Structure–function analysis and genetic interactions of the yeast branchpoint binding protein Msl5

Jonathan Chang[1], Beate Schwer[2] and Stewart Shuman[1],*

[1]Molecular Biology Program, Sloan-Kettering Institute and [2]Microbiology and Immunology Department, Weill Cornell Medical College, New York, NY 10065, USA

ABSTRACT

Saccharomyces cerevisiae Msl5 (branchpoint binding protein) orchestrates spliceosome assembly by binding the branchpoint sequence 5′-UACUAAC and establishing cross intron-bridging interactions with other components of the splicing machinery. Reciprocal tandem affinity purifications verify that Msl5 exists *in vivo* as a heterodimer with Mud2 and that the Msl5–Mud2 complex is associated with the U1 snRNP. By gauging the ability of mutants of Msl5 to complement *msl5Δ*, we find that the Mud2-binding (amino acids 35–54) and putative Prp40-binding (PPxY[100]) elements of the Msl5 N-terminal domain are inessential, as are the C-terminal proline-rich domain (amino acids 382–476) and two zinc-binding CxxCxxxxHxxxxC motifs (amino acids 273–286 and 299–312). A subset of conserved branchpoint RNA-binding amino acids in the central KH-QUA2 domain (amino acids 146–269) are essential pairwise (Ile198–Arg190; Leu256–Leu259) or in trios (Leu169–Arg172–Leu176), whereas other pairs of RNA-binding residues are dispensable. We used our collection of viable Msl5 mutants to interrogate synthetic genetic interactions, in *cis* between the inessential structural elements of the Msl5 polypeptide and in *trans* between Msl5 and yeast splicing factors (Mud2, Nam8 and Tgs1) that are optional for vegetative growth. The results suggest a network of important but functionally buffered protein–protein and protein–RNA interactions between the Mud2–Msl5 complex at the branchpoint and the U1 snRNP at the 5′ splice site.

INTRODUCTION

Regulated pre-mRNA splicing figures prominently in the control of eukaryal gene expression and defects in pre-mRNA splicing underlie many human diseases (1,2).

The yeast *Saccharomyces cerevisiae* has been a powerful model system to elucidate the composition and function of the core splicing machinery and the dynamic changes in the spliceosome that orchestrate splicing chemistry and fidelity (3–5). The yeast spliceosome transits through multiple steps—assembly, activation, catalysis, and disassembly—mediated by five snRNPs (U1, U2, U4, U5, U6) and scores of proteins that interact with the snRNPs and the pre-mRNA. The spliceosomal snRNAs are all essential for yeast viability. By contrast, many of the yeast splicing proteins are inessential for vegetative growth, especially factors implicated in early steps of spliceosome assembly. Classical genetics and synthetic genetic array analyses (6–12) have highlighted networks of mutational synergies between the inessential yeast splicing factors, whereby combining two viable null mutations results in synthetic lethality or 'sickness'. Such genetic interactions among individually dispensable players in a common pathway (splicing) meets an operational definition of redundancy (i.e. 'the provision of additional or duplicate systems, equipment, etc., that function in case an operating part or system fails'). Genetic redundancy does not necessitate that the synthetic interactor proteins perform the identical task, but rather that the process (e.g. spliceosome assembly) can be accomplished with the aid of either protein, conceivably via different microscopic sub-pathways.

This scenario is exemplified by the genetics of Tgs1, the enzyme responsible for synthesis of the signature 2,2,7-trimethylguanosine (TMG) cap structure of the U1, U2, U4 and U5 snRNAs (13,14). Tgs1 catalyzes two successive methyltransfer reactions from AdoMet to the N2 atom of 7-methylguanosine nucleotides via a distributive mechanism (10,14–16). The TMG cap is not essential for viability of eukaryal cells, insofar as a *tgs1Δ* mutant of fission yeast grows normally (17). The *tgs1Δ* mutation of budding yeast causes a growth defect at cold temperatures, though *tgs1Δ* cells grow as well as *TGS1* cells at 34–37°C (10,13). The *tgs1Δ* mutants of budding and fission yeast lack any detectable TMG caps on their U1, U2, U4 and U5 snRNAs, as gauged by IP-Northerns using an anti-TMG antibody (13,17),

signifying that there is no Tgs1-independent route to generate TMG caps. *tgs1Δ* yeast cells have apparently normal steady-state levels of snRNAs, and they display no overt aberrations in the RNA or protein contents of their spliceosomal snRNPs, except for the acquisition of the nuclear m[7]G cap-binding complex as a stoichiometric component of the U1 snRNP (13,18).

That fungi grow in the absence of Tgs1 suggested there might be backup mechanisms to ensure the function of the U snRNAs when the TMG modification is missing. This idea was confirmed by synthetic genetic array analysis in budding yeast, which revealed that the effects of ablating the TMG cap are buffered by spliceosome assembly factors that are themselves inessential for vegetative growth (12,13). The strongest genetic interactor with Tgs1 is Mud2. Whereas ablation of Mud2 *per se* has no apparent effect on yeast cell growth, the *tgs1Δ mud2Δ* combination resulted in unconditional synthetic lethality (13). Mud2 was initially identified by the Rosbash lab in a 'MUD screen' (mutant U1 die) for yeast mutations that cause synthetic lethality with otherwise viable mutations in the U1 snRNA (19). Mud2, a 527-amino acid polypeptide composed of a hydrophilic N-terminal domain and a C-terminal RRM3 domain, is the yeast homolog of human splicing factor U2AF65. Mud2 exists as a heterodimer with Msl5, the essential yeast homolog of mammalian branchpoint binding protein SF1 (20). Mud2 interacts with the pre-mRNA/U1 snRNP complex, in a manner that depends on the branchpoint sequence of the intron and the association of Mud2 with Msl5. Msl5 then facilitates recruitment of the U2 snRNP, binding of which displaces Mud2/Msl5 from the branchpoint (19–23). The synthetic lethality of *tgs1Δ* and *mud2Δ* mutations implies that spliceosome assembly during vegetative growth depends on *either* Mud2 *or* a TMG cap on one or more of the U snRNAs.

Mud2 interacts genetically with many other spliceosome assembly factors by the criteria of synthetic lethality and synthetic sickness. These include the U1 snRNP subunits Mud1 and Nam8, the nuclear cap-binding complex (implicated in cap-dependent recruitment of U1 snRNP to the 5′ splice site), the U2 snRNP components Lea1 and Msl1, and splicing factors Cwc21, Swm2 and Swt21 (6,7,9,11,12,16,24). We recently exploited the synthetic lethal *mud2Δ* interactions to delineate functional domains and structure-activity relations in yeast Mud2 (16). We found that the C-terminal RRM3 domain is necessary but not sufficient for Mud2 function in complementing growth of *tgs1Δ mud2Δ* and *mud1Δ mud2Δ* strains. Other changes in Mud2 elicited distinct phenotypes in *tgs1Δ* versus *mud1Δ* backgrounds. For example: (i) alanine-cluster mutation of the RNP2 and RNP1 motifs ([424]LLLLNC and [478]AGNIYIKF) of the RRM3 domain was lethal in *tgs1Δ*, but not in *mud1Δ*; (ii) the segment from amino acids 112–172 was critical for Mud2 function in *tgs1Δ*, but not in *mud1Δ*. The requirements for Mud2 function are apparently more stringent when yeast cells lack TMG caps on U snRNAs than when they lack Mud1 in the U1 snRNP.

The physical association of the inessential Mud2 protein with the essential Msl5 (20) raises the question of whether *mud2* phenotypes are mediated via Msl5. Msl5 is a 476-amino acid polypeptide composed of a central branchpoint RNA binding domain (amino acids 147–331, embracing KH, QUA2 and Zn knuckle modules; see Figure 1) flanked by an N-terminal domain that binds Mud2, and N-terminal and C-terminal proline-rich motifs that might contribute to the interaction of Msl5 with the Prp40 subunit of the U1 snRNP (21,22,25–27). Initial studies showed that a 2-hybrid clone expressing the Gal4 activation domain fused to Msl5-(10–362) was able to complement growth of an *msl5Δ* mutant, whereas a Gal4 fusion to Msl5-(10–183) could not (22). An N-terminal Msl5 deletion mutant lacking the first 56 amino acids was unable to bind stably to Mud2, but nonetheless complemented *msl5Δ* when the '*bbpΔ56*' allele was integrated into the yeast chromosome (20). These results indicate that neither the binding of Mud2 to Msl5 nor the putative interactions of the Msl5 proline-rich C domain are essential *per se* for vegetative growth. Here we map at higher resolution the margins of the active Msl5 protein and we then show that the N and C domains of Msl5 are genetically redundant.

We examine the issue of whether Mud2 phenotypes are Msl5-mediated by surveying the effects of Msl5 mutations in a *mud2Δ* background. The rationale here is that *mud2Δ* synthetic phenotypes resulting from absence of Mud2 from the Msl5-Mud2 heterodimer ought to be phenocopied by Msl5 mutations that efface its Mud2 binding site. We also interrogate the effects of Msl5 mutations in the putative Prp40-binding motif and the KH, QUA2 and Zn knuckle modules on complementation of *msl5Δ* and *msl5Δ tgs1Δ* and *msl5Δ nam8Δ* yeast strains. Our results reveal a rich network of intramolecular and intermolecular genetic interactions of the yeast branchpoint binding protein.

MATERIALS AND METHODS

Tandem affinity purification and peptide analysis

Yeast *MUD2-TAP* and *MSL5-TAP* strains were purchased from Open Biosystems. Tandem-affinity purifications from whole cell extracts (derived from 6-l yeast cultures grown in YPD medium to A_{600} of 2.0–2.5) were performed as described (18). The polypeptides in the MUD2-TAP and MSL5-TAP preparations were resolved by SDS–PAGE; the lanes were sliced horizontally into 10 segments. *In situ* trypsin digestion of polypeptides in each gel slice, LC-MS/MS analysis and protein/peptide identifications from the LC-MS/MS data were performed as described in detail elsewhere (18). Total numbers of peptide spectra assigned to specific yeast polypeptides (obtained by merging information obtained from all the gel slices in the lane) are compiled in Table 1.

Msl5 mutants

CEN LEU2 plasmids bearing wild-type or mutated *MSL5* genes under the control of the native *MSL5* promoter were constructed as follows. First, we made plasmid pRS415-MSL5(5′3′), a derivative of the yeast shuttle
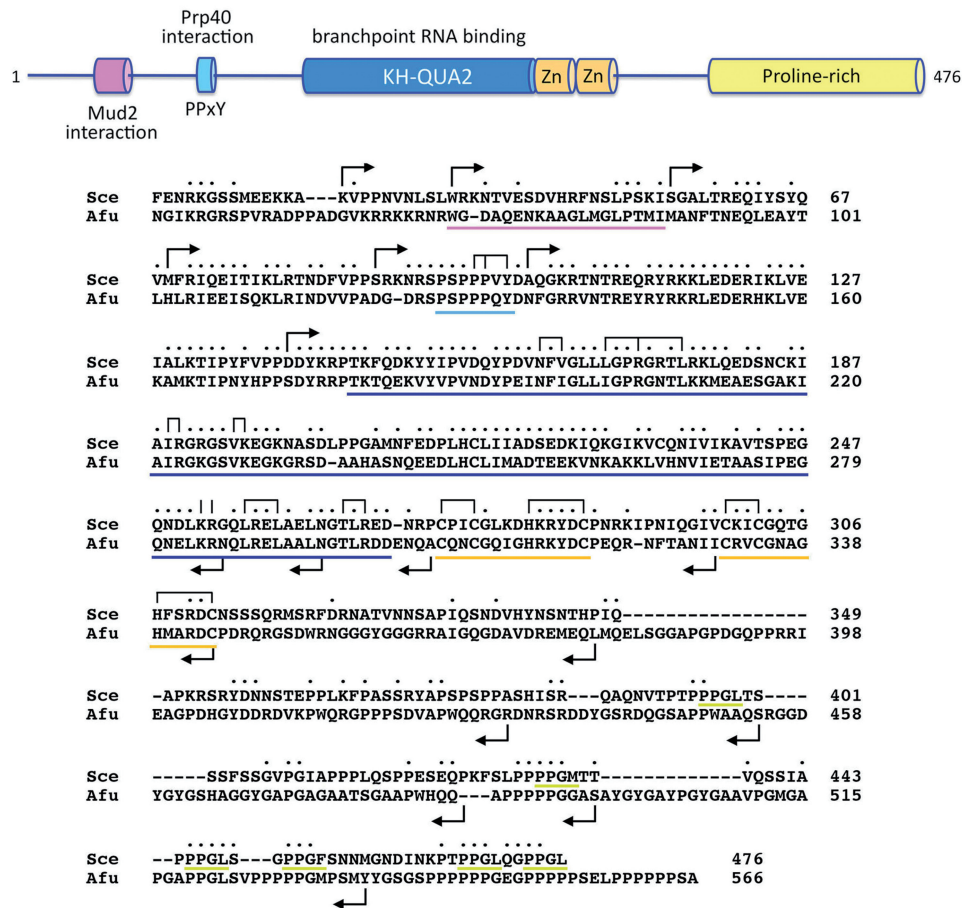
Figure 1. Domain organization and primary structure of yeast Msl5. The 476-amino acid Msl5 polypeptide is depicted in the top panel as a linear array with the N-terminus at left and the C-terminus at right and the known or imputed domains drawn as cylinders spanning their segments of the primary structure. The amino acid sequence of *S. cerevisiae* Msl5 (Sce; Genbank accession NP_013217) from amino acids 11–476 is aligned to that of the homologous 566-amino acid polypeptide from *Aspergillus fumigatus* (Afu; Genbank accession XP_754535) in the bottom panel. Positions of amino acid site chain identity/similarity are denoted by dots above the sequence. The segments corresponding to the various domains are underlined according to the color scheme in the top panel. Gaps in the alignment are denoted by dashes. Forward and reverse arrowheads indicate the boundaries of the N-terminal and C-terminal truncations, respectively. Vertical lines and brackets above the sequence signify the amino acids subjected to single-alanine or alanine-cluster mutagenesis.

vector pRS415 (*CEN LEU2*) with a genomic DNA segment spanning the 509-bp immediately upstream of the *MSL5* ORF cloned between the XhoI and BamHI sites and a genomic DNA segment comprising the 264-bp segment immediately downstream of the *MSL5* ORF inserted between the NotI and SacI sites. The wild-type *MSL5* ORF was PCR-amplified with primers that introduced a BamHI site preceding the start codon and a NotI site following the stop codon. N-terminal truncation mutants *MSL5-(25–476)*, *MSL5-(35–476)*, *MSL5-(55–476)*, *MSL5-(69–476)*, *MSL5-(88–476)*, *MSL5-(102–476)* and *MSL5-(140–476)* were PCR-amplified with sense-strand primers that introduced a BamHI site and a methionine codon in place of the codons for Ala24, Leu34, Ile54, Val68, Pro87, Asp101 or Pro139. The C-terminal truncation mutants *MSL5-(1–458)*, *MSL5-(1–437)*, *MSL5-(1–425)*, *MSL5-(1–401)*, *MSL5-(1–381)*, *MSL5-(1–347)*, *MSL5-(1–312)*, *MSL5-(1–298)*, *MSL5-(1–272)*, *MSL5-(1–263)* and *MSL5-(1–254)* were PCR-amplified with anti-sense

strand primers that introduced a stop codon (and a flanking NotI site) in lieu of the codons for Gly459, Val438, Lys426, Ser402, Ser382, Ile348, Asn313, Cys299, Cys273, Gly264 or Gln255. Combined N- and C-terminal truncation mutants were PCR-amplified with the appropriate primers. Alanine mutations were introduced into *MSL5* and *MSL5-(55–476)* by two-stage PCR overlap extension using mutagenic primers. The various PCR products were digested with BamHI and NotI and then inserted between the BamHI and NotI sites of pRS415-MSL5(5′3′). The inserts of the resulting pRS415-MSL5 plasmids were sequenced completely to exclude the presence of unwanted coding changes.

Assay of Msl5 activity *in vivo*

We developed plasmid shuffle assays to gauge mutational effects on Msl5 activity in *msl5Δ*, *tgs1Δ msl5Δ*, *mud2Δ msl5Δ* and *nam8Δ msl5Δ* strains that rely for viability on maintenance of *MSL5* on a *URA3* plasmid. To establish the complementation assay in the *msl5Δ* background,

Table 1. Tandem affinity purified Mud2 and Msl5

	Proteins	Mud2-TAP (# spectra)	Msl5-TAP (# spectra)
Mud2–Msl5 complex	Mud2	**407**	467
	Msl5	289	**229**
U1 snRNP	Snu71	7	7
	Snp1	9	11
	Snu56	13	20
	Luc7	8	13
	Prp40	8	11
	Prp39	6	4
	Mud1	7	10
	Nam8	3	9
	Yhc1	3	4
	Prp42	1	3
	SmD1	3	4
	SmD2	2	3
	SmD3	0	2
	SmB	0	2
Others	Prp43	2	6
	Ubp3	40	29
	Bre5	14	18
	Ebs1	29	25
	Pab1	19	62
	Pbp1	20	17
	Ded1	20	23
	Ecm16	12	10
	Lsm12	12	10
	Yra1	3	11
	Set2	15	20
	Nop14	3	5
	eIF4G	2	6
	Pat1	3	2
	Npl3	1	7
	Xrn1	3	8
	Sto1	0	3
	Cdc33	0	2
	Smy2	0	9

we transformed a heterozygous *MSL5 msl5::kanMX* diploid with p360-MSL5, a *CEN URA3* plasmid in which expression of *MSL5* is driven by a yeast *TPI1* promoter. The diploid was sporulated, asci were dissected and haploid *msl5Δ* Ura$^+$ progeny were recovered. The *msl5Δ* Ura$^+$ cells were resistant to geneticin and unable to grow on medium containing FOA (5-fluoroorotic acid, a drug that selects against the *URA3* plasmid).

To assay complementation in the double-mutant backgrounds, we generated a heterozygous *TGS1 tgs1::natMX MSL5 msl5::kanMX* diploid, a heterozygous *MUD2 mud2::natMX MSL5 msl5::kanMX* diploid, and a heterozygous *NAM8 nam8::natMX MSL5 msl5::kanMX* diploid by crossing the respective haploid deletion strains with the *msl5Δ* p360-MSL5 strain. The diploids were sporulated, asci were dissected, and haploid *tgs1Δ msl5Δ* Ura$^+$, *mud2Δ msl5Δ* Ura$^+$ and *nam8Δ msl5Δ* Ura$^+$ progeny were recovered that were resistant to both geneticin and clonNAT and unable to grow on medium containing FOA.

For plasmid shuffle assays, the test strains were transformed with the pRS415-MSL5 plasmids. Individual Leu$^+$ transformants were streaked on agar medium containing 1 mg/ml FOA. Growth was scored after incubation for 7 days at 18, 25, 30 or 37°C. Lethal mutants were those that

failed to form colonies at any temperature. Individual FOA-resistant colonies with viable *MSL5* alleles were grown to mid-log phase in YPD broth and adjusted to A_{600} of 0.1. Aliquots (3 μl) of serial 10-fold dilutions were spotted on YPD agar plates, which were then incubated at 18, 25, 30 and 37°C. (The *tgs1Δ* strains were tested for growth on YPD agar at 25, 30 and 37°C; *tgs1Δ* cells are inviable at 18°C.) Growth was assessed as follows: (+++) colony size was indistinguishable from strains bearing wild-type MSL5; (++) slightly reduced colony size; (+) only pinpoint macroscopic colonies were formed; (±) only microscopic colonies were formed; (−) no growth.

RESULTS

Tandem affinity purification of Mud2 and Msl5 confirms their association *in vivo* and highlights their interaction with the U1 snRNP

We used the tandem affinity purification (TAP) method to isolate Mud2- and Msl5-containing complexes from soluble whole-cell extracts of *MUD2–TAP* and *MSL5–TAP* yeast cells (18,28). The protein contents of the samples were analyzed in parallel by SDS–PAGE, then digested *in situ* with trypsin. The resulting peptides were analyzed by liquid chromatography-coupled tandem mass spectrometry and correlated to specific peptide fragments identified in the *S. cerevisiae* proteome. The total numbers of peptide spectra assigned to individual yeast proteins with >95% confidence were compiled and are presented in Table 1. (We omitted from this compilation any ribosomal proteins, heat-shock proteins, protein chaperones and metabolic enzymes.) Mud2 and Msl5 had the highest spectral counts in both TAP preparations; indeed the numbers of Mud2-derived spectra were similar in the Mud2-TAP and Msl5-TAP samples (407 and 467, respectively) as were the numbers of Msl5-derived spectra (289 and 229 in the Mud2-TAP and Msl5-TAP samples, respectively) (Table 1). These findings are in accord with the previously reported analyses of ^{35}S-methionine-labeled Mud2-TAP and Msl5-TAP preparations (20) and provide further evidence that Mud2 and Msl5 comprise a stable complex *in vivo*. [Also in agreement with previous studies (20), we found that heat shock protein Ssb2 and ribosomal protein Rpl4 were the next most abundant sources of peptide spectra in both TAP samples: Mud2-TAP yielded 147 Ssb2 spectra and 69 Rpl4 spectra; Msl5-TAP yielded 132 Ssb2 spectra and 79 Rpl4 spectra.]

Both Mud2-TAP and Msl5-TAP were associated with all 10 of the U1-specific protein subunits of the U1 snRNP (Prp39, Prp40, Snu71, Snu56, Snp1, Mud1, Luc7, Prp42, Nam8 and Yhc1) and with several subunits of the Sm protein ring. By contrast, we detected no peptides in either TAP preparation derived from the U2 snRNP subunits Lea1, Rse1, Hsh155, Cus1 and Msl1 or from the U5 snRNP subunits Prp8 and Snu114. These findings suggest that affinity purified yeast Msl5-Mud2 heterodimer is specifically associated with the U1 snRNP, conceivably because the TAP purification recovered early intermediates in spliceosome assembly

that contained U1 snRNP at the 5′ splice site and Msl5-Mud2 at the branchpoint.

Also compiled in Table 1 under the category of 'Others' are sundry non-ribosomal proteins implicated in RNA transactions or RNA binding that were detected in the Mud2-TAP and Msl5-TAP preparations. We include this information for the sake of transparency and for those who might be interested in these yeast proteins. Note that these 'others' were also present in TAP preparations of bulk Sm-containing U snRNPs, isolated U1 and U2 snRNPs, and C/D and H/ACA snoRNPs (18), often with spectral counts similar to those reported in Table 1, which might mean that they are associated generically with RNAs in yeast ribonucleoproteins.

Effects of structure-guided mutations in the putative Mud2–Msl5 interface

The C-terminal RRM3 domains of human U2AF65 and yeast Mud2 are implicated in protein–protein interactions with branchpoint binding protein rather than, or in

addition to, RNA binding (22,29,30). RRM3-like domains, renamed U2AF homologs motifs (UHMs), have been identified in many other proteins, where they are proposed to comprise a protein interaction module (31). We exploited the NMR structure of a binary complex of human U2AF65 RRM3 bound to a peptide ligand [13]PSKKRKRSRWNQD[25] derived from the N-terminal domain of SF1 (30) (Figure 2A) to interrogate the homologous Mud2–Msl5 interface in yeast. The RRM3 fold comprises a four-stranded β-sheet flanked on one side by two α-helices. The central β1 and β3 strands of the sheet (referred to as the RNP2 and RNP1 motifs, respectively) typically make atomic contacts with bound RNA in classical RRM–RNA complexes (32), but this surface of the β-sheet is occupied in the U2AF65-SF1 complex by a C-terminal α-helix (Figure 2B). SF1 binds along the opposite side of the RRM3 fold, principally by insertion of a conserved SF1 tryptophan side chain (Trp22 in human SF1; Trp35 in yeast Msl5) into a hydrophobic pocket of U2AF65 formed by the 'back' side of the β-sheet
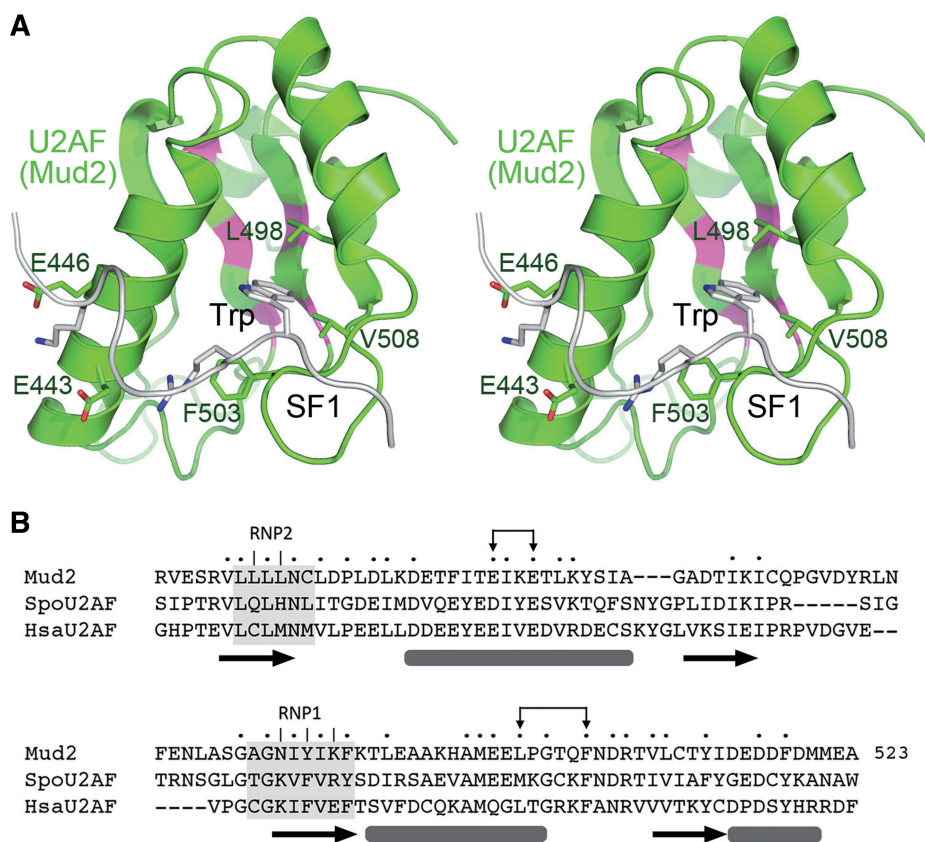

Figure 2. Structure-guided mutations of the Mud2-Msl5 interface. (**A**) Stereo view of the crystal structure of the RRM3 domain of human U2AF65 (colored green) bound to a peptide ligand derived from human SF1 (colored gray) as reported in pdb 1OPI (30). The human RRM3 domain is homologous to the RRM3 domain of yeast Mud2. Amino acids in the RRM3 domain that comprise the SF1 binding site are shown as stick models and numbered according to their equivalent positions in Mud2. The tryptophan side chain (Trp22) of human SF1 peptide that is critical for binding to human U2AF65 is shown as a gray stick model. The putative equivalent position in yeast Msl5 is Trp35. (**B**) Alignment of the primary structures of the RRM3 domains of *S. cerevisiae* Mud2 (Accession AAA64215), *S. pombe* U2AF59 (Accession NP_595396) and *Homo sapiens* U2AF65 (Accession P26368). Positions of side chain identity/similarity in all three proteins are indicated by dots above the alignment. The secondary structure elements are depicted below the alignment (arrows for β strands and bars for α helices). The RNP2 and RNP1 motifs are highlighted in gray boxes. The five RNP2 and RNP1 residues of Mud2 that were mutated previously to alanine (16) are indicated by vertical lines above the alignment. [The equivalents in human U2AF65 are colored magenta in (A)]. The putative branchpoint binding protein-interacting residues of Mud2—Glu443, Glu445, Leu498 and Phe503—that were subjected to mutation, pair-wise and *en masse*, are denoted by bracketed arrows.

and the two α-helices (Figure 2A). The U2AF65 side chains Glu397, Glu400, Leu449, Phe454 and Val459 that form the SF1 binding site are conserved in Mud2 (as Glu443, Glu446, Leu498, Phe503 and Val508) and are numbered according to the Mud2 counterparts in Figure 2A. Here we tested the effects of two double-alanine mutations E443A-E446A and L498A-F503A, plus a quadruple-alanine change E443A-E446A-L498A-F503A on Mud2 activity in two synthetic lethal genetic backgrounds—*tgs1Δ mud2Δ* and *mud1Δ mud2Δ*—in which cell viability requires a functional Mud2 protein. The mutations were introduced into the fully active *MUD2-(41–527)* gene (16) carried on a *CEN* plasmid under the control of the native *MUD2* promoter and the 'wild-type' and mutant alleles were tested for complementation by plasmid shuffle, after which the viable FOA-resistant strains were tested for growth on YPD agar at 18, 25, 30 and 37°C. The key findings were that neither of the double-alanine changes in the putative Msl5 interface affected Mud2 activity, as gauged by colony size compared to the 'wild-type' Mud2 control (Supplementary Table S1). Indeed, even a double charge inversion mutation—*E443K-E446K*—did not affect complementation (Supplementary Table S1). The quadruple-alanine mutant *E443A-E446A-L498A-F503A* retained complementation activity in both synthetic lethal strains, albeit with modest hypomorphic phenotypes, whereby *E443A-E446A-L498A-F503A mud1Δ* cells formed smaller colonies at 37°C (scored as ++ growth) and *E443A-E446A-L498A-F503A tgs1Δ* cells displayed ++ growth at 25, 30 and 37°C (Supplementary Table S1).

To further probe the interface, we tested the effect of an alanine mutation at Trp35 of yeast Msl5 on Msl5's function *in vivo*, as gauged by complementation of the otherwise inviable *msl5Δ* mutant strain. The *W35A* allele (and all other *MSL5* mutants tested in this study) was installed on a *CEN* plasmid under the control of the native *MSL5* promoter and transformed into a *msl5Δ* p(*CEN URA3 MSL5*) strain. Viable FOA-resistant *W35A* cells were recovered after plasmid shuffle and grew as well as wild-type *MSL5* cells on YPD agar at 18, 25, 30 and 37°C, as gauged by colony size (Table 2). If Trp35 is essential for Msl5 interaction with Mud2, and Mud2 activity is contingent on this interaction, then we might expect the *MSL5-W35A* allele to mimic *mud2Δ* with respect to its synthetically lethality with *tgs1Δ*. However, this was not the case (Table 2).

Taken together, these results suggest that either the interaction of Mud2 with Msl5 is not significantly affected by the structure-guided mutations (e.g. because the interface of the yeast proteins *in vivo* is different or more complex than that revealed by the NMR structure of the human RRM3 domain and the SF1-derived peptide) or that the binding of Mud2 to Msl5 is not essential for Mud2 activity in the *mud1Δ* and *tgs1Δ* backgrounds.

Mapping the proximal margins of the functional Msl5 protein

We used a primary structure alignment of *S. cerevisiae* and *Aspergillus fumigatus* branchpoint binding proteins (Figure 1) to guide a deletional analysis of the Msl5 polypeptide. A series of N-terminal truncation alleles were tested for *msl5Δ* complementation. We found that deleting 24, 34 or 54 amino acids from the N-terminus had no apparent impact on yeast growth. [This agrees with the viability of the '*bbp-Δ56*' mutant reported previously (20).] Truncations of 68 or 87 amino acids resulted in moderately slower growth (scored as ++ in Table 2). Extending the deletions to 101 and 139 amino acids (i.e. to the proximal border of the Msl5 KH domain) resulted in severe conditional lethality, whereby *msl5-(102–476)* and

Table 2. Effects of N- and C-terminal mutations on Msl5 functions *in vivo*

MSL5 allele	msl5Δ complementation				msl5Δ tgs1Δ complementation		
	18°	25°	30°	37°	25°	30°	37°
WT	+++	+++	+++	+++	+++	+++	+++
W35A	+++	+++	+++	+++	+++	+++	+++
25–476	+++	+++	+++	+++	+++	+++	+++
35–476	+++	+++	+++	+++	+++	+++	+++
55–476	+++	+++	+++	+++	–	–	–
69–476	+++	++	++	++	–	–	–
88–476	++	++	++	++	–	–	–
102–476	–	–	–	++	–	–	–
140–476	–	–	–	++	–	–	–
1–458	+++	+++	+++	+++	++	+++	+++
1–437	+++	+++	+++	+++	–	++	+++
1–425	+++	+++	+++	+++	–	–	++
1–401	+++	+++	+++	+++	–	–	–
1–381	+++	+++	+++	+++	–	–	–
1–347	+++	+++	+++	+++	–	–	–
1–312	+++	+++	+++	+++	–	–	–
1–298	+++	+++	+++	+++	–	–	–
1–272	+	++	++	++	–	–	–
1–263	–	–	–	–	–	–	–
1–254	–	–	–	–	–	–	–

msl5-(140–476) cells grew slowly at 37°C (++) and were unable to form colonies at 18, 25 or 30°C (Table 2). Thus, the segment between amino acids 88 and 101 includes constituents that are either essential *per se* or are required for Msl5 activity when the proximal N-terminal domain is missing.

The Msl5-(55–476) mutant should be unable to form a heterodimeric complex with Mud2 (20). If Mud2 acts via the Msl5-Mud2 heterodimer, then we expect the *MSL5-(55–476)* allele to phenocopy *mud2Δ* with respect to its synthetically lethality with *tgs1Δ*. Indeed, this was the case, insofar as *MSL5-(55–476)* did not complement a *msl5Δ tgs1Δ* strain (Table 2). As one might expect, the even shorter mutants *MSL5-(69–476)* and *MSL5-(88–476)* that complemented *msl5Δ* also displayed synthetic lethality with *tgs1Δ* (Table 2). By contrast, the deletion alleles *MSL5-(35–476)* and *MSL5-(25–476)* were as active as wild-type *MSL5* in the *tgs1Δ* background (Table 2). These findings, and those above, implicate the segment from amino acids 35–54 (but not Trp35 *per se*) as necessary for Msl5 interaction with Mud2.

Distal margins of the functional Msl5 protein

We constructed a series of 11 C-terminal truncation alleles of *MSL5* (depicted in Figure 1) and tested them for *msl5Δ* complementation. *MSL5-(1–298)* and all of the longer versions sustained normal growth at all temperatures tested (Table 2). This signifies that the C-terminal 178-amino acid segment—which includes the proline-rich domain, the distal zinc knuckle and the linker domain connecting them—is dispensable for Msl5 activity. Truncating the C-terminus to amino acid 272, which deleted the proximal zinc knuckle, resulted in slowed growth of the *MSL5-(1–272)* strain at all temperatures tested (Table 2). Further deletions into the QUA2 domain (placing the C-terminus at position 263 or 254) abolished Msl5 activity (Table 2).

The effects of incremental C-terminal deletions were significantly different in a *tgs1Δ* background, whereby truncations to or beyond amino acid 401 were unconditionally lethal (Table 2). Thus, the proline-rich C-terminal segment from amino acids 402–476 becomes essential in the absence of TMG capping of the U snRNAs. Contained within the proline-rich domain are multiple copies of the peptide motif PPG[M/L/F] that mediates protein–protein interactions with GYF-domain proteins such as yeast Smy2 (20,33,34). Lesser C-terminal deletions of Msl5 progressively regained function in the *tgs1Δ* background. For example: *MSL5-(1–425) tgs1Δ* was slow growing at 37°C and inviable at 25 or 30°C; *MSL5-(1–437) tgs1Δ* grew as well as *MSL5 tgs1Δ* at 37°C, was slow growing at 30°C (++), and was inviable at 25°C; and *MSL5-(1–458) tgs1Δ* displayed +++ growth at 30 and 37°C and ++ growth at 25°C (Table 2). The severity of the growth defects in *tgs1Δ* correlated with the number of C-terminal PPG[M/L/F] motifs, with six copies in full-length Msl5, four copies in Msl5-(1–458), two copies in Msl5-(1–437) and just one motif in the synthetic lethal Msl5-(1–401) variant (Figure 1). We considered the

prospect that the synthetic lethal and sick phenotypes of C-terminal Msl5 deletions in a *tgs1Δ* background might reflect the loss of a putative Msl5-Smy2 interaction, in which case we would expect a synthetic genetic interaction between Smy2 and Tgs1. Accordingly, we tested the effects of ablating *SMY2* (an inessential yeast gene) in a *tgs1Δ* background, but observed no synthetic growth phenotype for *smy2Δ tgs1Δ* cells (data not shown). We surmise that the need for the Msl5 C-terminus in the absence of TMG caps is not exerted via Smy2.

Effect of mutating the putative Prp40-binding motif of Msl5

Direct binding of the U1 snRNP protein Prp40 to Msl5 is thought to bridge the 5′ splice site and branchpoint segments of the intron and stabilize the commitment complex, an intermediate in spliceosome assembly (19,34). Tandem WW domains in yeast Prp40 bind to the isolated [94]PSPPPVYDA[102] nonapeptide of yeast Msl5, which contains a PPxY[100] motif recognized by WW domain proteins (27). To probe the role of the Prp40-binding site, we introduced a triple-alanine mutation P97A-P98A-Y100A and tested its effect on Msl5 activity *in vivo*. The finding that *MSL5-(P97A-P98A-Y100A)* cells grew as well as *MSL5* cells at all temperatures (Table 3) implies that a Prp40-Msl5 interaction via the PPxY[100] motif is not essential *per se* in budding yeast under standard laboratory growth conditions, perhaps because there are redundant systems in place to provide the bridging interactions between U1 snRNP and the branchpoint. The inessentiality of the PPxY[100] motif was also noteworthy in light of the fact that an incremental deletion of amino acids 88–101 containing the motif was lethal at 18–30°C (Table 2), which raised the prospect that the loss of the Prp40-binding motif might be buffered by contributions of the upstream Msl5 N-terminal domain. To evaluate this scenario, we introduced the P97A-P98A-Y100A mutation into the otherwise fully active truncated allele *MSL5-(55–476)* that lacks the Mud2-interaction module. Testing *msl5Δ* complementation by plasmid shuffle yielded microscopic colonies of *MSL5-(55–476)-(P97A-P98A-Y100A)* cells that, after amplification on rich medium, grew very poorly on YPD agar at all temperatures (forming only microscopic colonies after ≥7 days; scored as ± growth in Table 4). A similar synthetic genetic interaction was observed when we tested the *msl5Δ* complementation activity of full-length *MSL5-(P97A-P98A-Y100A)* in a *mud2Δ* background (Table 5). *MSL5-(P97A-P98A-Y100A) mud2Δ* cells were viable but slow growing at 37°C (++), formed only microscopic colonies after prolonged incubations at 25 and 30°C (± growth), and failed to grow at 18°C (Table 5). These results suggest functional redundancy in the bridging interactions of Msl5 with the U1 snRNP: either via Prp40 binding to the Msl5 PPxY[100] motif or via interactions of the Msl5(Nterm)-Mud2 heterodimer with one or more U1 snRNP components (34).

Table 3. Effects of PPxY, KH-QUA2 and Zn knuckle mutations on Msl5 functions *in vivo*

MSL5 allele	msl5Δ complementation				msl5Δ tgs1Δ complementation		
	18°	25°	30°	37°	25°	30°	37°
WT	+++	+++	+++	+++	+++	+++	+++
P97A P98A Y100A	+++	+++	+++	+++	±	±	±
N163A V165A	+++	+++	+++	+++	−	−	−
L169A R172A L176A	−	−	−	−	−	−	−
I189A R190A	−	−	−	−	−	−	−
V195A K196A	+++	+++	+++	+++	−	−	++
L256A L259A	−	−	−	−	−	−	−
T265A R267A	++	+++	+++	+++	−	−	+
K252A R253A	++	+++	+++	+++	−	−	−
K252R	+++	+++	+++	+++	−	−	+
R253K	+++	+++	+++	+++	+++	+++	+++
K252R R253K	+++	+++	+++	+++	−	−	+
C273A C276A	+++	+++	+++	+++	+++	+++	+++
H281A C286A	+++	+++	+++	+++	++	+++	++
C299A C302A	+++	+++	+++	+++	+++	+++	+++
H307A C312A	+++	+++	+++	+++	+++	+++	+++

Table 4. Functional redundancies of Msl5 domains

MSL5 allele	msl5Δ complementation			
	18°	25°	30°	37°
55–476	+++	+++	+++	+++
55–458	+++	+++	+++	+++
55–437	+++	+++	+++	+++
55–425	+++	+++	+++	++
55–401	++	+++	++	+
55–381	++	+++	++	+
55–347	++	++	++	+
55–312	+	++	+	−
55–298	−	−	−	−
55–272	−	−	−	−
MSL5-(55–476) plus indicated mutations				
P97A P98A Y100A	±	±	±	±
C273A C276A	+++	+++	+++	+++
H281A C286A	+++	+++	+++	++
C299A C302A	+++	+++	+++	+++
H307A C312A	+	+	+	+

Table 5. Effect of Msl5 mutations on a *mud2Δ* background

MSL5 allele	msl5Δ mud2Δ complementation			
	18°	25°	30°	37°
WT	+++	+++	+++	+++
55−476	+++	+++	+++	+++
1–458	+++	+++	+++	+++
1–437	+++	+++	+++	+++
1–425	+++	+++	+++	+++
1–401	+	+	++	++
1–381	+	+	++	++
1–347	±	+	++	+
1–312	−	−	−	−
P97A P98A Y100A	−	±	±	++
C273A C276A	+++	+++	+++	+++
H281A C286A	++	+++	+++	+++
C299A C302A	+++	+++	+++	+++
H307A C312A	+++	+++	+++	+++

Effects of structure-guided mutations in the KH and QUA2 domains of Msl5

Recognition of the intron branchpoint by branchpoint binding protein is accomplished by an autonomous RNA-binding module composed of a proximal KH domain fused to a distal QUA2 (Quaking homology 2) domain. The NMR structure of human SF1 bound to an RNA (5′-AUACUAACAA) containing the consensus yeast branchpoint sequence 5′-UACUAAC (35) reveals an extensive network of atomic contacts between conserved amino acid side chains (many of them hydrophobic) and the RNA nucleobases and sugars, while the RNA phosphates are surface-exposed and make relatively few protein contacts (Figure 3). Here we used the SF1–RNA structure to guide a mutational analysis of the KH and QUA2 modules of yeast Msl5.

The KH domain (colored green in Figure 3) engages the CUAACAA-3′ segment of the RNA ligand (the branchpoint adenosine is colored yellow in Figure 3). Here we introduced double-alanine or triple-alanine mutations in clusters of conserved KH residues that line the RNA binding groove, these being: N163A-V165A; L169A-R172A-L176A; I189A-R190A and V195A-K196A. The mutant alleles were tested for *msl5Δ* complementation (Table 3). Given that these side chains are identical in SF1 and Msl5 (Figure 3), and that the RNA–protein contacts observed for these side chains in the SF1 structure are to a canonical yeast branchpoint RNA recognition element, it is our presumption that the same RNA contacts are very likely to be made by the equivalent Msl5 side chains, allowing us to interpret mutational outcomes accordingly.

The results provide new insights to which RNA contacts are most important. *MSL5-(N163A-V165A)* and *MSL5-(V195A-K196A)* cells grew as well as *MSL5* cells at all temperatures. Thus, the contacts of Asn163 with the cytosine base, of Val165 and Lys196 with the adenine base preceding the branchpoint adenosine, and of Val195 with the branchpoint adenine and the 3′-terminal
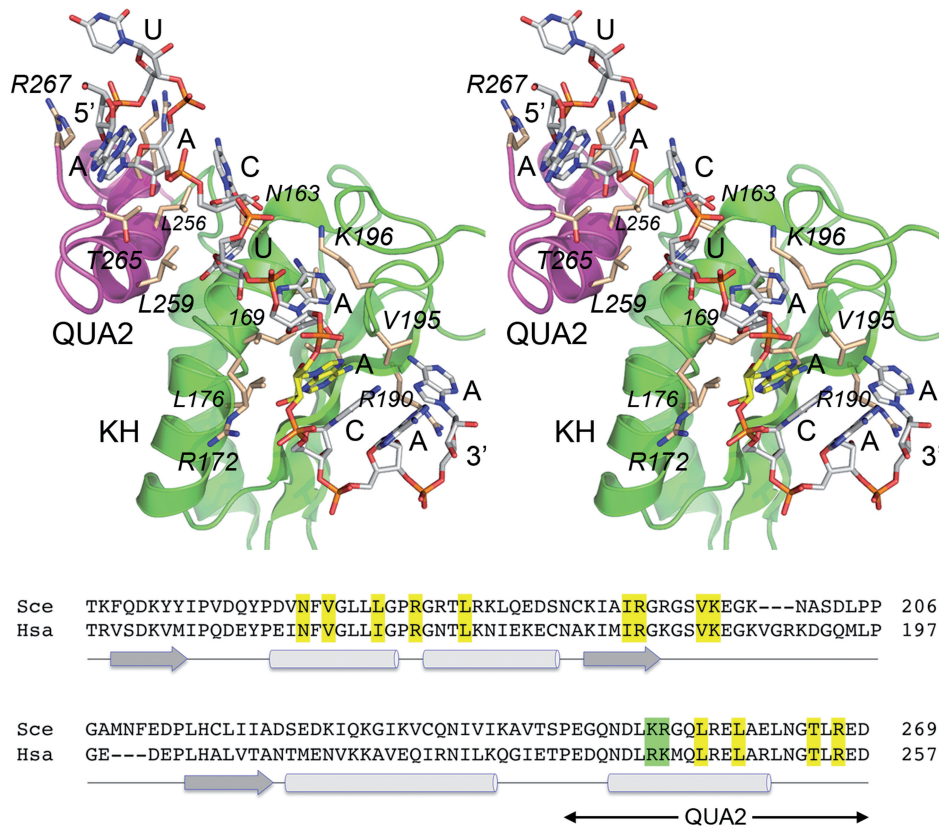
Figure 3. Structure-guided mutations of the KH and QUA2 domains of yeast Msl5. The top panel shows a stereo view of the NMR structure (35; pdb 1K1G) of the KH and QUA2 domains of human SF1, colored green and magenta, respectively, bound to an RNA oligonucleotide 5′-AUACUAACAA that contains the consensus yeast intron branchpoint (underlined). The RNA is rendered as a stick model with gray carbons, except for the branchpoint adenosine (yellow carbons) that transesterifies to the 5′ splice site to form the lariat intermediate. SF1 side chains that comprise the RNA docking site and are conserved in yeast Msl5 are depicted as stick models with beige carbons and are numbered according to their positions in yeast Msl5. In the bottom panel, the amino acids sequences of the KH-QUA2 segments of *S. cerevisiae* (Sce) Msl5 and *Homo sapiens* (Hsa) SF1 (Genbank accession NP_004621) are aligned. The secondary structure elements are depicted below the alignment (arrows for β strands and cylinders for α helices). The conserved residues chosen for mutagenesis are highlighted in yellow boxes. The green box highlights a dipeptide in the QUA2 domain (KR in yeast; RK in human) that was suggested (26) to be a determinant of yeast versus human branchpoint sequence recognition; this dipeptide in Msl5 was subjected to mutagenesis presently.

adenine of the RNA ligand are all inessential for Msl5 function. By contrast, the *L169A-R172A-L176A* and *I189A-R190A* mutations were lethal (Table 3). Leu169 is imputed to intercalate between and make van der Waals contacts to the branchpoint adenine and the upstream flanking adenine (Figure 3). Arg172 contacts the 3′-phosphate of the branchpoint adenosine. Leu176 contacts both the adenine N3 atom and the ribose 2′OH of the branchpoint adenosine. Ile189 contacts the N1 atom of the branchpoint adenine. Arg190 contacts the 3′-terminal adenine base (Figure 3).

The QUA2 domain (colored magenta in Figure 3) binds to the proximal 5′-AUACU segment of the RNA. We tested two double-alanine mutations in the Msl5 QUA2 module, L256A-L259A and T265A-R267A, targeted to side chains that are identical in SF1 and comprise the SF1 QUA2-RNA interface (Figure 3). *MSL5-(T265A-R267A)* cells grew as well as *MSL5* cells at 25–37°C, but were slower growing (++) at 18°C (Table 3). We surmise that the RNA interactions of Thr265 (a van der Waals contact with the ApC phosphate) and Arg267 (with the ribose 2′OH and 3′ phosphate of the

5′-terminal adenosine and with the N6, C5 and N7 atoms of the downstream adenine) are not essential for Msl5 function *in vivo*. By contrast, the *MSL5-(L256A–L259A)* allele was unconditionally lethal (Table 3). Leu256 makes multiple close van der Waals contacts (2.9–3.1 Å) to the ApC dinucleotide, specifically to the adenosine ribose and the cytidine base and ribose (Figure 3). Leu259 makes close van der Waals contacts (3.0 Å) to the uracil base and ribose of the flanking uridine nucleoside (Figure 3).

We also targeted mutations to the [252]KR[253] dipeptide in the QUA2 domain of yeast Msl5, which is the counterpart of the [240]RK[241] dipeptide in human SF1 (Figure 3). Domain swaps and local sequence swaps have implicated the sequence of this simple dipeptide as the determining factor in the specificity of yeast Msl5 for the conserved yeast branchpoint motif 5′-UACUAAC (26). The structure of the complex of human SF1 with the yeast branchpoint sequence does not illuminate how this might be, insofar as Arg240 is located 7.2 Å away from the closest atom in the RNA. However, Lys241 in SF1 does make multiple van der Waals contacts with the

C2, N1 and N6 atoms of the first adenine base in the UAC UAAC motif (35) and this adenosine is a key determinant of the binding affinity of yeast Msl5 for the branchpoint (26). Thus, one might expect that subtracting the equivalent basic residue in yeast Msl5 (Arg253), in tandem with the vicinal Lys252 residue, would affect Msl5 activity *in vivo*. To the contrary, the *MSL5-(K252A–R253A)* allele was as active as wild-type *MSL5* in complementation of *msl5Δ* at 25–37°C (+++ growth) and sustained ++ growth at 18°C (Table 3). We also performed a yeast-to-human dipeptide swap [the reciprocal of the human-to-yeast swap (26)], along with single amino acid swaps, and found that *MSL5-(K252R–R253K)*, *MSL5-(K252R)* and *MSL5-(R253K)* cells grew as well as wild-type *MSL5* cells at all temperatures (Table 3). Thus, our findings do not support an important role for the [252]KR[253] dipeptide for Msl5 function in an otherwise wild-type background.

Functional redundancies of the N- and C-terminal domains of Msl5

The dispensability of the N-terminal 54-amino acid and C-terminal 213-amino acid domains raised the issue of whether they might make genetically redundant contributions to Msl5 activity *in vivo*. To address this question, we deleted the N-terminal 54 amino acids from our series of otherwise viable C-terminal truncation alleles and tested them for *msl5Δ* complementation (Table 4). *MSL5-(55–458)* and *MSL5-(55–437)* cells grew as well as *MSL5* at all temperatures, while *MSL5-(55–425)* cells displayed +++ growth at 18–30°C and ++ at 37°C. Thus, we surmise that the distal segment of the proline-rich domain that includes five of the PPG[M/L/F] motifs is not functionally redundant with the N-terminal Mud2-interaction domain. However, further truncation of the C-terminus to amino acid 401 results in + growth at 37°C and ++ growth at 18 and 30°C (Table 4). The growth defect was exacerbated by deleting amino acids 313–347; such that *MSL5-(55–312)* cells did not grow at 37°C and displayed + growth at 18 and 30°C and ++ growth at 25°C (Table 4). Finally, the *msl5-(55–298)* allele was unconditionally lethal (Table 4). These results imply functional overlap between the N-terminal module and an internal region that includes the distal zinc knuckle. At this point, the basis for the functional overlap is undefined; indeed, because we have not directly measured the intracellular localization and steady state levels of the biologically inactive Msl5-(55–298) protein *in vivo* (such measurements would have to be performed in a yeast strain coexpressing a functional Msl5 protein to sustain cell growth), we cannot rule out that the functional overlap of the N- and C-terminal domains is at the level of Msl5 folding, stability or localization, versus an overlap in directing spliceosome assembly.

If the mutational synergies between the N-terminal domain deletion and the C-terminal deletion series reflects the loss of Mud2 binding to the N-terminus, then we might expect a similar pattern of mutational synergies between the absence of Mud2 and the C-terminal deletion series. To interrogate this scenario, we tested the C-terminal deletions for complementation in an *msl5Δ mud2Δ* background (Table 5). As an initial control, we verified that the *MSL5-(55–476)* allele was as active as wild-type *MSL5* in *msl5Δ mud2Δ* complementation (Table 5), i.e. that loss of the Mud2-binding domain of Msl5 does not synergize with *mud2Δ* (which itself elicits no obvious growth phenotype). We found that *MSL5-(1–425) mud2Δ* cells displayed +++ growth, signifying a lack of functional overlap between Mud2 and the segment of the proline-rich domain that includes five of the PPG[M/L/F] motifs. The effects of further Msl5 truncations in the *mud2Δ* background echoed, but did not exactly replicate, the hierarchy of effects seen in the context of Msl5-(55–476) (compare Tables 4 and 5). For example, the C-terminal truncations to amino acids 401, 381 and 347 elicited growth defects in *mud2Δ* cells that were more severe at 18 and 25°C than at 30 and 37°C (Table 5). The *MSL5-(1–312)* allele, which retains the distal zinc knuckle and is active in *msl5Δ* complementation, was synthetically lethal with *mud2Δ* (Table 5).

Mutational analysis of the zinc knuckles

To focus on the potential roles and genetic interactions of the zinc knuckles, we introduced alanine mutations pairwise in lieu of the putative zinc-binding cysteine and histidine residues (Figure 1). In the context of full-length Msl5, the *C273A–C276A* and *H281A–C286A* mutations that disrupt the proximal knuckle and the *C299A–C302A* and *H307A–C312A* mutations that disrupt the distal knuckle had no impact on *msl5Δ* complementation (Table 3), in keeping with our earlier results that deletion of the distal zinc knuckle in *MSL5-(1–298)* did not affect *msl5Δ* complementation and that *MSL5-(1–272)* cells lacking both zinc knuckles were viable albeit slow-glowing.

We then proceeded to introduce the zinc knuckle alanine mutations into *MSL5-(55–476)* and test them for *msl5Δ* complementation (Table 4). In this context, the knuckle mutations *C273A–C276A* and *C299A–C302A* supported +++ growth at all temperatures, while *H281A–C286A* sustained +++ growth at 18–30°C and ++ growth at 37°C. However, the most distal mutation *H307A–C312A* in the second knuckle displayed severe mutational synergy with the N-terminal deletion, whereby *MSL5-(55–476)-(H307A–C312A)* cells displayed + growth at all temperatures (Table 4). Because the *C299A–C302A* and *H307A–C312A* changes are expected to exert synonymous effects on zinc-binding by the second knuckle, we suspect that the deleterious effects of *MSL5-(55–476)-(H307A-C312A)* are caused by a zinc-independent structural perturbation of this region of the protein.

New mutational synergies between Msl5 PPxY and KH-QUA2 mutations and TMG capping

We tested our collection of Msl5 alanine-mutants for their ability to complement growth in an *msl5Δ tgs1Δ* background. The *MSL5-(P97A-P98A-Y100A)* allele, which was fully active in *msl5Δ* complementation, was extremely feeble in complementing *msl5Δ tgs1Δ*, such that

MSL5-(P97A-P98A-Y100A) tgs1Δ cells formed only microscopic colonies after prolonged incubations at 25, 30 and 37°C (± growth; Table 3). Thus, mutation of the Prp40-binding PPxY^{100} motif of Msl5 elicited strong synthetic phenotypes in three different backgrounds: (i) *tgs1Δ*; (ii) *mud2Δ* and (iii) *MSL5-(55–476)*, wherein the Mud2–Msl5 interaction is ablated. These results fortify the already substantial genetic connections between the Msl5(N-term)-Mud2 heterodimer, the protein components of the U1 snRNP (i.e. Prp40, Nam8 and Mud1) and the snRNA TMG cap structure.

A new and instructive finding was that four double-alanine mutations in the KH-QUA2 domain that had no apparent effect on *msl5Δ* complementation were synthetically lethal in the *tgs1Δ* background: these were *N163A-V165A* (lethal at 25, 30 and 37°C), *V195A-K196A* (lethal at 25 and 30°C; ++ growth at 37°C), *T265A-R267A* (lethal at 25 and 30°C; + growth at 37°C) and *K252A-R253A* (lethal at 25, 30 and 37°C) (Table 3). These results indicate that the essentiality or dispensability of many constituents of the imputed interface of Msl5 with the branchpoint RNA is context-dependent and can be influenced by the presence or absence of other components of the splicing machinery (in this case, the TMG cap). The simplest interpretation is that proper spliceosome assembly *in vivo* relies on *either* the contacts of Asn163, Val165, Val195, Lys196, Thr265 and Arg267 with the conserved yeast branchpoint RNA sequence *or* the presence of the TMG cap.

Further insights to the mutational synergy of *MSL5-(K252A-R253A)* and *tgs1Δ* came from the finding that the swap-allele *MSL5-(K252R-R253K)* was lethal in *tgs1Δ* cells at 25 and 30°C and sufficient for only + growth at 37°C (Table 3). Moreover, whereas the single-mutant *MSL5-R253K* was fully functional in the *tgs1Δ* background, the *MSL5-K252R tgs1Δ* strain was inviable at 25 and 30°C and displayed feeble growth (+) at 37°C (Table 3). We surmise that the mutational synergy is caused by the replacement of Lys252 with either alanine or arginine. This is notable given that Lys252 in yeast Msl5 is predicted not to interact directly with the branchpoint RNA (35).

Finally, we found that the double-alanine mutations of the zinc-binding residues of the knuckle modules displayed little or no mutational synergy with *tgs1Δ* (Table 3). In sum, these studies reveal new genetic interactions of the essential Msl5 protein with the TMG cap that could not have been appreciated by a synthetic genetic array of *tgs1Δ* against the yeast knockout collection (10).

Mutational synergies with U1 snRNP component Nam8

Bridging interactions between constituents of the U1 snRNP engaged at the 5′ splice site and Msl5 bound to the branchpoint RNA sequence are thought to be critical for spliceosome assembly. Emerging evidence, here and elsewhere (36), suggests that the cross-intron bridging contacts are multiple and functionally buffered, such that elimination of a single pairwise interaction (e.g. between branchpoint binding protein and U1

snRNP, U1 and the 5′ splice site, or branchpoint binding protein and the branchpoint RNA) might not elicit an overt phenotype unless other cross-intron contacts are also impaired. To explore this idea, we surveyed several of our Msl5 mutants for their ability to complement growth in an *msl5Δ nam8Δ* background (Table 6). Nam8 is a stoichiometric protein component of the yeast U1 snRNP that is inessential for vegetative yeast growth. The 523-amino acid Nam8 protein contains three tandem RRM domains (RRMs 1, 2 and 3) flanked by N-terminal leader and C-terminal tail segments. Whereas the leader, tail and RRM1 are dispensable for vegetative Nam8 function in multiple synthetic lethal genetic backgrounds (including *tgs1Δ*), Nam8 activity in synthetic lethal situations is compromised by mutations in the putative RNA binding sites of the RRM2 and RRM3 domains (24). We thought that absence of Nam8 might synergize with a subset of our viable Msl5 mutants, especially ones with mutations in the RNA-binding KH-QUA2 domain and the Prp40-binding motif PPxY^{100}.

Indeed, we found that *nam8Δ* was unconditionally synthetic lethal with *MSL5-(P97A-P98A-Y100A)* (Table 6), suggesting that the otherwise dispensable contacts of the PPxY^{100} motif with U1 component Prp40 are rendered essential for spliceosome assembly in the absence of U1 component Nam8. Also, *nam8Δ* was synthetically lethal with several otherwise benign mutations of the RNA binding site in the KH-QUA2 module, these being *N163A-V165A* and *K252A-R253A* (unconditionally lethal), *V195A-K196A* (lethal at 18 and 25°C; + and +++ growth at 30 and 37°C, respectively), and *T265A-R267A* (lethal at 18, 25 and 30°C; + growth at 37°C) (Table 6). These results show how constituents of the protein–protein and protein–RNA interfaces of Msl5 are either essential or not, depending on the composition of the U1 snRNP. It is notable that the synergies of KH-QUA2 and PPxY mutations in the absence of Nam8 track closely with those seen in the absence of TMG caps (compare Tables 3 and 6), hinting that an altered state of the U1 snRNP in *tgs1Δ* cells (18) might

Table 6. Effect of Msl5 mutations on a *nam8Δ* background

MSL5 allele	*msl5Δ nam8Δ* complementation			
	18°	25°	30°	37°
WT	+++	+++	+++	+++
55–476	+++	+++	+++	–
1–458	+++	+++	+++	+++
1–437	+++	+++	+++	+++
1–425	+	+	+	+++
1–401	–	–	–	+
1–381	–	–	–	+
1–347	–	–	–	+
1–312	–	–	–	–
P97A P98A Y100A	–	–	–	–
N163A V165A	–	–	–	–
V195A K196A	–	–	+	+++
T265A R267A	–	–	–	+
K252A R253A	–	–	–	–
K252R	+	+	++	+++
R253K	+++	+++	+++	+++

account for their reliance on otherwise optional features of Msl5.

Ablation of Nam8 also caused a severe synthetic growth defect with a C-terminal truncation of Msl5 to amino acid 401, resulting in lethality at 18, 25 and 30°C and + growth at 37°C (Table 6). A partial gain-of-function in the *nam8Δ* background was seen when the C-terminus was extended to amino acid 425, with +++ growth at 37°C, but only + growth at ≤30°C (Table 6), while *MSL5-(1–437) nam8Δ* cells displayed +++ growth at all temperatures tested. Finally, it was noteworthy that *MSL5-(55–476) nam8Δ* cells were viable and grew normally at 18, 25 and 30°C, though they failed to grow at 37°C (Table 6). By contrast, the *mud2Δ nam8Δ* combination is unconditionally lethal (16). This is an instance in which elimination of the Mud2-interaction domain of Msl5 did not phenocopy the ablation of Mud2. Because the former maneuver is less severe in *nam8Δ* than the latter, we suggest that Mud2 may have a context-dependent role in splicing that does not require its heterodimerization with the Msl5 N-terminal module.

DISCUSSION

The present study enhances our understanding of the domain organization and essential structural features of Msl5, a critical organizer of spliceosome assembly that recognizes the intron branchpoint and makes manifold interactions with other components of the splicing machinery, especially Mud2 and the U1 snRNP. By characterizing the ability of deletion mutants and missense mutants to complement *msl5Δ* growth, we showed that: (i) ablation of the Mud2-binding (amino acids 35–54) and Prp40-binding (PPxY^{100}) elements of the Msl5 N-terminal domain does not affect yeast growth under laboratory conditions; (ii) the C-terminal proline-rich domain (amino acids 382–476), the distal zinc knuckle (amino acids 299–312) and the intervening linker are also dispensable for Msl5 function in an otherwise wild-type background; (iii) a subset of conserved branchpoint RNA-binding amino acids in the KH-QUA2 domain are essential pairwise (Ile198–Arg190; Leu256–Leu259) or in trios (Leu169–Arg172–Leu176) for Msl5 activity, whereas others are dispensable (Asn163–Val165; Val195–Lys196; Thr265–Arg267) and (iv) the QUA2 domain ^{252}KR^{253} dipeptide, purported to act as a governor of yeast versus human branchpoint sequence recognition (based on effects of QUA2 domain and dipeptide swaps on branchpoint RNA binding *in vitro*; *26*), is also dispensable. Indeed, the structural and *in vivo* mutational data point away from the idea that this dipeptide is decisive for yeast branchpoint recognition *in vivo*. We find that neither of the two zinc-binding CxxCxxxxHxxxxC motifs (amino acids 273–286 and 299–312) flanking the KH-QUA2 domain is important *per se* for *msl5Δ* complementation. Whereas, the distal knuckle can be deleted without overt consequence for growth, extending the deletion to include the proximal knuckle is lethal. These results are consistent with previous inferences from RNA binding studies *in vitro* (25) that the proximal knuckle contributes to intron RNA binding, independent of its ability to coordinate zinc via the knuckle cysteines.

Because Msl5 is required for growth of *S. cerevisiae*, and relatively few mutants of Msl5 had been studied previously *in vivo* (20,37), it has escaped analysis by synthetic array approaches that have proven so fruitful in illuminating the genetic connections among the many inessential yeast splicing factors (8–10,12). Here we took advantage of our collection of viable Msl5 mutants to interrogate synthetic genetic interactions, between the inessential structural elements of the Msl5 polypeptide and between Msl5 and several yeast splicing factors that are optional for vegetative growth. Both approaches highlighted multiple redundancies of the dispensable components of the Msl5 protein, as summarized in Figure 4 and discussed briefly below.

Redundancies within the N-terminal segment of Msl5 are evinced by the lethality of simultaneously ablating the Mud2-binding module and the Prp40-binding PPxY^{100}
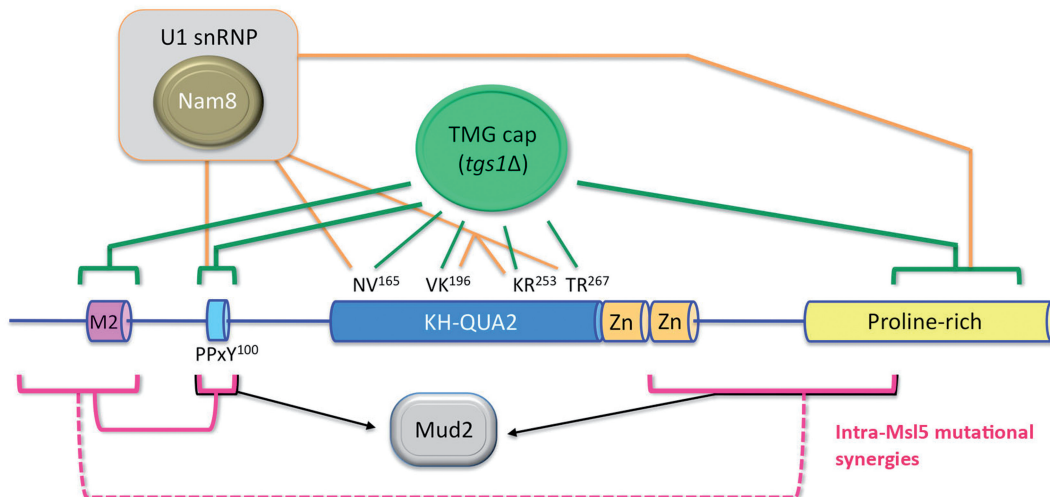

Figure 4. Network of genetic interactions of yeast Msl5 domains.

motif (see the 'intra-Msl5' mutational synergies denoted by red lines in Figure 4). Moreover, mutating the PPxY[100] motif is synthetically lethal with *mud2Δ*. (Mud2 mutational synergies are depicted as black lines in Figure 4). The simplest inference here is that parallel functionally buffered contacts between the Mud2–Msl5 complex at the branchpoint and the U1 snRNP at the 5′ splice site aid in spliceosome assembly *in vivo*, entailing Msl5 interaction with Prp40 while Mud2 interacts with another U1 protein, e.g. the essential U1 component Snu56 (38). Our finding that mutation of the PPxY[100] motif is lethal in a *nam8Δ* background suggests that the Msl5–Prp40 interaction is itself buffered by Nam8, an otherwise inessential U1 snRNP subunit. (Nam8 synthetic interactions are shown as gold lines in Figure 4.) Otherwise benign mutations of the Mud2-binding and Prp40-binding sites of Msl5 are lethal in a *tgs1Δ* background in which U1 snRNA lacks the TMG cap structure. (Tgs1 synthetic interactions are denoted by green lines in Figure 4.) These results, together with prior findings of synthetic lethality between *tgs1Δ*, *nam8Δ*, *mud2Δ* and *mud1Δ* (9,10,16,24), underscore the plentiful overlapping genetic connections between the Msl5-Mud2 heterodimer and the U1 snRNP.

The redundancies of the C-terminal segment of Msl5 that we observe suggest the presence of two distinct functional units. Deletion of the proline-rich domain from amino acids 402–437 elicits synthetic lethality in *tgs1Δ* and *nam8Δ* backgrounds (Figure 4), but does not synergize comparably with *mud2Δ* or with intra-Msl5 deletion of the N-terminal Mud2-binding site. It is conceivable that this segment of the proline-rich domain mediates additional interactions with the U1 snRNP that are buffered by Nam8 and TMG caps. Extending the Msl5 C-terminal deletions inward to amino acid 312 and amino acid 298 imparts synthetic lethality with *mud2Δ* and the Msl5-NΔ54 deletion, respectively (Figure 4). To our knowledge, no specific physical interactions (protein–protein or protein–RNA) have been ascribed to amino acids 402–437 and amino acids 299–401 of Msl5 that could easily explain the genetic synergies. Absent any structural information for the Msl5 C-terminal domain, there is little impetus for speculation.

By contrast, there is an NMR structure available of the KH-QUA2 module of human SF1 bound to the consensus yeast intron branchpoint RNA element (35) that provided a blueprint for the mutagenesis of RNA-binding residues of yeast Msl5. It was the collection of viable alanine-cluster mutants in the KH and QUA2 domains that yielded novel insights to genetic redundancies between Msl5's RNA–protein contacts and its interaction with the U1 snRNP, as inferred from the many synthetic sick and synthetic phenotypes of otherwise benign KH-QUA2 mutations in the *nam8Δ* (and *tgs1Δ*) backgrounds (Figure 4). It is worth noting that the original identification of Msl5 by Abovich and Rosbash (21), entailing a genetic screen for synthetic lethality with an N-terminal deletion mutant of Mud2, yielded an allele with a single Gly230Ser mutation in the third α-helix of the KH domain. The equivalent residue in the human

SF1 structure (Ala218) is buried in the hydrophobic core of the KH fold, quite far from the branchpoint RNA. Rutz and Séraphin (37) reported a cold-sensitive *msl5* phenotype associated with coding changes V195D in the KH domain and E258V in the QUA2 domain. As noted earlier, the conserved Val195 side chain interacts directly with the branchpoint RNA, but is clearly inessential for function, as gauged by the fact that alanine substitution has no effect on *msl5Δ* complementation, even at cold temperature (Table 3). Thus, we surmise that the contribution of the V195D change to the *cs* defect reflects the deleterious impact of ectopic negative charge, rather than the loss of the van der Waals contacts of the valine to the vicinal RNA nucleobases. The Glu258 side chain is located on the face of the QUA2 α-helix opposite that which binds the branchpoint RNA. The upshot is that structurally guided missense mutations of amino acids that directly bind the branchpoint are well-suited to dissect the genetic redundancies of the Msl5-branchpoint interface, whether by screening candidate mutants for synergies (as done here) or by genome-wide approaches (12).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Nilsen,T.W. and Graveley,B.R. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature*, **463**, 457–463.
2. Tazi,J., Bakkour,N. and Stamm,S. (2009) Alternative splicing and disease. *Biochim. Biophys. Acta*, **1792**, 14–26.
3. Fabrizio,P., Dannenberg,J., Dube,P., Kastner,B., Stark,H., Urlaub,H. and Lührmann,R. (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell*, **36**, 593–608.
4. Warkocki,Z., Odenwälder,P., Schmitzová,J., Platzmann,F., Stark,H., Urlaub,H., Ficner,R., Fabrizio,P. and Lührmann,R. (2009) Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nat. Struct. Mol. Biol.*, **16**, 1237–1243.
5. Hoskins,A.A., Friedman,L.J., Gallagher,S.S., Crawford,D.J., Anderson,E.G., Wombacher,R., Ramirez,N., Cornish,V.W., Gelles,J. and Moore,M.J. (2011) Ordered and dynamic assembly of single spliceosomes. *Science*, **331**, 1289–1295.
6. Tang,J., Abovich,N. and Rosbash,M. (1996) Identification and characterization of a yeast gene encoding the U2 small nuclear

ribonucleoprotein particle B" protein. *Mol. Cell. Biol.*, **16**, 2787–2795.

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

8. Tong,A.H., Evangelista,M., Parsons,A.B., Xu,H., Bader,G.D., Pagé,N., Robinson,M., Raghibizadeh,S., Hogue,C.W., Bussey,H. *et al.* (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, **294**, 2364–2368.

9. Wilmes,G.M., Bergkessel,M., Bandyopadhyay,S., Shales,M., Braberg,H., Cagney,G., Collins,S.R., Whitworth,G.B., Kress,T.L., Weissman,J.S. *et al.* (2008) A genetic interaction map of RNA-processing factors reveals links between Sem1/ Dss1-containing complexes and mRNA export and splicing. *Mol. Cell*, **32**, 735–746.

10. Hausmann,S., Zheng,S., Costanzo,M., Brost,R.L., Garcin,D., Boone,C., Shuman,S. and Schwer,B. (2008) Genetic and biochemical analysis of yeast and human cap trimethylguanosine synthase: functional overlap of TMG caps, snRNP components, pre-mRNA splicing factors, and RNA decay pathways. *J. Biol. Chem.*, **283**, 31706–31718.

11. Khanna,M., Van Bakel,H., Tang,X., Calarco,J.A., Babak,T., Guo,G., Emili,A., Greenblatt,J.F., Hughes,T.R., Krogan,N.J. *et al.* (2009) A systematic characterization of Cwc21, the yeast ortholog of the human spliceosomal protein SRm300. *RNA*, **15**, 2174–2185.

12. Costanzo,M., Baryshnikova,A., Bellay,J., Kim,Y., Spear,E.D., Sevier,C.S., Ding,H., Koh,J.L., Toufighi,K., Mostafavi,S. *et al.* (2010) The genetic landscape of a cell. *Science*, **327**, 425–431.

13. Mouaikel,J., Verheggen,C., Bertrand,E., Tazi,J. and Bordonné,R. (2002) Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol. Cell*, **9**, 891–901.

14. Hausmann,S. and Shuman,S. (2005) Specificity and mechanism of RNA cap guanine-N2 methyltransferase (Tgs1). *J. Biol. Chem.*, **280**, 4021–4024.

15. Benarroch,D., Jankowska-Anyszka,M., Stepinski,J., Darzynkiewicz,E. and Shuman,S. (2010) Cap analog substrates reveal three clades of cap guanine-N2 methyltransferases with distinct methyl acceptor specificities. *RNA*, **16**, 211–220.

16. Chang,J., Schwer,B. and Shuman,S. (2010) Mutational analyses of trimethylguanosine synthase (Tgs1) and Mud2: proteins implicated in pre-mRNA splicing. *RNA*, **16**, 1018–1031.

17. Hausmann,S., Ramirez,A., Schneider,S., Schwer,B. and Shuman,S. (2007) Biochemical and genetic analysis of RNA cap guanine-N2 methyltransferases from *Giardia lamblia* and *Schizosaccharomyces pombe*. *Nucleic Acids Res.*, **35**, 1411–1420.

18. Schwer,B., Erdjument-Bromage,H. and Shuman,S. (2011) Composition of yeast snRNPs and snoRNPs in the absence of trimethylguanosine caps reveals nuclear cap binding protein as a gained U1 component implicated in the cold-sensitivity of *tgs1Δ* cells. *Nucleic Acids Res.*, **39**, 6715–6728.

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

20. 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.

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

22. Rain,J.C., Rafi,Z., Legrain,P. and Krämer,A. (1998) Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA*, **4**, 551–565.

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

24. Qiu,Z.R., Schwer,B. and Shuman,S. (2011) Determinants of Nam8-dependent splicing of meiotic pre-mRNAs. *Nucleic Acids Res.*, **39**, 3427–3445.

25. Garrey,S.M., Voelker,R. and Berglund,J.A. (2006) An extended RNA binding site for the yeast branch point-binding protein and the role of its zinc knuckle domains in RNA binding. *J. Biol. Chem.*, **281**, 27443–27453.

26. Garrey,S.M., Cass,D.M., Wandler,A.M., Scanlan,M.S. and Berglund,J.A. (2008) Transposition of two amino acids changes a promiscuous RNA binding protein into a sequence-specific RNA binding protein. *RNA*, **14**, 78–88.

27. Wiesner,S., Stier,G., Sattler,M. and Macias,M.J. (2002) Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. *J. Mol. Biol.*, **324**, 807–822.

28. Rigaut,G., Shevchenko,A., Rutz,B., Wilm,M., Mann,M. and Séraphin,B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.*, **17**, 1030–1032.

29. Berglund,J.A., Abovich,N. and Rosbash,M. (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.*, **12**, 858–867.

30. Selenko,P., Gregorovic,G., Sprangers,R., Stier,G., Rhani,Z., Krämer,A. and Sattler,M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol. Cell*, **11**, 965–976.

31. Kielkopf,C.L., Lücke,S. and Green,M.R. (2004) U2AF homology motifs: protein recognition in the RRM world. *Genes Dev.*, **18**, 1513–1526.

32. Sickmier,E.A., Frato,K.E., Shen,H., Paranawithana,S.R., Green,M.R. and Kielkopf,C.L. (2006) Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol. Cell*, **23**, 49–59.

33. Ash,M.R., Faelber,K., Kosslick,D., Albert,G.I., Roske,Y., Kofler,M., Schuemann,M., Krause,E. and Freund,C. (2010) Conserved β-hairpin recognition by the GYF domains of Smy2 and GIGYF2 in mRNA surveillance and vesicular transport complexes. *Structure*, **18**, 944–954.

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

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

36. Görnemann,J., Barrandon,C., Hujer,K., Rutz,B., Rigaut,G., Kotovic,K.M., Faux,C., Neugebauer,K.M. and Séraphin,B. (2011) Cotranscriptional spliceosome assembly and splicing are independent of the Prp40p WW domain. *RNA*, **17**, 2119–2129.

37. Rutz,B. and Séraphin,B. (2000) A dual role for BBP/ScSF1 in nuclear pre-mRNA retention and splicing. *EMBO J.*, **19**, 1873–1886.

38. Balzer,R.J. and Henry,M.F. (2008) Snu56p is required for Mer1p-activated meiotic splicing. *Mol. Cell. Biol.*, **28**, 2497–2508.