secondary structure in mutually exclusive splicing. Nat. Struct. Mol. Biol. *18*, 159-168.

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

Zhang,C., Baudino,T.A., Dowd,D.R., Tokumaru,H., Wang,W., and MacDonald,P.N. (2001). Ternary complexes and cooperative interplay between NCoA-62/Ski-

interacting protein and steroid receptor coactivators in vitamin D receptor-mediated transcription. J. Biol. Chem. *276*, 40614-40620.

Zhang,C., Dowd,D.R., Staal,A., Gu,C., Lian,J.B., van Wijnen,A.J., Stein,G.S., and MacDonald,P.N. (2003). Nuclear coactivator-62 kDa/Ski-interacting protein is a nuclear matrix-associated coactivator that may couple vitamin D receptor-mediated transcription and RNA splicing. J. Biol. Chem. *278*, 35325-35336.

Zhang,X. and Schwer,B. (1997). Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. Nucleic Acids Res. *25*, 2146-2152.

Zhou,S., Fujimuro,M., Hsieh,J.J., Chen,L., Miyamoto,A., Weinmaster,G., and Hayward,S.D. (2000). SKIP, a CBF1-associated protein, interacts with the ankyrin repeat domain of NotchIC To facilitate NotchIC function. Mol. Cell Biol. *20*, 2400-2410.

Zhou,Z., Licklider,L.J., Gygi,S.P., and Reed,R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature *419*, 182-185.

Zuker,M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31*, 3406-3415.

APPENDICES

List of appendices:

A. Sequences of splicing reporters

A.1 Sequences of splicing reporters obtained from other labs

A.2 Sequences of splicing reporters prepared in this study

B. Publications (available on the attached CD)

B.1 Publication 1

Gahura O., Abrhámová K., Skružný M., Valentová A., Munzarová V., Folk P., Půta F.
Prp45 affects Prp22 partition in spliceosomal complexes and splicing efficiency of non-consensus substrates.
J Cell Biochem 2009, 106(1):139-151 (IF 3.381)

B.2 Publication 2

Gahura O., Hammann Ch., Valentová A., Půta F., Folk P.
Secondary structure is required for 3' splice site recognition in yeast.
Nucleic Acids Research, in press (IF 7.836)

B.3 Publication 3

Wang X., Wu F., Xie Q., Wang H., Wang Y., Yue Y., Gahura O., Ma S., Liu L., Cao Y., Jiao Y., Puta F., McClung C. R., Xu X., Ma L.
SKIP is a splicing factor linking alternative splicing and circadian clock in *Arabidopsis*
Plant Cell, manuscript submitted (IF 9.396)

A. Sequences of splicing reporters

A.1 Sequences of splicing reporters obtained from other labs

All sequences start at the ATG start codon and end just before the CUP1 coding sequence.

pMM1C (canonical ACT1-CUP1 „recombinant“)
ATGGATTCTGGTATGTTCTAGCGCATGCACCATCCCATTTAACTGTAAGAAGAATTGCACGGTCCCAA
TTGCTCGAGAGATTTCTCTTTTACCTTTTTTTACTATTTTTCACTCTCCCATAACCTCCTATATTGAC
TGATCTGTAATAACCACGATATTATTGGAATAAATAGGGGCTTGAAATTTGGAAAAAAAAAAAACTGA
AATATTTTCGTGATAAGTGATAGTGATATTCTTCTTTTATTTGCTACTGTGTCTCATGTACTAACATC
GATTGCTTCATTCTTTTTGTTGCTATATTAATGTAGAATATTCATTCTCC

pMM4C (3´ss gAG ACT1-CUP1 „recombinant“)
ATGGATTCTGGTATGTTCTAGCGCTTGCACCATCCCATTTAACTGTAAGAAGAATTGCACGGTCCCAA
TTGCTCGAGAGATTTCTCTTTTATCTTTTTTTTACTATTTTTCACTCTCCCATAACCTCCTATATTGAC
TGATCTGTAATAACCACGATATTATTGGAATAAATAGGGGCTTGAAATTTGGAAAAAAAAAAAACTGA
AATATTTTCGTGATAAGTGATAGTGATATTCTTCTTTTATTTGCTACTGTGTCTCATGTACTAACATC
GATTGCTTCATTCTTTTTGTTGCTATATTAATGGAGAATATTCATTCTCC

pRH20 (5´ss A3C ACT1-CUP1 „recombinant“)
ATGGATTCTGGTCTGTTCTAGCGCTTGCACCATCCCATTTAACTGTAAGAAGAATTGCACGGTCCCAA
TTGCTCGAGAGATTTCTCTTTTACCTTTTTTTTACTATTTTTCACTCTCCCATAACCTCCTATATTGAC
TGATCTGTAATAACCACGATATTATTGGAATAAATAGGGGCTTGAAATTTGGAAAAAAAAAAAACTGA
AATATTTTCGTGATAAGTGATAGTGATATTCTTCTTTTATTTGCTACTGTGTCTCATGTACTAACATC
GATTGCTTCATTCTTTTTGTTGCTATATTAATGTAGAATATTCATTCTCC

pMA2C (5´ss A3C ACT1-CUP1 „actin“)
ATGGATTCTGGTCTGTTCTAGCGCTTGCACCATCCCATTTAACTGTAAGAAGAATTGCACGGTCCCAA
TTGCTCGAGAGATTTCTCTTTTACCTTTTTTTTACTATTTTTCACTCTCCCATAACCTCCTATATTGAC
TGATCTGTAATAACCACGATATTATTGGAATAAATAGGGGCTTGAAATTTGGAAAAAAAAAAAACTGA
AATATTTTCGTGATAAGTGATAGTGATATTCTTCTTTTATTTGCTACTGTGTCTCATGTACTAACATC
GATTGCTTCATTCTTTTTGTTGCTATATTATATGTTTAGAGGTTGCTGCTTTGGTTATTGATAACGGT
TCTGGTATGTGTAAAGCCGGTACCGG

pMA42 (BP-C ACT1-CUP1 „actin“)
ATGGATTCTGGTATGTTCTAGCGCTTGCACCATCCCATTTAACTGTAAGAAGAATTGCACGGTCCCAA
TTGCTCGAGAGATTTCTCTTTTACCTTTTTTTTACTATTTTTCACTCTCCCATAACCTCCTATATTGAC
TGATCTGTAATAACCACGATATTATTGGAATAAATAGGGGCTTGAAATTTGGAAAAAAAAAAAACTGA
AATATTTTCGTGATAAGTGATAGTGATATTCTTCTTTTATTTGCTACTGTGTCTCATGTACTACCATC
GATTGCTTCATTCTTTTTGTTGCTATATTATATGTTTAGAGGTTGCTGCTTTGGTTATTGATAACGGT
TCTGGTATGTGTAAAGCCGGTACCGG

pCQ04 (BP-G ACT1-CUP1 „actin“) – *expected, not confirmed by sequencing*
ATGGATTCTGGTATGTTCTAGCGCTTGCACCATCCCATTTAACTGTAAGAAGAATTGCACGGTCCCAA
TTGCTCGAGAGATTTCTCTTTTACCTTTTTTTTACTATTTTTCACTCTCCCATAACCTCCTATATTGAC
TGATCTGTAATAACCACGATATTATTGGAATAAATAGGGGCTTGAAATTTGGAAAAAAAAAAAACTGA
AATATTTTCGTGATAAGTGATAGTGATATTCTTCTTTTATTTGCTACTGTGTCTCATGTACTAGCATC
GATTGCTTCATTCTTTTTGTTGCTATATTATATGTTTAGAGGTTGCTGCTTTGGTTATTGATAACGGT
TCTGGTATGTGTAAAGCCGGTACCGG

pC256A (C256A ACT1-CUP1 „actin“)
ATGGATTCTGGTATGTTCTAGCGCTCGCACCATCCCATTTAACTGTAAGAAGAATTGCACGGTCCCAA
TTGCTCGAGAGATTTCTCTTTTACCTTTTTTTTACTATTTTTCACTCTCCCATAACCTCCTATATTGAC
TGATCTGTAATAACCACGATATTATTGGAATAAATAGGGGCTTGAAATTTGGAAAAAAAAAAAACTGA
AATATTTTCGTGATAAGTGATAGTGATATTCTTCTTTTATTAGCTACTGTGTCTCATGTAATAACATC
GATTGCTTCATTCTTTTTGTTGCTATATTATATGTTTAGAGGTTGCTGCTTTGGTTATTGATAACGGT
TCTGGTATGTGTAAAGCCGGTACCGG

pJP40 (3´ss aAG ACT1-CUP1 „actin")
ATGGATTCTGGTATGTTCTAGCGCTTGCACCATCCCATTTAACTGTAAGAAGAATTGCACGGTCCCAA
TTGCTCGAGAGATTTCTCTTTTACCTTTTTTTACTATTTTTCACTCTCCCATAACCTCCTATATTGAC
TGATCTGTAATAACCACGATATTATTGGAATAAATAGGGGCTTGAAATTTGGAAAAAAAAAAAACTGA
AATATTTTCGTGATAAGTGATAGTGATATTCTTCTTTTATTTGCTACTGTGTCTCATGTACTAACATC
GATTGCTTCATTCTTTTTGTTGCTATATTATATGTTAAGAGGTTGCTGCTTTGGTTATTGATAACGGT
TCTGGTATGTGTAAAGCCGGTACCGG

pJP56 (3´ss cAG ACT1-CUP1 „actin")
ATGGATTCTGGTATGTTCTAGCGCTTGCACCATCCCATTTAACTGTAAGAAGAATTGCACGGTCCCAA
TTGCTCGAGAGATTTCTCTTTTACCTTTTTTTACTATTTTTCACTCTCCCATAACCTCCTATATTGAC
TGATCTGTAATAACCACGATATTATTGGAATAAATAGGGGCTTGAAATTTGGAAAAAAAAAAAACTGA
AATATTTTCGTGATAAGTGATAGTGATATTCTTCTTTTATTTGCTACTGTGTCTCATGTACTAACATC
GATTGCTTCACTCTTTTTGTTGCTATATTATATGTTCAGAGATTGCTGCTTTGGTTATTGATAACGGT
TCTGGTATGTGTAAAGCCGGACCGGGTTCAGCGAATAATAAC

A.2 Sequences of splicing reporters prepared in this study

CUP1 fusion genes are shown from the BamHI restriction site in the 5´UTR used for the cloning into the p423GPD vector and end before the CUP1 coding sequence.

All cof1-HFM1 constructs are shown including the HFM1 second exon fragments ending with the SalI cloning site.

ACT1 (pOG58)
GGATCCTTTTAGATTTTTCACGCTTACTGCTTTTTTCTTCCCAAGATCGAAAATTTACTGAATTAACA
ATGGATTCTGGTATGTTCTAGCGCTTGCACCATCCCATTTAACTGTAAGAAGAATTGCACGGTCCCAA
TTGCTCGAGAGATTTCTCTTTTACCTTTTTTTACTATTTTTCACTCTCCCATAACCTCCTATATTGAC
TGATCTGTAATAACCACGATATTATTGGAATAAATAGGGGCTTGAAATTTGGAAAAAAAAAAAACTGA
AATATTTTCGTGATAAGTGATAGTGATATTCTTCTTTTATTTGCTACTGTTACTAAGTCTCATGTACT
AACATCGATTGCTTCATTCTTTTTGTTGCTATATTATATGTTTAGAGGTTGCTGCTTTGGTTATTGAT
AACGGTTCTGGTATGTGTAAAGCCGAA

TUB3
GGATCCAAGCGACTTGAGACAATGAGAGAGGTCATTAGTATTAATGGTATGTATGCGTTCCTTTTTTT
GTTCAATATTCGCAACCAATGGCACCTGTGGGACAGGGAAAGAAGTTTGATCTGATCTGGTTTGATTC
ATTCCCAATTGGTCACCATCTGGTTGATTTACGGCAAATAATTTGACTTGTACCAGCACAGTTTACTA
ACAGTTTCTTTTTCTCCATTTTTTCTGGGCATACTCGGACGAAAAAGCTCATAATTGACCTCATTACA
TGGGGAGTGATTTTTGTGTCTTCTTCTTCGGAGGATTGCTGGAACTTTTGTTATTTTTCTTTTTTACA
ACAGTTGGTCAAGCAGGTTGTCAAATAGGTAATGCATGCTGGGAA

PFY1
GGATCCAACTACGATCGCAAATTATGTCTTGGCAAGGTATGTGAACGAGACAATTATCAATTGATTAA
GAAAGAAATGAGTCGGAGGTTAGCTTGTGTGACAATGTTTGGCAATGCCCGATTTTTGTTGATGCGCG
TAATTTCGAAGATTAACCACTCAGAGTAAATTACTAACTGGAATATCAAAAAACATATGAAATTTCAA
ACATGAATTTCTTTCCGTTTTTTTCTCCTACTTTTAAACAGCATACACTGATAACTTAATAGGAACCG
AA

RPL30
GGATCCTACAGCTTATTCAATTAATCAATATACGCAGAGATGGTCAGTATAACATGATTTTATAACTA
TTTCGCTTTATTCTTTCATGGGATTAGCAAGAAGGCTTGGAATACAAACATGAATTTTCAAAGTTTAT
TCCTGCTCTTGTGTTTGAGAGAGTCCGTTCAAAGGATGCTTCTCCATGACTGTGAAATAAAAGTGAGA
TATGCAAAAACTTTGAATAATCTACTAACAAGTTAATACGTTATTTTTTATCGTTTACATTTCAACAG
GCCCCAGTTAAATCCCAAGAA

CPT1
GGATCCATGGATTCGAAAACAGTATTACTAACTGATTCAGGGGAAACAATGGGATTCTTTATTCCTCA
GAGTAGTTTGGGAAACTTGAAGCTTTACAAGTATGTTGCTTATCTTATTGCACCCTAAATCTTCTGTG
GCAAGAAAAATAAATACTAACTATTGTAACATTAAATTACTAATCTGACTTTAGGTATCAAAGTGAA

COF1
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATAAGAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAA

cof1 G149A
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATAAAAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAA

cof1 C80T
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTTTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATAAGAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAA

cof1 G149A+C80T
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTTTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATAAAAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAA

cof1 A148G
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATAGGAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAA

cof1 A148C
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATACGAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAA

cof1 A148T
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATATGAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAA

cof1(Δ91-140)
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTATTCATAAGAAGCC
AGGAAAATGTGAAGTTTCTGTGTAGTGTTGCTGTTGCTGAA

cof1(Δ91-140)G149A
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTATTCATAAAAAGCC
AGGAAAATGTGAAGTTTCTGTGTAGTGTTGCTGTTGCTGAA

cof1(Δ91-140)G149A+C80T
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTTTTATGTCCTTATTCATAAAAAGCC
AGGAAAATGTGAAGTTTCTGTGTAGTGTTGCTGTTGCTGAA

cof1(Δ76-152)
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCCAGGAAAATGTGAAGTTTCTGTGTAGTG
TTGCTGTTGCTGAA

cof1(Δ76-176)
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGTAGTGTTGCTGTTGCTGAA

cof1-hel
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCCAAGTATCTCTCCTTTCTTTCCTCCCTTTT
TAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTGAGATACTTGCAGGAAAATGTGAAGTTTCT
GTGTAGTGTTGCTGTTGCTGAA

UBC13-CUP1
GGATCCATTGTAACATAGTTAGAAATGGCATCATTACCCAAGAGAATAATCAAGGTATGTTCAAACAA
GAGCAGTGGATTATATACGAAATCTGAAGTGACGTTTTCGTCTGAGCCGTCACACTCTACACTGTTTC
CTCTTGATAATGGCTAGAACTTACATACTAACGCAAAACTTCCAAAACAAGAAGGCTTTTTTTGTCCA
GTTTGTAGCAGCGTTCCGCCTTCCCTTTGTCCATATCCGCAAACGCTCATATGGAGGGAATTCAACAG
AACAAACTACTAGGGCAGAAAACGCAAAGTCTTTGTGAACTTTCATATAGGAAACTGAGAAATTAGTA
AGTGAA

ubc13 disordered
GGATCCATTGTAACATAGTTAGAAATGGCATCATTACCCAAGAGAATAATCAAGGUAUGUUCAAACAA
GAGCAGUGGAUUAUAUACGAAAUCUGAAGUGACGUUUUCGUCUGAGCCGUCACACUCUACACUGUUUC
CUCUUGAUAAUGGCUAGAACUUACAUACUAACGCAAAACUUCCAAAACAAGAAGGCUUUUUUUGUCCA
GUUUGUAGCAGAGUUCCGCCUUCCCUUUGUCCAUAUCCGCAAACCAUCUAAUCACGCUCAUUCAACAG
AAACAUCUACUACUCAUCUUAGCGCAAAGUCUUUGUGAACUUUCAUAUAGGAAACTGAGAAATTAGTA
AGTGAA

ubc13 reordered
GGATCCATTGTAACATAGTTAGAAATGGCATCATTACCCAAGAGAATAATCAAGGUAUGUUCAAACAA
GAGCAGUGGAUUAUAUACGAAAUCUGAAGUGACGUUUUCGUCUGAGCCGUCACACUCUACACUGUUUC
CUCUUGAUAAUGGCUAGAACUUACAUACUAACGCAAAACUUCCAAAACAAGAAGGCUUUAAGGUGAGA
GAUGUUAGCAGCGUUCCGCCUGAGCUUUGGUGAUUACCGCAAACGCUCUAAUCACGCUCAUUCAACAG
AAACAUCUACUACUCAUCUUAGCGCAAAGUCUUUGUGAACUUUCAUAUAGGAAACTGAGAAATTAGTA
AGTGAA

ubc13 G232A
GGATCCATTGTAACATAGTTAGAAATGGCATCATTACCCAAGAGAATAATCAAGGUAUGUUCAAACAA
GAGCAGUGGAUUAUAUACGAAAUCUGAAGUGACGUUUUCGUCUGAGCCGUCACACUCUACACUGUUUC
CUCUUGAUAAUGGCUAGAACUUACAUACUAACGCAAAACUUCCAAAACAAGAAGGCUUUUUUUGUCCA
GUUUGUAGCAGCGUUCCGCCUUCCCUUUGUCCAUAUCCGCAAACGCUCAUAUGGAGGGAAUUCAACAG
AACAAACUACUAGAGCAGAAAACGCAAAGUCUUUGUGAACUUUCAUAUAGGAAACTGAGAAATTAGTA
AGTGAA

ubc13 G232A+C148T
GGATCCATTGTAACATAGTTAGAAATGGCATCATTACCCAAGAGAATAATCAAGGUAUGUUCAAACAA
GAGCAGUGGAUUAUAUACGAAAUCUGAAGUGACGUUUUCGUCUGAGCCGUCACACUCUACACUGUUUC
CUCUUGAUAAUGGCUAGAACUUACAUACUAACGCAAAACUUCCAAAACAAGAAGGCUUUUUUUGUTCA
GUUUGUAGCAGCGUUCCGCCUUCCCUUUGUCCAUAUCCGCAAACGCUCAUAUGGAGGGAAUUCAACAG
AACAAACUACUAGAGCAGAAAACGCAAAGUCUUUGUGAACUUUCAUAUAGGAAACTGAGAAATTAGTA
AGTGAA

COF1-HFM1/100
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATAAGAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAATTCCAGTCTGCTACATTTAAAAGAAATAAGAAAAATAGCCGCCAA
CCAAGTTTTAAAGTAGGTTTGTCTTATAACAGTTTACTGGATGATTGAGTCGAC

COF1-HFM1/200
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATAAGAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAATTCCAGTCTGCTACATTTAAAAGAAATAAGAAAAATAGCCGCCAA
CCAAGTTTTAAAGTAGGTTTGTCTTATAACAGTTTACTGGATGATTGTGATGATGAAAATGAAACAGA
GGAAATATTTGAGGGCCGTGGTCTGCAATTCTTCGACAAAGACGATAATTTTTCAATAACAGCAGATG
ATACACAAGTAACTAGTAAATGAGTCGAC

COF1-HFM1/400
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATAAGAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAATTCCAGTCTGCTACATTTAAAAGAAATAAGAAAAATAGCCGCCAA
CCAAGTTTTAAAGTAGGTTTGTCTTATAACAGTTTACTGGATGATTGTGATGATGAAAATGAAACAGA
GGAAATATTTGAGGGCCGTGGTCTGCAATTCTTCGACAAAGACGATAATTTTTCAATAACAGCAGATG
ATACACAAGTAACTAGTAAATTATTTGATCACGATTTAGAGCAGACACCTGATGAAGAGGCGAAGAAA
CCTAAGAAGGTCACGATAAGGAAATCAGCGAAGAAATGTTTGAGCACGACCATATTACCAGATTCATT
CAGAGGGGTTTTTAAATTTACAGAATTTAATAAAATGCAATCAGAGGCTTTTCCTAGTATTTATGAGA
GTAACGAGAACTGAGTCGAC

COF1-HFM1/600
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATAAGAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAATTCCAGTCTGCTACATTTAAAAGAAATAAGAAAAATAGCCGCCAA
CCAAGTTTTAAAGTAGGTTTGTCTTATAACAGTTTACTGGATGATTGTGATGATGAAAATGAAACAGA
GGAAATATTTGAGGGCCGTGGTCTGCAATTCTTCGACAAAGACGATAATTTTTCAATAACAGCAGATG
ATACACAAGTAACTAGTAAATTATTTGATCACGATTTAGAGCAGACACCTGATGAAGAGGCGAAGAAA
CCTAAGAAGGTCACGATAAGGAAATCAGCGAAGAAATGTTTGAGCACGACCATATTACCAGATTCATT
CAGAGGGGTTTTTAAATTTACAGAATTTAATAAAATGCAATCAGAGGCTTTTCCTAGTATTTATGAGA
GTAACGAGAACTGCATAATTTCTTCACCAACCGGATCAGGCAAGACTGTATTATTTGAATTGGCAATA
TTACGTCTTATAAAGGAAACAAATAGTGATACCAATAACACCAAAATTATATACATCGCGCCAACCAA
ATCTCTATGTTACGAGATGTACAAGAACTGAGTCGAC

COF1-HFM1/1000
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATAAGAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAATTCCAGTCTGCTACATTTAAAAGAAATAAGAAAAATAGCCGCCAA
CCAAGTTTTAAAGTAGGTTTGTCTTATAACAGTTTACTGGATGATTGTGATGATGAAAATGAAACAGA
GGAAATATTTGAGGGCCGTGGTCTGCAATTCTTCGACAAAGACGATAATTTTTCAATAACAGCAGATG
ATACACAAGTAACTAGTAAATTATTTGATCACGATTTAGAGCAGACACCTGATGAAGAGGCGAAGAAA
CCTAAGAAGGTCACGATAAGGAAATCAGCGAAGAAATGTTTGAGCACGACCATATTACCAGATTCATT
CAGAGGGGTTTTTAAATTTACAGAATTTAATAAAATGCAATCAGAGGCTTTTCCTAGTATTTATGAGA
GTAACGAGAACTGCATAATTTCTTCACCAACCGGATCAGGCAAGACTGTATTATTTGAATTGGCAATA
TTACGTCTTATAAAGGAAACAAATAGTGATACCAATAACACCAAAATTATATACATCGCGCCAACCAA
ATCTCTATGTTACGAGATGTACAAGAACTGGTTTCCTTCCTTTGTGAATCTTTCTGTTGGTATGCTTA
CTAGTGATACCTCTTTTCTGGAAACTGAAAAGGCTAAAAAATGTAACATAATAATTACGACACCAGAA
AAATGGGATTTGTTAACAAGAAGATGGTCCGATTACAGCCGATTATTCGAATTGGTCAAACTTGTCCT
AGTTGATGAGATTCATACCATAAAGGAGAAAAGAGGAGCATCTTTGGAAGTAATTCTGACAAGGATGA
ATACAATGTGTCAAAACATTCGGTTTGTTGCTTTAAGTGCAACAGTTCCAAATATAGAAGACCTAGCA
TTGTGGCTCAAAACTAACAACGAGCTTCCTGCAAATATTCTTTCTTTTGACGAATCGTACAGGCAAGT
TCAGTTAACGAAGTTTGTTTATGGGTACTCCTTCAATTGCAAAAATGACTGAGTCGAC

COF1-HFM1/2000
GGATCCTAACAAAAGAAGATGTCTAGATCTGGGTATGCTAAATTTCATTTGTACTCCTGTGTAAACTA
TCAATTACTAACTAATTGACATGATGCTGTTTTATCTCGCTCTCTTATGTCCTTTCTTTCCTCCCTTT
TTAACATGGAACTTTAAAAGGAATGTGGGGGAAAAAATTCATAAGAAGCCAGGAAAATGTGAAGTTTC
TGTGTAGTGTTGCTGTTGCTGAATTCCAGTCTGCTACATTTAAAAGAAATAAGAAAAATAGCCGCCAA
CCAAGTTTTAAAGTAGGTTTGTCTTATAACAGTTTACTGGATGATTGTGATGATGAAAATGAAACAGA

GGAAATATTTGAGGGCCGTGGTCTGCAATTCTTCGACAAAGACGATAATTTTTCAATAACAGCAGATG
ATACACAAGTAACTAGTAAATTATTTGATCACGATTTAGAGCAGACACCTGATGAAGAGGCGAAGAAA
CCTAAGAAGGTCACGATAAGGAAATCAGCGAAGAAATGTTTGAGCACGACCATATTACCAGATTCATT
CAGAGGGGTTTTTAAATTTACAGAATTTAATAAAATGCAATCAGAGGCTTTTCCTAGTATTTATGAGA
GTAACGAGAACTGCATAATTTCTTCACCAACCGGATCAGGCAAGACTGTATTATTTGAATTGGCAATA
TTACGTCTTATAAAGGAAACAAATAGTGATACCAATAACACCAAAATTATATACATCGCGCCAACCAA
ATCTCTATGTTACGAGATGTACAAGAACTGGTTTCCTTCCTTTGTGAATCTTTCTGTTGGTATGCTTA
CTAGTGATACCTCTTTTCTGGAAACTGAAAAGGCTAAAAAATGTAACATAATAATTACGACACCAGAA
AAATGGGATTTGTTAACAAGAAGATGGTCCGATTACAGCCGATTATTCGAATTGGTCAAACTTGTCCT
AGTTGATGAGATTCATACCATAAAGGAGAAAAGAGGAGCATCTTTGGAAGTAATTCTGACAAGGATGA
ATACAATGTGTCAAAACATTCGGTTTGTTGCTTTAAGTGCAACAGTTCCAAATATAGAAGACCTAGCA
TTGTGGCTCAAAACTAACAACGAGCTTCCTGCAAATATTCTTTCTTTTGACGAATCGTACAGGCAAGT
TCAGTTAACGAAGTTTGTTTATGGGTACTCCTTCAATTGCAAAAATGACTTCCAAAAAGATGCTATAT
ATAATTCCAAATTGATTGAAATAATTGAAAAGCATGCCGATAATCGTCCCGTACTAATATTTTGTCCG
ACTAGGGCTTCAACTATATCAACAGCAAAATTTTTATTAAATAATCATATTTTTTCAAAAAGTAAGAA
AAGATGTAATCATAATCCCTCCGATAAAATATTAAATGAATGTATGCAACAAGGTATCGCTTTCCATC
ATGCTGGAATTTCTTTAGAAGACCGTACTGCCGTTGAGAAAGAATTTCTTGCAGGTTCAATTAATATA
TTGTGTTCAACTTCAACGTTAGCTGTTGGTGTGAACCTTCCCGCATATTTGGTCATTATCAAAGGCAC
CAAAAGTTGGAACTCTTCTGAAATCCAAGAGTATTCAGACCTAGATGTACTGCAAATGATTGGGCGGG
CAGGAAGACCTCAATTTGAAACTCACGGGTGCGCAGTAATCATGACAGACTCTAAAATGAAGCAAACA
TATGAAAACTTAATTCATGGAACTGACGTTTTGGAAAGTTCATTACATTTGAATTTGATCGAGCATTT
AGCAGCAGAAACCTCTTTGGAAACTGTATATTCAATTGAAACTGCAGTAAATTGGCTACGAAACACAT
TCTTTTACGTTAGATTTGGGAAGAATCCAGCTGCATATCAAGAGGTGAACAGATACGTTAGTTTTCAT
TCCGTTGAAGACTCTCAGATAAATCAATTCTGTCAGTATTTGCTTGACACCTTAGTAAAAGTGAAAAT
AATCGATATCAGCAATGGGGAGTACAAATCTACAGCGTACGGAAATGCAATGACGCGACATTATATAT
CCTTTGAGTCGATGAAACAATTCATTAACGCAAAAAAATTTCTATCATTACAAGGGATACTAAATTTA
CTGGCCACTTCTGAAGAGTTTTCAGTTATGAGGGTGAGGCACAATGAAAAAAAATTGTTCAAAGAAAT
CAATTTATCGCCACTTTTGAAATATCCATTTTGAGTCGAC

The PhD thesis is accompanied by a CD containing following materials:

- A copy of the thesis in the PDF format
- A summary of the thesis in English and in Czech
- Sequences of splicing reporters used in the study
- Author´s curriculum vitae
- Electronic versions of all articles related to the thesis, including supplementary files:

Publication 1:
Gahura O., Abrhámová K., Skružný M., Valentová A., Munzarová V., Folk P., Půta F.
Prp45 affects Prp22 partition in spliceosomal complexes and splicing efficiency of non-consensus substrates.
J Cell Biochem 2009, 106(1):139-151 (IF 3.381)

Publication 2:
Gahura O., Hammann Ch., Valentová A., Půta F., Folk P.
Secondary structure is required for 3' splice site recognition in yeast.
Nucleic Acids Research, in press (IF 7.836)

Publication 3:
Wang X., Wu F., Xie Q., Wang H., Wang Y., Yue Y., Gahura O., Ma S., Liu L., Cao Y., Jiao Y., Puta F., McClung C. R., Xu X., Ma L.
SKIP is a splicing factor linking alternative splicing and circadian clock in *Arabidopsis*
Plant Cell, manuscript submitted (IF 9.396)