RNA (1998), *4*:1–10. Cambridge University Press. Printed in the USA.

Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits

HALLER IGEL, SANDRA WELLS, RHONDA PERRIMAN, and MANUEL ARES, JR.

Center for the Molecular Biology of RNA, Biology Department, Sinsheimer Laboratories, University of California at Santa Cruz, Santa Cruz, California 95064, USA

ABSTRACT

Human SAP 49, a subunit of the multimeric splicing factor 3b (SF3b), contains two RNA recognition motifs (RRMs) and binds another SF3b subunit called SAP 145, whose yeast homologue is *CUS1*. Here we show that the predicted yeast open reading frame YOR319w (*HSH49*) encodes an essential yeast splicing factor. Using bacterially expressed proteins, we find that yeast HSH49 binds CUS1. Mutations that alter putative RNA-binding residues of either HSH49 RRM are lethal in vivo, but do not prevent binding to CUS1 in vitro, suggesting that the predicted RNA-binding surfaces of HSH49 are not required for interaction with CUS1. In vivo interaction tests show that HSH49 and CUS1 associate primarily through the N-terminal RRM of HSH49. Recombinant HSH49 protein has a general RNA-binding activity that does not require CUS1. The parallels in structure and interaction between two SF3b subunits from yeast implies that the mechanism of SF3b action is highly conserved.

Keywords: RNA-binding proteins; RNA processing; *Saccharomyces cerevisiae*; SAP 49

INTRODUCTION

Removal of introns from nuclear primary transcripts (pre-mRNA) occurs in the spliceosome, a fluid assemblage of snRNPs and extrinsic protein factors that bind to the pre-mRNA. Spliceosome assembly and function involves multiple RNA–RNA, RNA–protein, and protein–protein interactions that are established and dissolved during the course of the reaction (for reviews see Moore et al., 1993; Madhani & Guthrie, 1994; Ares & Weiser, 1995; Kramer, 1996). Early steps of spliceosome assembly do not require ATP and result in the formation of initial complexes (commitment complex or E complex) composed of the pre-mRNA, the U1 snRNP, and numerous proteins that influence subsequent steps in spliceosome assembly (reviewed in Moore et al., 1993).

The first ATP-dependent step in spliceosome assembly is the stable binding of the U2 snRNP to these initial complexes and represents formation of the "prespliceosome" (reviewed in Moore et al., 1993). Biochemical fractionation of mammalian extracts has identified a number of protein factors required for the addition of U2 snRNP during prespliceosome formation, including two multimeric protein complexes, SF3a and SF3b (for review see Kramer, 1996). Independent studies aimed at snRNP purification revealed nine proteins associated with the mammalian U2 snRNP under low-salt conditions (Behrens et al., 1993); included in this group are the seven proteins currently thought to comprise SF3a and SF3b (Brosi et al., 1993). Yet a third strategy involving purification of assembled splicing complexes showed that the same set of proteins are present (and in some cases can be crosslinked to pre-mRNA) in preparations of purified spliceosomes (called SAPs, or spliceosome associated proteins, Bennett et al., 1992; Staknis & Reed, 1994). In addition to revealing proximity to the branchpoint region of pre-mRNA in the assembled spliceosome, Reed and colleagues have provided evidence that the SF3a and SF3b proteins enter the spliceosome at the same time as U2 RNA and remain until the splicing reactions are completed (Bennett et al., 1992; Champion-Arnaud & Reed, 1994; Staknis & Reed, 1994). Taken together, the work from all three approaches indicates that the SF3a and SF3b splicing factors can be considered reversibly salt-

Reprint requests to: M. Ares, Center for the Molecular Biology of RNA, Biology Department, Sinsheimer Laboratories, University of California at Santa Cruz, Santa Cruz, California 95064, USA; e-mail: ares@biology.ucsc.edu.

dissociable U2 snRNP proteins that comprise a large part of the protein environment near the catalytic core of the spliceosome.

Genetic and biochemical studies in yeast have revealed that a set of proteins required for splicing interact functionally with the most conserved parts of U2 snRNA. Based on amino acid sequence similarity, domain swapping, immunological cross reactivity, and conserved protein–protein interactions, the products of the yeast *PRP9*, *PRP11*, and *PRP21* genes are homologues of the three SF3a subunits SAP 61, SAP 62, and SAP 114 (reviewed in Hodges & Beggs, 1994; Kramer, 1996). A similar pattern is emerging for the SF3b set of proteins. A suppressor of a cold-sensitive mutation in stem IIa of U2 RNA called *CUS1* is a yeast splicing factor homologous to the SAP 145 subunit of SF3b (Gozani et al., 1996; Wells et al., 1996). Another SF3b subunit, SAP 49, is known to bind SAP 145 (Champion-Arnaud & Reed, 1994), and the potential to encode a SAP 49 homologue has been identified in the yeast genome (*HSH49* or YOR319w, Wells et al., 1996, Fromont-Racine et al., 1997). In this report, we show that *HSH49* is a yeast splicing factor. Each of two conserved RNA recognition motifs (RRMs, Kenan et al., 1991; Birney et al., 1993; Nagai et al., 1995) are essential for viability and share at least one essential function that must be executed from the same polypeptide chain. A protein–protein interaction formally analogous to that observed between human SAP 145 and SAP 49 (Champion-Arnaud & Reed, 1994) occurs between yeast CUS1 and HSH49 proteins in vitro. We find that this interaction is mediated by the first RRM of HSH49. The highly conserved structures of these proteins and their parallel interactions indicate that there are strong similarities in the mechanism of SF3b action during splicing in yeast and mammals.

RESULTS

HSH49 is member of a conserved family of SAP 49 homologues

We previously identified yeast and worm (*Caenorhabditis elegans*) sequences with potential to encode proteins homologous to SAP 49 (Wells et al., 1996), by using the human SAP 49 sequence (Champion-Arnaud & Reed, 1994) to search the public database. Recently, studies have revealed *Drosophila* (emb | X97197; gb | AA263487), *Schizosaccharomyces pombe* (VanHoy & Wise, 1996) and mouse (Ruiz-Lozano et al., 1997) homologues as well. Alignment of representatives of these family members (Fig. 1) indicates that the predicted yeast HSH49 protein possesses many of the structural features shared by the metazoan SAP 49 homologues. Each protein contains two segments homologous to the RRM domain (also known as the RNP, RBD, or CS-RNP80 domain) found in a large family of proteins demonstrated (or presumed) to bind RNA (Kenan et al., 1991; Birney et al., 1993; Nagai et al., 1995). The N-terminal RRM (RRM1) is more conserved than the C-terminal RRM (RRM2) when the yeast sequence is included in the comparison (Fig. 1). A C-terminal proline-rich domain of varying length is found in the metazoan proteins, but is absent in yeast (Wells et al., 1996). The length of the C-terminal proline-rich sequence accounts for most of the difference in size of the different family members. The proline-rich domain has been shown to be dispensable for the protein–protein interaction between human SAP 145 and SAP 49 (Champion-Arnaud & Reed, 1994).

The metazoan SAP 49 proteins show greater similarity to each other than to the yeast HSH49 protein. The similarities extend beyond the presence of the

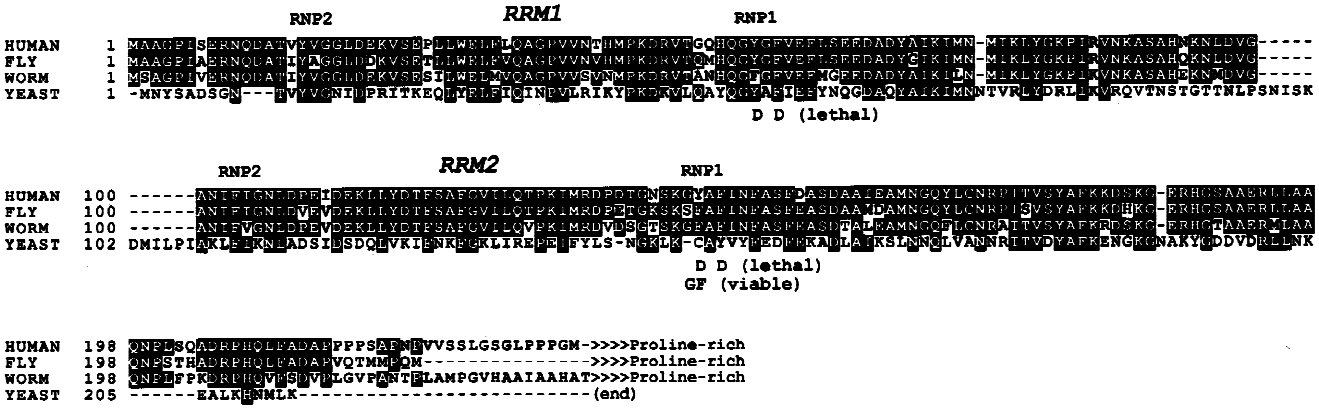

FIGURE 1. Alignment of SAP 49 homologues from different organisms. Positions of the conserved RNP2 and RNP1 sequences of each RRM are indicated. Mutations in the RNP1 sequence of each RRM and their phenotypes are listed below the yeast sequence. A variable-length proline-rich domain extends C-terminal to the metazoan sequences not shown.

proline-rich domain and span the entire protein. For example, of the 215 amino acids N-terminal to the proline-rich domain, 167 (78%) are identical and 210 (98%) are similar when the human SAP 49 sequence is compared to that from the two invertebrates (Fig. 1). Comparison of the yeast sequence with the metazoan sequences reveals that HSH49 is 35% identical and 55% similar to the metazoan SAP 49 family members, exclusive of the proline-rich domain. Searches performed on the complete yeast genome show that the HSH49 open reading frame represents by far the best match to the metazoan SAP 49 sequences. We conclude that HSH49 is the yeast member of the SAP 49 family.

HSH49 is an essential gene

To study HSH49 as a model for the role of SAP 49-like proteins in splicing, we prepared primers designed to amplify a segment of the yeast genome spanning the *HSH49* coding region and including flanking sequences presumably required for transcription. The gene was amplified, cloned in a plasmid vector, and completely sequenced to confirm that amplification was accurate. To determine whether the *HSH49* gene is essential, a disrupted *HSH49* allele was generated by inserting the *HIS3* gene in place of sequences in the middle of *HSH49* (*hsh49::HIS3*). In addition, the wild-type *HSH49* gene was cloned into centromere plasmids marked with either *URA3* or *TRP1*.

We cotransformed a *ura3*$^-$, *his3*$^-$ haploid yeast strain with linear DNA representing the *hsh49::HIS3* disruption and a centromere plasmid carrying *URA3* and the wild-type *HSH49* gene. The Ura$^+$, His$^+$ transformants were tested for growth on 5-fluoroorotic acid (5-FOA) and numerous transformants unable to give rise to Ura$^-$ cells were obtained (data not shown), indicating that *HSH49* is essential. To confirm this, a wild-type copy of *HSH49* on a *TRP1* plasmid was introduced into these strains. Upon selection with 5-FOA, Ura$^-$, His$^+$ colonies could now be obtained, and these invariably carried the *TRP1-HSH49* plasmid (data not shown), demonstrating that yeast cannot survive without a functional *HSH49* gene. Southern blots probed with *HSH49* sequences confirmed the presence of the *hsh49::HIS3* fragment integrated at the chromosomal locus of *HSH49* (data not shown). We conclude that *HSH49* is an essential gene.

Repression of HSH49 expression inhibits splicing

To determine whether expression of *HSH49* is required for splicing in vivo, we constructed a glucose-repressible derivative of *HSH49* by fusing the *GAL1* promoter to the *HSH49* coding region (*GAL:HSH49*). A high copy plasmid carrying the fusion is able to complement the *hsh49::HIS3* disruption when the growth medium lacks glucose and contains galactose (not shown), however, such strains cannot grow well on medium containing glucose (Fig. 2A). To determine whether this growth inhibition is accompanied by a splicing defect, we shifted strains carrying a wild-type *HSH49* gene or the *GAL:HSH49* construct from galactose to glucose, isolated RNA, and measured levels of unspliced pre-U3 RNA by primer extension (Fig. 2B). As cells carrying the *GAL1:HSH49* gene cease growth, unspliced U3 RNA accumulates, indicating that efficient splicing in vivo requires *HSH49* expression. This experiment provides functional evidence that HSH49 protein plays a role in splicing in vivo. We conclude that HSH49 is an essential yeast splicing factor.

Both RRMs are essential for HSH49 function

To determine whether either or both RRMs in HSH49 are essential, we designed mutations in the putative RNA-binding surface of each RRM, and tested their ability to complement the *hsh49::HIS3* disruption by plasmid shuffling. We chose to alter two residues at positions typically occupied by aromatic side chains in the RNP-1 (beta 3 strand) of each RRM (Fig. 3A). Based on structural analysis of other members of the RRM family of RNA-binding domains, in particular U1A, the targeted residues (in RRM1: Y52 and F54; in RRM2: C150 and Y152, positionally equivalent to Q54 and F56 in U1A) are expected to project into solvent from the RNA-binding face of the four-stranded beta sheet, where they may stack with bases in RNA (Kenan et al., 1991; Oubridge et al., 1994; Nagai et al., 1995; see Fig. 3A). Similar double aspartate substitutions have been shown to disrupt RRM function (Caceres & Krainer, 1993). We constructed the desired mutations in a centromere plasmid carrying the *TRP1* marker and tested them by plasmid shuffling (Guthrie & Fink, 1991; see Fig. 3B). The results are tabulated in

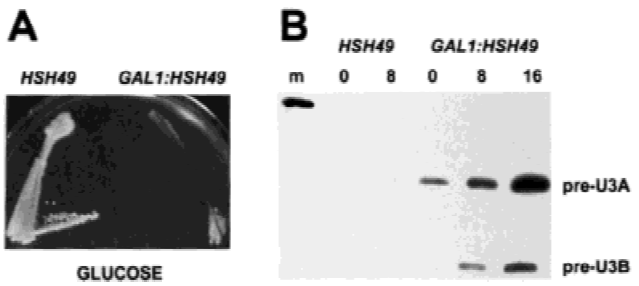

FIGURE 2. Repression of HSH49 expression inhibits growth and splicing. **A:** Growth inhibition. Yeast carrying wild-type (left) or a *GAL:HSH49* gene were grown in galactose medium, streaked on glucose medium, and incubated at 30 °C for 3 days. **B:** Inhibition of U3 splicing. Yeast carrying wild-type (left) or a *GAL:HSH49* gene were grown in galactose medium, and shifted to glucose medium at time zero. RNA was extracted at indicated times (h). Splicing of pre-U3 RNA was analyzed using reverse transcriptase and a primer complementary to the second exon of U3A and U3B snoRNA.

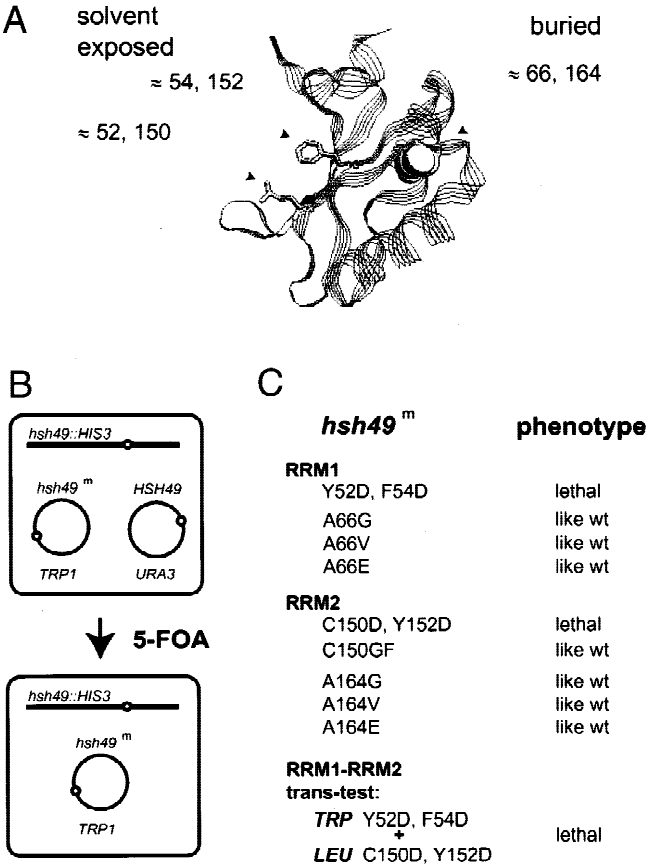

FIGURE 3. Both RRM domains are necessary for an essential function of HSH49. **A:** Design of mutations in the RRMs of HSH49. Using an alignment and representation of U1A, residues of HSH49 were targeted for mutagenesis as described in the text. **B:** Plasmid shuffling strategy. A strain carrying a chromosomal disruption of *HSH49* (*hsh49::HIS3*) and a wild-type copy of the gene on a *URA3* plasmid was transformed with the indicated mutants (*hsh49*^{m}) on a *TRP1* plasmid. Growth and phenotype of cells on 5-FOA medium reveals the phenotype of the tested mutant. **C:** Mutant phenotypes. Growth of strains carrying the indicated HSH49 amino acid substitutions is indicated in the table. Lethal, no growth on 5-FOA; like wt, growth of strains carrying the indicated mutation was like wild type on 5-FOA, and on rich medium at different temperatures. The *trans*-test employed the two indicated mutant genes carried on separate plasmids.

Figure 3C. The RRM1 substitution Y52D, F54D is lethal, indicating that RRM1 is essential for function. RRM2 is unusual but not unique in lacking the first conserved glycine of the RNP-1 octamer and in having cysteine in place of the first aromatic residue (see Birney et al., 1993). Replacement of C150 with two amino acids (glycine and phenylalanine, GF) creates a metazoan "consensus" sequence in RNP1 of RRM2 (worm has GF, human GY, and fly SF at this position, see Fig. 1). This substitution causes no obvious growth defect, indicating that the unusual cysteine residue in the second RRM carries no essential yeast-specific function. To test for the functional requirement of RRM2, we replaced C150 and Y152 with aspartates, and this substitution is lethal (Fig. 3C), indicating that RRM2 is also essential. In the *C. elegans* SAP 49 homologue, RRM2 is primarily responsible for an RNA-binding activity assayed in vitro (Tanaka et al., 1997).

To determine whether the two RRMs might have distinct nonoverlapping functions that could be *trans*-complemented, we created a strain in which the RRM1 mutation was carried on one plasmid and the RRM2 mutation was carried on a separate plasmid. The presence of both plasmids in the same cell did not complement the *hsh49::HIS3* disruption. This suggests that the two RRMs of HSH49 do not have divisible functions that can be executed from separate polypeptide chains. We conclude that each RRM of HSH49 is required for at least one essential function in yeast, and that at least one of these functions necessitates the presence of both RRMs on the same polypeptide chain.

In an effort to obtain temperature-sensitive alleles of HSH49, we altered alanine residues in the conserved hydrophobic core of the RRM fold (A66 for HSH49 RRM1 and A164 for HSH49 RRM2, positionally equivalent to U1A residue A68; Oubridge et al., 1994). A substitution of this alanine results in temperature sensitivity of another essential yeast RRM protein (Henry et al., 1996). In the case of both of the HSH49 RRMs, changing these putatively core-positioned alanine residues to glycine, valine, or even glutamate produced no detectable heat- or cold-sensitive growth defect.

HSH49 binds CUS1, the yeast homologue of SAP 145

SAP 49 binds to SAP 145 (Champion-Arnaud & Reed, 1994). The yeast homologue of SAP 145 is the splicing factor CUS1 (Gozani et al., 1996; Wells et al., 1996), identified as a suppressor of U2 RNA mutations (Wells et al., 1996; Yan & Ares, 1996). To determine whether the protein–protein interaction observed between human SAP 49 and SAP 145 is conserved in yeast, we prepared CUS1 and HSH49 protein from bacteria. Expression of untagged CUS1, CUS1 tagged at the C-terminus with the HA epitope and six histidines (Wells et al., 1996), and untagged HSH49 is relatively efficient and produces soluble protein in *Escherichia coli*. We bound tagged CUS1 to nickel-chelate matrix and determined the ability of untagged HSH49 protein to bind. Untagged HSH49 protein produced in bacteria binds to the matrix only if tagged CUS1 has been prebound (Fig. 4A). We have also been able to bind untagged CUS1 to nickel-chelate matrix provided the matrix has been preloaded with His-tagged HSH49 (data not shown). The interaction between bacterially produced CUS1 and HSH49 is resistant to 400 mM NaCl (data not shown). We conclude that HSH49 binds to CUS1.

A heterogeneous substance that stains orange-brown with silver can be observed near the tops of the gel lanes containing tagged CUS1, but not in lanes that

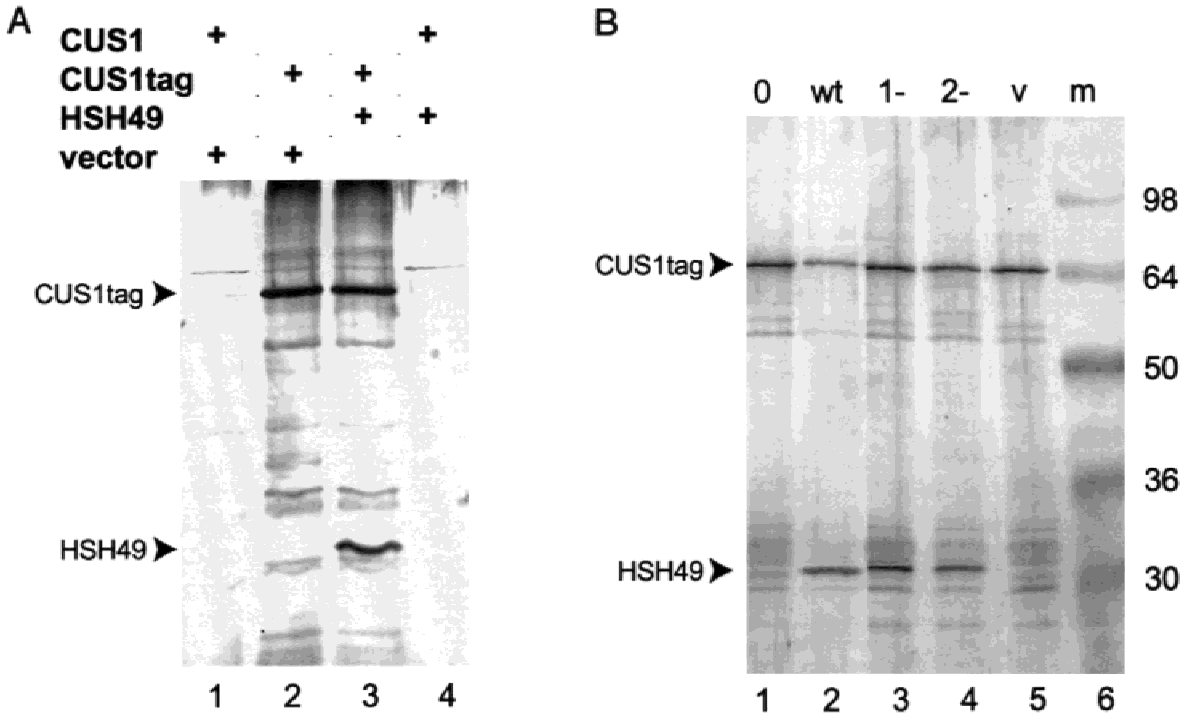

FIGURE 4. Bacterially expressed HSH49 binds to CUS1 in vitro. **A:** Binding of HSH49 is dependent on HSH49 expression and the tag on CUS1. Proteins retained on Ni-chelate matrix were eluted with imidazole, run on a gel and silver stained. **B:** RNP1 RRM mutations do not prevent interaction between HSH49 and CUS1. Proteins retained on Ni-chelate matrix in the presence of RNaseA were eluted with imidazole, run on a gel, and silver stained. All samples contained CUS1-tag extract plus the indicated extract. Lane 1, no second extract; lane 2, wild-type HSH49; lane 3, RRM1 mutant Y52D, F54D; lane 4, RRM2 mutant C150D, Y152D; lane 5, vector; lane 6, markers.

contain His-tagged HSH49 or no tagged bacterial protein. This substance does not stain with Coomassie and is identified as RNA based on its sensitivity to ribonuclease (data not shown, compare lanes 2 and 3 of Fig. 4A with all lanes in Fig. 4B), and additional experiments indicate that CUS1 binds RNA (M. Haynes Pauling & M. Ares, unpubl. obs.). We were concerned that RNA might mediate binding between CUS1 and HSH49 indirectly; however, their association is resistant to ribonuclease (Fig. 4B). Although this eliminates the possibility of indirect association through an RNA linker to which both proteins are bound, allosteric or cooperative effects of binding of RNA fragments to either protein on their ability to associate with each other remain possible.

To test the possibility that the lethal RRM mutations of HSH49 might block the protein–protein interaction between HSH49 and CUS1, we produced recombinant HSH49 carrying the mutations and tested them for binding to CUS1 in vitro (Fig. 4B). Both proteins readily bound to tagged CUS1 (lanes 3 and 4), suggesting that the lethal defects caused by the RRM mutations are not caused by a direct block in the association between CUS1 and HSH49. This also argues against a strong contribution by the putative RNA-binding surfaces of the HSH49 RRMs to the protein–protein interaction surface between HSH49 and CUS1.

HSH49-CUS1 interaction is mediated primarily through RRM1 of HSH49

To simplify the mapping of protein sequences required for the interaction between HSH49 and CUS1, we tested the ability of subdomains of HSH49 to support a two-hybrid interaction with CUS1 (Fig. 5). Results from exhaustive two-hybrid screening of the yeast genome show that the protein–protein interaction between HSH49 (YOR319w) and CUS1 can be detected

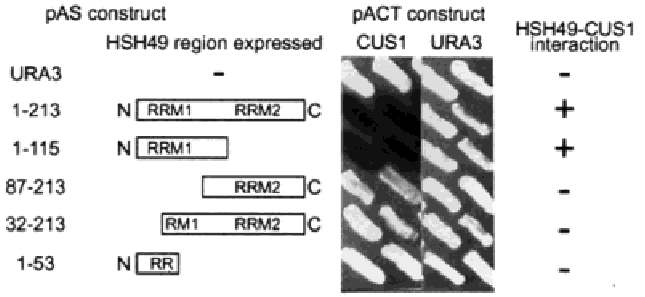

FIGURE 5. HSH49 RRM1 mediates CUS1 binding. HSH49 amino acids listed at left were fused to the C-terminus of the *GAL4* DNA binding domain and tested for interaction with CUS1 fused to the C-terminus of the *GAL4* activation domain in the two-hybrid system. Constructs fused to *URA3* were used as controls. Cells were patched on agar plates, grown, and overlaid with X-gal. Interaction is indicated by dark color of the patched cells.

by this technique (Fromont-Racine et al., 1997). A construct containing only the RRM1 of HSH49 interacts with CUS1 nearly as well as the complete protein (Fig. 5). Splitting RRM1 into two pieces destroyed the interaction with CUS1. An RRM2 fusion is unable to mediate the interaction. A fusion containing RRM1 with the lethal double aspartic acid substitution supported the interaction (data not shown), in agreement with the biochemical results above (Fig. 4B). We conclude that the protein–protein interaction between HSH49 and CUS1 occurs primarily through RRM1 of HSH49, but we cannot yet assign this interaction to a smaller part of RRM1. The negative results from the smaller constructs do not allow us to determine why the fusions failed to interact, but it seems likely that destroying the RRM fold prevents formation of the CUS1 interaction surface.

A GST-HSH49 fusion protein binds RNA

Because of the tight association of CUS1 with HSH49 (Fig. 4), and the unusual nature of the second RRM (Fig. 1), we sought to demonstrate whether HSH49 possesses an RNA-binding activity in the absence of CUS1 protein. We fused the coding region of HSH49 to GST, expressed the protein in *E. coli*, and purified the expressed protein using glutathione agarose affinity chromatography. A 309-nt U2 transcript consisting of the conserved essential 5′ and 3′ structural elements of yeast U2 snRNA binds to GST-HSH49 protein with an apparent K_d of less than 400 nM under the binding conditions used (Fig. 6). Under the same conditions, an equivalently sized RNA derived from transcription of a polylinker does not bind. Although we have not yet determined whether one or both RRMs are necessary for this binding activity, we know that other RNAs, such as pre-actin mRNA, are also able to bind (data not shown). Thus, this experiment indicates only that HSH49 has a general RNA-binding activity in the absence of CUS1, but does not convincingly identify the natural target of either RRM.

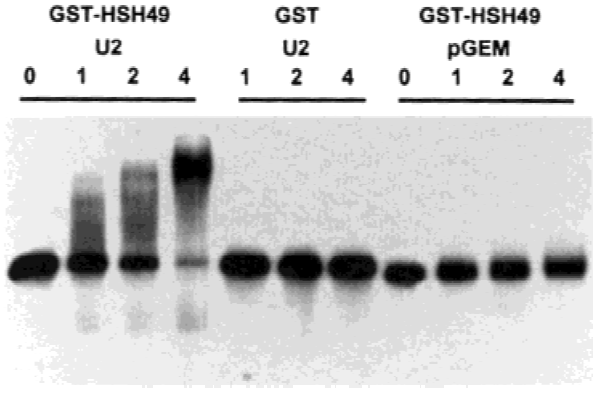

FIGURE 6. GST-HSH49 binds RNA. RNA (20 fmol) and protein (indicated at top in pmol) were mixed, incubated in 10 μL (protein concentrations are 100–400 nM), and run on a native polyacrylamide gel. Left: GST-HSH49 mixed with a U2 RNA. Center: GST alone mixed with a U2 RNA. Right: GST-HSH49 mixed with a polylinker RNA. Complex formation is indicated by retarded migration of the probe.

DISCUSSION

We have presented evidence that HSH49, the yeast homologue of human SAP 49/SF3b50, functions in splicing (Fig. 2). Although strikingly conserved among metazoans, the yeast protein has diverged, more so in the second RRM (Fig. 1). Mutations tested in the yeast protein demonstrate that key residues at conserved positions likely mediating RNA binding are essential for function (Fig. 3). A distinct region contained in residues spanning RRM1 is necessary for a protein–protein interaction between HSH49 and CUS1 (Fig. 5), the yeast homologue of SAP 145/SF3b145 (Gozani et al., 1996). This interaction is not disrupted by lethal mutations in putative RNA-binding regions of HSH49 and is resistant to RNaseA, indicating that the interaction between the two proteins is not mediated by RNA (Fig. 4). The CUS1-HSH49 interaction is not required to support a general RNA-binding activity of HSH49 detected using a recombinant GST-HSH49 fusion protein (Fig. 6). The conservation of structure and interactions between two essential subunits of yeast SF3b indicate that this multimeric splicing factor and spliceosomal component plays conserved roles in splicing.

Conservation of subunit structure in SF3b

SF3b is a splicing factor required for U2 snRNP binding to the prespliceosome at an early step in splicing complex formation (Kramer, 1996). Four major polypeptides are represented in the most purified fractions: SAP 49/SF3b50, SAP 145/SF3b145, SAP 155/SF3b155, and SAP 130/SF3b130, however, it is not certain that this list of SF3b subunits is complete. Along with SF3a, the four identified SF3b subunits comprise seven of the nine additional U2 snRNP proteins that are found in the 17S U2 snRNP (Behrens et al., 1993; Brosi et al., 1993; Staknis & Reed, 1994). Yeast SF3b subunit homologues are highly similar to the mammalian subunits. For example, CUS1 is 43% identical and 65% similar to SAP 145 in a central 200-amino acid segment of each protein (Gozani et al., 1996; Wells et al., 1996). HSH49 is also essential and conserved, showing 35% identity and 55% similarity to the metazoan homologues, exclusive of a nonconserved proline-rich domain (Fig. 1). In addition to sequence similarity, a protein–protein interaction between HSH49 and CUS1 is observed (Figs. 4, 5; Fromont-Racine et al., 1997), analogous to that observed between SAP 49 and SAP 145 (Champion-Arnaud & Reed, 1994). Interaction between the human proteins has only been restricted to the half of SAP 49 containing both RRMs (Champion-Arnaud

& Reed, 1994); interaction between the yeast proteins is mediated primarily through the first HSH49 RRM (Fig. 5). Like the mammalian SF3b proteins (Brosi et al., 1993), yeast CUS1 protein is required for stable U2 snRNP addition to the spliceosome (Wells et al., 1996). Taken together, these results suggest that yeast SF3b protein complex is essential for growth and will work together with U2 RNA in a fashion conserved throughout eukaryotes.

Indivisible functions for the RRMs of HSH49

Aromatic amino acid residues on the solvent-exposed surface of the beta 3 strand of the RRM fold within the U1A protein–U1 stem-loop A RNA co-crystal structure are involved in RNA binding (Oubridge et al., 1994). Mutation of the corresponding residues of either RRM of HSH49 to aspartate are lethal (Fig. 3). These mutations do not prevent the protein–protein interaction with CUS1 as assayed in bacterial extracts (Fig. 4) or in the two-hybrid system (data not shown). The second RRM of the yeast protein has an unusual but not unique replacement of a G(F/Y) dipeptide with a single cysteine residue. The devolutionary replacement of this C in RRM2 with GF did not disrupt function in yeast (Fig. 3). Consistent with the findings of Tanaka et al. (1997), for the *C. elegans* protein, we find that purified recombinant yeast HSH49 has a general RNA-binding activity. Together with the efficient crosslinking of RNA to human SAP 49 (Champion-Arnaud & Reed, 1994; Gozani et al., 1996), these observations strongly suggest that each RRM of HSH49 is involved in a conserved, essential RNA-binding event. Furthermore, these essential binding events cannot be separated to distinct polypeptide chains, as indicated by the failure of a *trans*-complementation test in which cells producing two mutant HSH49 proteins, each with a different RRM inactivated, were unable to survive (Fig. 3).

The precise RNA targets of the HSH49 RRMs are not known. RNAs that bind to RRM2 of the *C. elegans* SAP 49 homologue have been identified by selection (Tanaka et al., 1997), but these resemble neither a conserved U2 sequence, nor the "anchoring" sequence of pre-mRNA, shown to crosslink to human SAP 49 in purified splicing complexes (Gozani et al., 1996). Both RRMs could be required to bind a single target, or the protein could bind two different RNAs. Human SAP 49 crosslinks efficiently to pre-mRNA upstream of the branchpoint (Gozani et al., 1996) and, in yeast RRM1 of HSH49, is in close association with CUS1 (Fig. 5), which suppresses U2 mutations (Wells et al., 1996). If HSH49 binds two RNAs in the spliceosome, it might bind U2 RNA through RRM1, and the pre-mRNA through RRM2 when the U2 snRNP is bound to the branchpoint. Definition of the RNA ligands of the HSH49 RRMs will be necessary to understand how HSH49 functions during splicing.

Interaction between HSH49 RRM1 and CUS1

RRM1 of HSH49 is able to bind CUS1 in the two-hybrid system (Fig. 5), and a lethal alteration of putative RNA-binding residues in this RRM (Y52D, F54D) does not block the interaction with CUS1 (Fig. 4B and data not shown). This suggests that HSH49 residues required for interaction with CUS1 are not located on the putative RNA-binding face of RRM1 and that the function defined by Y52D, F54D is likely not CUS1 binding. Recent studies indicate that evolutionarily conserved CUS1 sequences are required for HSH49 binding (M. Haynes Pauling & M. Ares, unpubl.), suggesting that conserved sequence elements of HSH49 may also be involved. In considering the location of the subunit interface, hydrophobic patches composed of conserved amino acids can be found on the outer faces of both alpha helices of HSH49 RRM1. Loss of the interaction observed in two overlapping truncations of RRM1 (Fig. 5) suggests that no single small region of HSH49 is responsible, and that the integrity of the RRM fold plays a role in presenting the surface that interacts with CUS1. The ability of Y52D, F54D mutant HSH49 to bind CUS1 also suggests that RNA need not be bound to RRM1 for the interaction with CUS1 to take place.

Expression of wild-type HSH49 from a GAL promoter on a high copy plasmid results in a slight growth defect (not shown) and the accumulation of unspliced pre-U3 (Fig. 2B). We have also noted a dominant negative growth phenotype associated with increased expression of RRM1 alone (data not shown). These observations suggest that the amount of HSH49 produced in the cell must be regulated with respect to other components, for example, CUS1 and any other yeast SF3b subunits, and possibly the putative RNA targets of HSH49. If HSH49 binds multiple ligands (at least CUS1 protein plus at least one RNA) simultaneously, then too much HSH49 could increase the concentration of many different HSH49 binary complexes at the expense of a critical multimeric complex containing HSH49, for example, SF3b or the U2 snRNP.

MATERIALS AND METHODS

Protein sequence analysis

Sequences homologous to SAP 49 were identified through BLAST (Altschul et al., 1990) searches at the National Center for Biotechnology Information at the National Library of Medicine (http://www.ncbi.nlm.nih.gov). SAP 49 family members were aligned using CLUSTAL W (Higgins et al., 1996) through the Baylor College of Medicine Search Launcher (http://dot.imgen.bcm.tmc.edu:8088/search-launcher/launcher.html), and shaded using BOXSHADE (http://ulrec3.unil.ch/software/BOX_form.html). Identification of putative solvent-exposed and buried residues of the HSH49 RRMs was done by aligning each HSH49 RRM with the

sequence of human U1A protein and noting the positions of the corresponding U1A residues in the crystal structure of the U1A protein using RASMOL (Sayle & Milner-White, 1995).

Yeast strains and growth

The yeast genomic sequences surrounding a region predicted to encode a protein homologous to human SAP 49 (YOR319w Wells et al., 1996; Fromont-Racine et al., 1997) was amplified from wild-type yeast DNA using Vent DNA polymerase (New England Biolabs) and the following primers: "montana," 5′-GAAGATCTCATTAACGCCGTTTTCC-3′; and "rice," 5′-GAAGATCTGCCAAGATTGTGTCCG-3′, which are designed to place *Bgl* II sites at the ends of the product. The product was cleaved with *Bgl* II and cloned into the *Bam*H I site of pGEM7zf(+). One clone was entirely sequenced and matches the sequence of the yeast genome. This fragment was subcloned into centromere plasmids carrying URA3 or TRP1. An HA+6HIS-tagged (Peterson et al., 1994) *GAL*-controlled *HSH49* gene was made using the primers "deion," 5′-AATATGCGGCCGCTAAGATTAACATTCA TGAAC-3′, and "ronnie," 5′-AATATGCGGCCGCTTTCAA CATGTTGTGCTT-3′, which are designed to fuse the C-terminus of HSH49 to the HA tag (Kolodziej & Young, 1991) plus six histidines (Crowe et al., 1994) under *GAL* control in the pTAG vector (Wells et al., 1996). The disrupted *HSH49* allele was generated by inserting the *HIS3 Bam*H I fragment in place of the *Nhe* I–*Bsm* I fragment of *HSH49* by ligation of fragments made blunt ended by the "Klenow" DNA polymerase. This removes 196 base pairs of the *HSH49* coding region, to create *hsh49::HIS3*. Mutant derivatives were made by oligonucleotide-directed mutagenesis (Kunkel et al., 1991).

To determine whether *HSH49* is essential, we transformed yeast HI227 (*MAT**a***, *ura3-52*, *leu2-3,112*, *trp1*, *his3-d200*, *lys2*Δ, *pep4-3*, *prb1*, *prc1*) with a mixture of both a linear DNA fragment spanning the *hsh49::HIS3* plasmid, and the *URA3* centromere plasmid carrying wild-type *HSH49*, using standard procedures (Guthrie & Fink, 1991). This results in His^+ Ura^+ transformants, primarily of two classes. In one class, the linear fragment is integrated by homologous recombination into the centromere plasmid, disrupting the plasmid-borne *HSH49* gene. In the second class, the linear fragment is integrated at and disrupts the chromosomal locus of *HSH49*, and the plasmid remains intact. The two classes and the essential nature of the gene are easily distinguished by examining cells that spontaneously lose the centromere plasmid from a number of independent transformants on 5-FOA medium, which selects for cells that have lost the *URA3* gene (Guthrie & Fink, 1991). The first class of transformant readily generates cells that lose the plasmid and become Ura^-; these become His^- as well because the *hsh49::HIS3* fragment is integrated into the plasmid. The behavior of the second class of transformant reveals whether the gene is essential or not. If the gene is not essential, Ura^- cells are easily recovered, and they remain His^+. An essential gene such as *HSH49* gives rise to transformants that do not grow on 5-FOA. Transformants of this class could grow on 5-FOA if provided with a second *HSH49* gene on a *TRP1* plasmid. Southern blot analysis indicates that the *hsh49::HIS3* fragment has replaced the wild-type gene. This strain was used to test the function of different *HSH49* mutants on the *TRP1* plasmid pRS314 (Fig. 3). A similar strain was constructed to test the effects of repression of HSH49 synthesis using the *GAL1:HSH49* gene carried on the pTAG vector (high copy, *URA3*), and the *hsh49::URA3* disruption. In this case, the plasmid lacks homology with the ends of the linear fragment, so that most transformants are disrupted in the chromosome. Because the cells require galactose to maintain expression of HSH49, this transformation was plated on galactose medium. Southern blots confirmed the correct integration of the disruption.

Glucose repression, RNA isolation, and primer extension with a U3 oligonucleotide was done as described previously (Wells et al., 1996).

Protein–protein interactions

Two-hybrid plasmids were constructed by subcloning *HSH49* restriction fragments or amplifying desired segment, and cloning them into pAS2-1 (Durfee et al., 1993; Harper et al., 1993). The entire *CUS1* coding region was cloned into pACT2 by fusing an *Nde* I site upstream of CUS1 into the *Nde* I site of pACT2 (Durfee et al., 1993; Harper et al., 1993). As controls, the *Nco* I–*Sal* fragment of the URA3 protein was cloned into each two-hybrid vector. DNAs from pairs of plasmids were mixed and cotransformed into Y187 (Durfee et al., 1993; Harper et al., 1993) and selected on synthetic complete medium lacking leucine and tryptophan (SCD −leu −trp). Multiple transformants from each pairwise combination of plasmids were patched in a grid on the same medium, allowed to grow, and overlaid with molten 0.5% agarose, 0.5 M $NaPO_4$, pH 7.0, 0.1% SDS, 2% dimethylformamide, 0.2% w/v X-gal, and incubated at 37 °C.

Recombinant untagged HSH49, CUS1, and CUS1-tag proteins were expressed in *E. coli* BL21 using pET vector constructs. A DNA fragment encoding CUS1 with an HA-tag plus six histidines at its C-terminus was subcloned from pGCT (Wells et al., 1996) into pET24b (Novagen). The 3′ end of this gene in this construct was replaced with that from the wild-type gene to create an untagged CUS1-expressing plasmid. An untagged *HSH49* gene and the mutant derivatives were cloned into pET11 (Novagen) using a *Bsp*H I site at the start codon in a clone derived by amplification using the deion primer. To produce protein, 2-mL overnights of expressing clones were used to seed 50-mL cultures in LB to $A_{600} = 0.05$ and cultures were incubated at 37 °C until $A_{600} = 0.6$. IPTG was added to 10 mM and cultures were incubated an additional 3 h, collected, and resuspended in 5 mL 50 mM HEPES-K^+, pH 7.5. Lysozyme was added to 100 μg/mL and Triton-X100 was added to 0.1%. After incubation at 30 °C for 20 min, the cells were chilled to 0 °C and lysed by sonication. The lysate was clarified by centrifugation at 10,000 × *g* for 10 min. Aliquots were frozen and stored at −70 °C. Protein isolated from *E. coli* BL21 carrying pET24b was produced as a control.

To test protein binding in vitro, Nickel-NTA resin (Qiagen) was equilibrated in 100 mM KCl, 50 mM HEPES-K^+, pH 7.6, 30 mM imidazole and suspended as a ~50% slurry. Cleared sonicates, prepared as described above, were thawed, made to 100 mM KCl, 30 mM imidazole, and 450-μL aliquots were mixed together in pairwise combinations. Mixtures were bound in batch to 30-μL aliquots of equilibrated 50% slurry Ni-NTA resin at 4 °C for 2 h with constant rotation. Resin was washed three times with 1 mL of 100 mM KCl, 50 mM HEPES-

K^+, pH 7.6, 30 mM imidazole. After removal of the last wash, bound proteins were eluted by incubation for 10 min on ice in 100 mM KCl, 50 mM HEPES-K^+, pH 7.6, 250 mM imidazole. Seven to ten microliters of eluate was run on a gel and silver stained. To demonstrate RNase resistance, extracts were mixed with 10 μg/mL RNaseA during the 2-h binding. Binding of HSH49 to CUS1 is retained after washing in 400 mM KCl, 50 mM HEPES-K^+, pH 7.6, 30 mM imidazole, indicating salt resistance of the interaction.

RNA binding

RNA binding by HSH49 was assessed by incubating recombinant HSH49 fusion protein with radiolabeled RNA. Recombinant HSH49 was produced as a GST fusion. The HSH49 coding region was amplified using the "deion" and "ronnie" oligonucleotides and the amplified product was cut with *Bsp*H I and *Not* I, the ends were filled in, and the fragment was ligated into the *Sma* site of pGEX-2T (Pharmacia Biotech). The plasmid was introduced into *E. coli* BL21 and expression of the GST-HSH49 fusion protein was induced with IPTG as above. Sonicated extracts of expressing clones of *E. coli* were bound to glutathione agarose (Sigma), washed, eluted with reduced glutathione, and desalted on a Microcon filter. GST produced in parallel was used as a control. Purity was estimated to be greater than 90% by gel electrophoresis. Protein concentrations were determined by the method of Bradford (1976). RNA probes were synthesized by in vitro transcription using T7 RNA polymerase (Milligan & Uhlenbeck, 1989). Protein (0–4 pmol) was incubated with labeled RNA (20 fmol) for 10 min at 25 °C in 10 μL of 20 mM HEPES, pH 7.9, 125 mM KCl, 1 mM dithiothreitol, 0.1% Triton X-100, 5% glycerol, 1 mM EDTA, and 50 ng tRNA (approximately 100-fold molar excess over probe). Labeled probe plus the tRNA was mixed, heated to 95 °C for 3 min, chilled on ice, and added to binding buffer and protein on ice. After incubation, 2 μL of 50% glycerol plus xylene cyanol was added to each and the samples were loaded on a prechilled 20 × 20 × 0.1 cm, 6% acrylamide (19:1 mono:bis) gel containing 10% glycerol and 0.5× TBE. The gel was run at 15 mA for 4 h in the cold room and exposed to film or a phosphorimager screen.

ACKNOWLEDGMENT

We thank Michelle Haynes Pauling for critical comments on the manuscript. This work was supported by grant GM40478 from the National Institutes of Health.

Manuscript accepted without revision October 14, 1997

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol 215*:403–410.

Ares M Jr, Weiser B. 1995. Rearrangement of snRNA structure during assembly and function of the spliceosome. *Prog Nucleic Acid Res Mol Biol 50*:131–159.

Behrens SE, Tyc K, Kastner B, Reichelt J, Luhrmann R. 1993. Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. *Mol Cell Biol 13*:307–319.

Bennett M, Michaud S, Kingston J, Reed R. 1992. Protein components specifically associated with prespliceosome and spliceosome complexes. *Genes & Dev 6*:1986–2000.

Birney E, Kumar S, Krainer AR. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res 21*:5803–5816.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem 72*:248–254.

Brosi R, Groning K, Behrens SE, Luhrmann R, Kramer A. 1993. Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science 262*:102–105.

Caceres JF, Krainer AR. 1993. Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *EMBO J 12*:4715–4726.

Champion-Arnaud P, Reed R. 1994. The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes & Dev 8*:1974–1983.

Crowe J, Dobeli H, Gentz R, Hochuli E, Stuber D, Henco K. 1994. 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods Mol Biol 31*:371–387.

Durfee T, Becherer K, Chen PL, Yeh SH, Yang Y, Kilburn AE, Lee WH, Elledge SJ. 1993. The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes & Dev 7*:555–569.

Fromont-Racine M, Rain JC, Legrain P. 1997. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet 16*:277–282.

Gozani O, Feld R, 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.

Guthrie C, Fink G. 1991. *Guide to yeast genetics and molecular biology*. San Diego, California: Academic Press.

Harper JW, Adami GR, Wei N, Keyomarsi K, Elledge SJ. 1993. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell 75*:805–816.

Henry M, Borland CZ, Bossie M, Silver PA. 1996. Potential RNA binding proteins in *Saccharomyces cerevisiae* identified as suppressors of temperature-sensitive mutations in NPL3. *Genetics 142*:103–115.

Higgins DG, Thompson JD, Gibson TJ. 1996. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol 266*:383–402.

Hodges PE, Beggs JD. 1994. RNA splicing. U2 fulfils a commitment. *Curr Biol 4*:264–267.

Kenan DJ, Query CC, Keene JD. 1991. RNA recognition: Towards identifying determinants of specificity. *Trends Biochem Sci 16*:214–220.

Kolodziej PA, Young RA. 1991. Epitope tagging and protein surveillance. *Methods Enzymol 194*:508–519.

Kramer A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem 65*:367–409.

Kunkel TA, Bebenek K, McClary J. 1991. Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol 204*:125–139.

Madhani HD, Guthrie C. 1994. Dynamic RNA–RNA interactions in the spliceosome. *Annu Rev Genet 28*:1–26.

Milligan JF, Uhlenbeck OC. 1989. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol 180*:51–62.

Moore MJ, Query CC, Sharp PA. 1993. Splicing of precursors to mRNA in the spliceosome. In: Gestland R, Atkins J, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 303–357.

Nagai K, Oubridge C, Ito N, Avis J, Evans P. 1995. The RNP domain: A sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem Sci 20*:235–240.

Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature 372*:432–438.

Peterson CL, Dingwall A, Scott MP. 1994. Five SWI/SNF gene products are components of a large multisubunit complex required for transcriptional enhancement. *Proc Natl Acad Sci USA 91*:2905–2908.

Ruiz-Lozano P, Doevendans P, Brown A, Gruber PJ, Chien KR. 1997.

Developmental expression of the murine spliceosome-associated protein mSAP49. *Dev Dyn 208*:482–490.
Sayle RA, Milner-White EJ. 1995. RASMOL: Biomolecular graphics for all. *Trends Biochem Sci 20*:374.
Staknis D, Reed R. 1994. Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. *Mol Cell Biol 14*:2994–3005.
Tanaka Y, Ohta A, Terashima K, Sakamoto H. 1997. Polycistronic expression and RNA-binding specificity of the *C. elegans* homologue of the spliceosome-associated protein SAP49. *J Biochem (Tokyo) 121*:739–745.
VanHoy RW, Wise JA. 1996. Molecular analysis of a novel *Schizosaccharomyces pombe* gene containing two RNP consensus-sequence RNA-binding domains. *Curr Genet 29*:307–315.
Wells SE, Neville M, Haynes M, Wang J, Igel H, Ares M Jr. 1996. CUS1, a suppressor of cold-sensitive U2 snRNA mutations, is a novel yeast splicing factor homologous to human SAP 145. *Genes & Dev 10*:220–232.
Yan D, Ares M Jr. 1996. Invariant U2 RNA sequences bordering the branchpoint recognition region are essential for interaction with yeast SF3a and SF3b subunits. *Mol Cell Biol 16*:818–828.