The EMBO Journal Vol.18 No.21 pp.6119–6133, 1999

Functional interaction of a novel 15.5kD [U4/U6·U5] tri-snRNP protein with the 5′ stem–loop of U4 snRNA

Stephanie Nottrott[1], Klaus Hartmuth[1], Patrizia Fabrizio[1], Henning Urlaub[1], Ivan Vidovic[1], Ralf Ficner[1] and Reinhard Lührmann[1,2,3]

[1]Institut für Molekularbiologie und Tumorforschung, Philipps-Universität Marburg, Emil-Mannkopff-Strasse 2, D-35037 Marburg and [2]Department of Cellular Biochemistry, Max-Planck-Institute of Biophysical Chemistry, Am Fassberg 11, D-37070 Göttingen, Germany

[3]Corresponding author
e-mail: luehrmann@imt.uni-marburg.de

Activation of the spliceosome for splicing catalysis requires the dissociation of U4 snRNA from the U4/U6 snRNA duplex prior to the first step of splicing. We characterize an evolutionarily conserved 15.5 kDa protein of the HeLa [U4/U6·U5] tri-snRNP that binds directly to the 5′ stem–loop of U4 snRNA. This protein shares a novel RNA recognition motif with several RNP-associated proteins, which is essential, but not sufficient for RNA binding. The 15.5kD protein binding site on the U4 snRNA consists of an internal purine-rich loop flanked by the stem of the 5′ stem–loop and a stem comprising two base pairs. Addition of an RNA oligonucleotide comprising the 5′ stem–loop of U4 snRNA (U4SL) to an *in vitro* splicing reaction blocked the first step of pre-mRNA splicing. Interestingly, spliceosomal C complex formation was inhibited while B complexes accumulated. This indicates that the 15.5kD protein, and/or additional U4 snRNP proteins associated with it, play an important role in the late stage of spliceosome assembly, prior to step I of splicing catalysis. Our finding that the 15.5kD protein also efficiently binds to the 5′ stem–loop of U4atac snRNA indicates that it may be shared by the [U4atac/U6atac·U5] tri-snRNP of the minor U12-type spliceosome.
Keywords: pre-mRNA splicing/RNA–protein interaction/snRNPs/U4 snRNA

Introduction

The removal of introns from nuclear mRNA precursors occurs via a two-step transesterification mechanism involving both 5′ and 3′ splice site cleavages and exon ligation (reviewed in Moore *et al*., 1993; Burge *et al*., 1999). The splicing reaction is catalyzed by a large ribonucleoprotein complex called the spliceosome. The spliceosome is composed of four small ribonucleoprotein particles (U1, U2, U5 and U4/U6 snRNPs) and numerous non-snRNP protein splicing factors, which assemble on the pre-mRNA in an ordered manner. The spliceosomal snRNPs contain five small nuclear RNAs (U1, U2, U4, U5 and U6 snRNA) and >50 distinct proteins that are either common to all snRNPs or specific for a given particle (reviewed in Will and Lührmann, 1997).

In the course of splicing, a network of both snRNA–snRNA and snRNA–pre-mRNA interactions is formed (Moore *et al.*, 1993; Madhani and Guthrie, 1994; Nilsen, 1998; Staley and Guthrie, 1998). During the early stages of spliceosome assembly, the U1 and U2 snRNPs recognize and base pair with the 5′ splice site and the branch site, respectively, thereby generating the so-called pre-spliceosome, or complex A. Mature spliceosomes (i.e. complexes B and C) are formed subsequently by the association of the [U4/U6.U5] tri-snRNP particle with the pre-spliceosome. Prior to the first step of splicing, the snRNAs, in particular the U4 and U6 snRNAs of the [U4/U6·U5] tri-snRNP complex, undergo dramatic conformational rearrangements, which are thought to be driven by spliceosomal proteins of the DEAD/DEXH-box family of ATP-dependent RNA unwindases (Staley and Guthrie, 1998). Within the tri-snRNP complex, the U4 and U6 snRNAs are base-paired and form a phylogenetically highly conserved Y-shaped U4/U6 interaction domain, consisting of stem I and II separated by the 5′ stem–loop of U4 snRNA (Bringmann *et al.*, 1984; Hashimoto and Steitz, 1984; Rinke *et al.*, 1985; Brow and Guthrie, 1988; see Figure 2). After association of the tri-snRNP with the pre-spliceosome, both stems of the U4/U6 interaction domain are disrupted. While the region of U6 snRNA constituting U4/U6 stem II forms a new intramolecular stem–loop, the region of U6 comprising stem I base pairs with U2 RNA to form part of the catalytic centre of the spliceosome (Datta and Weiner, 1991; Wu and Manley, 1991; Madhani and Guthrie, 1992; Sun and Manley, 1995). Moreover, the conserved ACAGAG sequence of U6 snRNA base pairs with the 5′ splice site of the pre-mRNA (Fabrizio and Abelson, 1990; Sawa and Abelson, 1992; Kandels-Lewis and Séraphin, 1993; Lesser and Guthrie, 1993; Sontheimer and Steitz, 1993). In contrast to U6, U4 snRNA is released from the spliceosome or remains only loosely attached to it (Lamond *et al.*, 1988; Yean and Lin, 1991). It is thought that following the dissociation of the post-spliceosomal complexes, the U4/U6 interaction domain re-forms for a new round of splicing (Moore *et al.*, 1993). Thus, U4 snRNA has been proposed to act as an RNA chaperone that delivers U6 to the spliceosome sequestering a catalytically active domain of U6 snRNA until the dissociation of the U4/U6 snRNA duplex activates this domain for its function in splicing (Brow and Guthrie, 1989). Similar structural rearrangements are also observed with the recently discovered U4atac and U6atac snRNAs of the so-called minor or U12-type spliceosome which are functional analogues of U4 and U6. Both RNAs also form a Y-shaped U4atac/U6atac interaction domain with the U4atac snRNA 5′ stem–loop separating the two intermolecular helices I and II (Tarn and Steitz, 1996). During minor spliceosome assembly, the U4atac and U6atac snRNAs also dissociate,

and U6atac snRNA engages in base pairing interactions with the U12 snRNA (the functional analogue of the U2 snRNA) and the 5′ splice site (Tarn and Steitz, 1996; Yu and Steitz, 1997; Incorvaia and Padgett, 1998).

U4 snRNA is not as phylogenetically conserved as U6 snRNA, except for the structural organization of the U4/U6 RNA interaction domain and the loop sequence of the U4 snRNA 5′ stem–loop (Brow and Guthrie, 1988; Guthrie and Patterson, 1988). Sequences of the U4 snRNA required for splicing have been investigated *in vivo* in yeast and *Xenopus* oocytes, as well as *in vitro* in HeLa cell splicing extracts. As expected, those regions of U4 snRNA known to base pair with U6 were found to be important for U4/U6 interaction and thus also for splicing. Moreover, distinct regions in the central domain of U4 snRNA appear to contribute to splicing, while the 3′-terminal Sm domain was shown to be dispensable for splicing *in vitro* (Vankan *et al.*, 1992; Wersig and Bindereif, 1992; and references therein). Consistent with its high degree of evolutionary conservation, the 5′ stem–loop of U4 snRNA was demonstrated to be essential for splicing in all systems investigated. For example, deletion of the 5′ stem–loop from the *Saccharomyces cerevisiae* U4 snRNA gene *SNR14* led to cell inviability (Bordonné *et al*., 1990). Mutational analyses of U4 snRNA in *Xenopus* oocytes and in the HeLa *in vitro* splicing system demonstrated that the 5′ stem–loop is essential for pre-mRNA splicing and spliceosome assembly (Vankan *et al.*, 1992; Wersig and Bindereif, 1992 and references therein). Since the U4 snRNA 5′ stem–loop is dispensable for U4/U6 RNA base pairing *in vitro*, it has been suggested that it functions in spliceosome assembly at a stage subsequent to U4/U6 snRNP formation (Wersig and Bindereif, 1992). This idea is also supported by the results obtained by mutational analysis of the U4 snRNA 5′ stem–loop in yeast (Bordonné *et al*., 1990; Hu *et al*., 1995).

While the 5′ stem–loop of U4 snRNA is clearly essential for U4 snRNA function in pre-mRNA splicing, it is not known whether its prime function is to interact directly with other RNAs in the spliceosome, or to recruit snRNP proteins to the spliceosome. Consistent with the latter idea, the U4 snRNA 5′ stem–loop is accessible to oligonucleotide-directed RNase H cleavage only after phenolization of snRNPs from both HeLa cell nuclear and yeast extracts (Black and Steitz, 1986; Xu *et al.*, 1990), suggesting that it is associated with one or more proteins. Aside from the Sm proteins, which are bound to the 3′-terminal Sm site of the U4 snRNA, relatively little is known about other U4 snRNA-associated proteins. The WD40 protein Prp4p of the yeast *S.cerevisiae* was the first protein shown to associate with U4/U6 snRNAs (Banroques and Abelson, 1989; Peterson-Bjørn *et al*., 1989). Interestingly, antibodies directed against Prp4p were shown to precipitate the 5′ portion of U4 snRNA, including the conserved 5′ stem–loop (Bordonné *et al*., 1990; Xu *et al*., 1990). However, direct interaction of Prp4p with U4 snRNA could not be demonstrated. A second U4 snRNP candidate protein in yeast is Prp3p, which is present in U4/U6 snRNPs and has been shown to interact directly with Prp4p (Anthony *et al*., 1997; Ayadi *et al*., 1998).

In the human 25S [U4/U6·U5] tri-snRNP, the 60kD and 90kD proteins have been identified as orthologues of the yeast Prp4p and Prp3p proteins, respectively, and were also found to be associated specifically with the U4/U6 snRNP particle (Horowitz *et al*., 1997; Lauber *et al*., 1997). The U4/U6-specific 60kD and 90kD proteins form a tight heteromeric complex with a 20kD cyclophilin, which has also been identified as a U4/U6-specific protein (Horowitz *et al*., 1997; Teigelkamp *et al*., 1998). It should be noted that none of these proteins contains a canonical RNA-binding motif such as a consensus RBD, a KH domain, an RGG box or an arginine-rich motif (reviewed in Burd and Dreyfuss, 1994; Cusack, 1999), nor has direct binding of the 20/60/90kD heteromer to U4 or U6 snRNA been reported so far.

Here we describe the molecular characterization of the 15.5kD protein of the HeLa 25S [U4/U6·U5] tri-snRNP. The 15.5kD protein belongs to a group of tri-snRNP proteins, with mol. wts of 15.5, 61 and 63/65 kDa, which have not yet been assigned to one of the snRNPs constituting the tri-snRNP complex (for a recent compilation of human tri-snRNP proteins, see Will and Lührmann, 1997). We show that the evolutionarily conserved 15.5kD protein binds directly to the 5′ stem–loop of human U4 snRNA as well as to U4atac snRNA in a specific manner. The latter finding indicates that the 15.5kD protein is also present in the [U4atac/U6atac·U5] tri-snRNP of the minor U12-type spliceosome. The 15.5kD protein is a new member of a family of RNP-associated proteins with a novel RNA recognition motif. By mutagenesis experiments, we have also characterized in detail the 15.5kD protein-binding site which consists of a conserved purine-rich structural element of the U4 snRNA 5′ stem–loop. Finally, we have investigated the functional importance of the interaction between the 15.5kD protein and the U4 snRNA 5′ stem–loop in pre-mRNA splicing. Our results indicate that the 15.5kD and/or additional proteins associated with it play an important role in the late stage of spliceosome assembly, after the [U4/U6·U5] tri-snRNP has been integrated into the spliceosome, but prior to the first catalytic step of splicing.

Results

Cloning of the HeLa 15.5kD tri-snRNP protein

The 15.5kD protein was isolated by fractionating proteins from purified HeLa 25S [U4/U6·U5] tri-snRNPs as described previously (Lauber *et al*., 1997). Microsequencing of the 15.5kD protein yielded two distinct peptide sequences (see Materials and methods). A database search with these peptide sequences identified several expressed sequence tags (ESTs) from different human tissues, as well as one human cDNA isolated from an embryonic brain cDNA library, that encoded both peptide sequences (Saito *et al*., 1996). Screening of a HeLa cDNA library with the EST R72749 led to the isolation of a clone with an open reading frame (ORF) encoding a putative protein of 128 amino acids (Figure 1A). The identity of this cDNA was verified by several criteria. First, both peptide sequences obtained by microsequencing of the 15.5kD tri-snRNP protein were present in the predicted amino acid sequence of the cDNA-encoded protein. Furthermore, antibodies raised against a recombinant fusion protein containing GST and the cDNA-encoded protein specifically recognized the endogenous 15.5kD protein on immunoblots when proteins from gradient-purified 25S

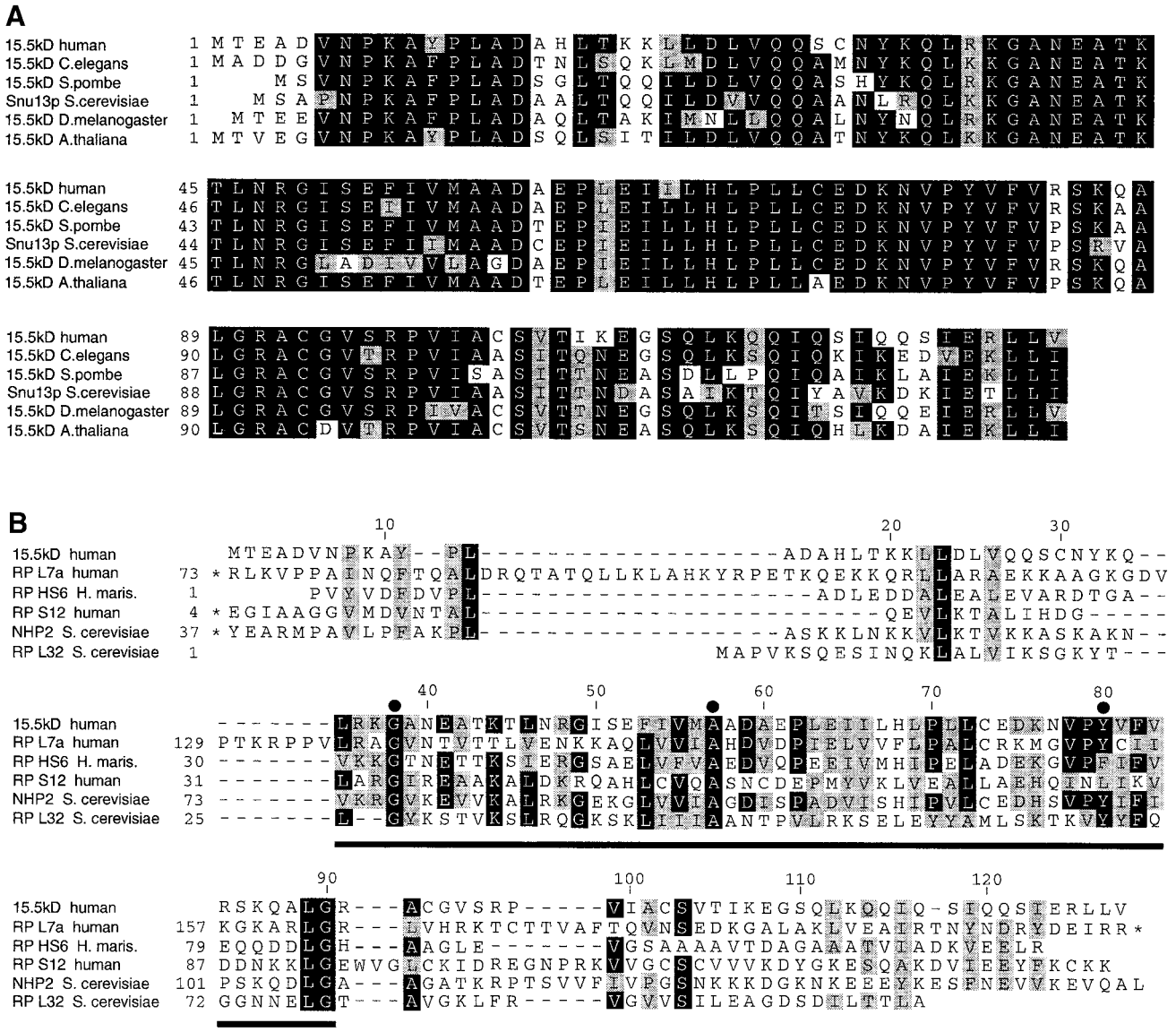

Fig. 1. Sequence alignment of the human 15.5kD protein with highly homologous proteins. (**A**) Human 15.5kD protein (AF155235) is aligned with its orthologues from *Caenorhabditis elegans* (Q21568), *Schizosaccharomyces pombe* (AF087136), *Saccharomyces cerevisiae* (P39990), *Drosophila melanogaster* (EST GH 03082) and *Arabidopsis thaliana* (sequence combined from EST T14197 and EST H76484). Identical residues are boxed in black and conserved residues (grouped DE, KRH, LIVAMPYFW and TSQCN) are shaded grey. The amino acid positions are indicated by numbers. (**B**) Human 15.5kD protein is aligned with human ribosomal protein L7a (P11518) and S12 (P47840), with *Haloarcula marismortui* ribosomal protein HS6 (P12743), *S.cerevisiae* protein NHP2 (P32495) and ribosomal protein L32 from *S.cerevisiae* (P14120). Residues identical in at least four out of six sequences are boxed in black and conserved residues are shaded grey. The sequence motif from position 35 to 90 of the human 15.5kD protein is underlined with a black bar. Amino acids of the human 15.5kD protein exchanged by site-directed mutagenesis are indicated by a black dot. The amino acid positions are indicated by numbers. Truncated sequences are marked with an asterisk at the point of truncation. Multiple sequence alignments were done using the Clustal method (Higgins and Sharp, 1988) and optimized by visual inspection.

HeLa [U4/U6·U5] tri-snRNPs were used as a source of antigen. Finally, *in vitro* translated 15.5kD protein co-migrated on SDS–polyacrylamide gels with the endogenous 15.5 kDa tri-snRNP protein (data not shown).

The 15.5kD tri-snRNP protein is evolutionarily highly conserved and belongs to a family of proteins with a novel putative RNA-binding domain

A database search with the HeLa 15.5kD tri-snRNP protein identified apparent orthologues in *Caenorhabditis elegans*, *S.cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Arabidopsis thaliana* which exhibited an extraordinarily high degree of homology with the human protein (71–77% identity, 83–89% similarity, Figure 1A). Consistent with the idea that it is a 15.5kD orthologue, the *S.cerevisiae* protein, termed Snu13p, is present in purified yeast [U4/U6·U5] tri-snRNPs (Gottschalk *et al.*, 1999; Stevens and Abelson, 1999). The human 15.5kD tri-snRNP protein shares significant homology with several other proteins in the database which are clearly not 15.5kD orthologues. These include the human ribosomal proteins L7a and S12, the yeast ribosomal protein L32, the ribosomal protein HS6 from *Haloarcula marismortui* and the yeast protein NHP2. NHP2 recently has been identified as a common protein of all H/ACA box small nucleolar

RNPs (Henras *et al*., 1998; Watkins *et al*., 1998). As previously noted by Henras *et al*. (1998), all of these proteins share a strongly conserved central region comprised of 56 amino acid residues (spanning positions 35–90 of the human 15.5kD protein, Figure 1B). It should be noted, however, that several short regions of homology are also detected between the various members of this family in the flanking N- and C-terminal regions (Figure 1B). Consistent with the fact that all of these proteins are components of RNP complexes, the N-terminal 25 amino acid residues of the conserved central region previously have been postulated to contain the consensus sequence of a novel, putative RNA-binding motif (Koonin *et al.*, 1994).

The 15.5kD protein binds specifically to the 5′ stem–loop of U4 snRNA

The observation that the 15.5kD protein belongs to a family of proteins sharing a putative RNA-binding motif (see above) prompted us to investigate whether it interacts directly with one of the snRNAs of the [U4/U6·U5] tri-snRNP. For this purpose, co-precipitation experiments (GST pull-downs) were performed with purified recombinant GST–15.5kD fusion protein. GST–15.5kD protein was first coupled to glutathione–Sepharose and then incubated with a mixture of snRNAs which were obtained by phenolizing immunoaffinity-purified HeLa snRNPs at 0°C to maintain U4/U6 base pairing (Bringmann *et al*., 1984). As shown in Figure 2A, the U4/U6 snRNA duplex was co-precipitated specifically with the GST–15.5kD protein, but not with GST alone (upper panel, lanes 2 and 3). This clearly demonstrates that the 15.5kD protein interacts directly with one or both of these two snRNAs. In a separate experiment, the snRNA mixture was incubated for 1 min at 90°C (in order to dissociate the U4/U6 snRNA duplex) prior to addition of the GST–15.5.kD protein. Under these conditions, exclusively U4 snRNA was co-precipitated, demonstrating that the U4 snRNA is sufficient for specific and stable complex formation with the 15.5kD protein (Figure 2A, upper panel, lane 4).

A specific interaction of the 15.5kD protein with the U4 snRNA was also observed with an alternative co-immunoprecipitation approach. In this assay, we incubated *in vitro* translated, ^{35}S-labelled 15.5kD protein with m^{7}G-capped, *in vitro* transcribed U1, U2, U4, U5 or U6 snRNA. The m^{7}G-capped snRNAs were then precipitated with the monoclonal anti-cap antibody H20. As monitored by SDS–PAGE, the *in vitro* translated 15.5kD protein was co-precipitated only with the U4 snRNA (Figure 2A, lower panel).

In order to narrow down the RNA-binding site of the 15.5kD protein on the U4 snRNA, we performed hydroxyl radical footprinting (Celander and Cech, 1990; Hüttenhofer and Noller, 1992; see Materials and methods) of the *in vitro* reconstituted, binary U4 snRNA–15.5kD protein complex. Primer extension analysis revealed that riboses in two regions of the U4 snRNA backbone were clearly protected from cleavage by hydroxyl radicals at increasing concentrations of recombinant 15.5kD protein (FP1 and FP2, Figure 2B). Intriguingly, both footprints map to the phylogenetically conserved 5′ stem–loop of U4 snRNA, encompassing positions A29–G35 (FP2) and G43–C47 (FP1) (Figure 2C).

Structural determinants of the U4 snRNA 5′ stem–loop required for specific recognition by the 15.5kD protein

An initial clue to understand how the structural elements in the U4 snRNA 5′ stem–loop might contribute to the specific recognition by the 15.5kD protein came from phylogenetic sequence comparisons (see also Guthrie and Patterson, 1988). Close inspection of the 5′ stem–loop sequences of U4 snRNAs currently available in the database revealed that nucleotides U31, G32, A33, G43 and A44 (nucleotide positions according to the sequence of the human U4 snRNA) are 100% conserved in all U4 snRNAs, except for U31 in *Trypanosoma brucei*. Furthermore, positions A29 and A30 are always purines, exceptions being A29 in *T.brucei* and *Physarum polycephalum*. Figure 3A shows the 5′ stem–loop sequences from human, yeast and *Drosophila* U4 snRNA as well as from human U4atac snRNA as selected examples to illustrate this phylogenetic conservation. The nucleotide corresponding to position 30 in human U4 snRNA is either an A or G. In addition, two base pairs potentially form between nucleotides G34 and C42, and G35 and C41, respectively. The potential G35–C41 base pair is phylogenetically conserved between the human, yeast and *Drosophila* loop sequences and is inverted to a C-G base pair in the human U4atac snRNA loop (Figure 3A). The presence of such a compensatory base change provides strong evidence for a base pair (reviewed in Michel and Costa, 1998). Similarly, the existence of the neighbouring base pair (G34–C42) is also supported by compensatory base mutations. That is, the canonical G–C base pair is replaced by a G·U base pair in human U4atac snRNA and by an A–U base pair in *Drosophila* U4 snRNA (Figure 3A). In all known U4 snRNA sequences, except in *T.brucei*, both base pairs show 100% covariation (data not shown). The nature of the remaining loop nucleotides (positions U36–U40 of the human U4 snRNA) varies significantly among the different U4 snRNAs with no apparent conservation in primary sequence. Finally, while the stem as such is conserved, its precise sequence is not, except for the polarity of the G–C base pair closing the loop, which is conserved in all U4 snRNA sequences (Figure 3A; data not shown).

To investigate whether the 5′ stem–loop is sufficient for binding the 15.5kD protein, we used a chemically synthesized mini stem–loop RNA oligonucleotide corresponding to nucleotides 26–47 of the human U4 snRNA (henceforth termed human U4SL) which contained an additional G–C base pair to stabilize the short stem. The interaction of recombinant 15.5kD protein with this RNA oligonucleotide was then analysed by electrophoretic mobility shift assays (EMSAs). As shown in lane 1 of Figure 3B, the 15.5kD protein effectively forms a complex with the human U4SL, demonstrating that the 5′ stem–loop of U4 snRNA is sufficient for 15.5kD protein binding. Efficient binding of the U4 snRNA 5′ stem–loop oligonucleotide was likewise observed in a pull-down experiment carried out with a GST–15.5kD fusion protein under the same conditions used for U4 snRNA–15.5kD protein complex formation as shown in Figure 2A (data not shown).

The importance of the nucleotides at positions 29–33, 43 and 44 of the human U4 snRNA 5′ stem–loop for complex formation with the 15.5kD protein was next

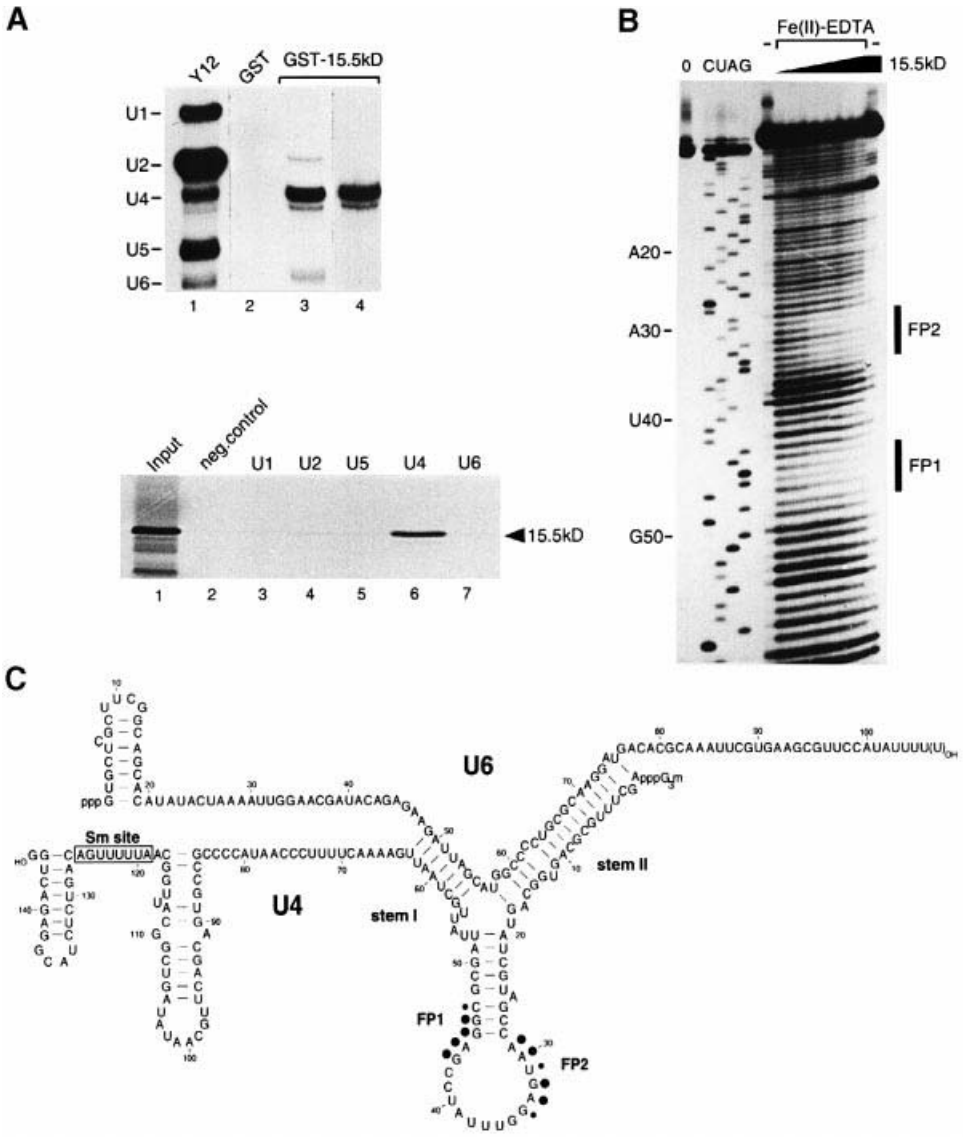

Fig. 2. The 15.5kD protein binds directly to the 5′ stem–loop of U4 snRNA. (**A**) Upper panel: U4 snRNA is co-precipitated specifically with GST–15.5kD in a GST pull-down assay. UsnRNAs were co-precipitated after incubation with immunoaffinity-purified HeLa snRNPs by monoclonal antibody Y12 (Lerner *et al.*, 1981) coupled to protein A–Sepharose (lane 1). Control precipitation of a mixture of native UsnRNAs with glutathione–Sepharose-coupled GST (lane 2). Co-precipitation of UsnRNAs after incubation of a mixture of native (lane 3) or heat-denatured (lane 4) UsnRNAs with glutathione–Sepharose-coupled GST–15.5kD. Bound UsnRNAs were recovered by phenol extraction and ethanol precipitation, 3′-end-labelled and separated on a denaturing 10% polyacrylamide–7 M urea gel. The identity of the precipitated UsnRNAs was verified by Northern blot analysis (data not shown). Lower panel: *in vitro* translated, ^{35}S-labelled 15.5kD protein was co-immunoprecipitated specifically with U4 snRNA by protein A–Sepharose-coupled monoclonal antibody H20 as monitored by SDS–PAGE. Lane 1: 10% input of ^{35}S-labelled 15.5kD protein. Co-immunoprecipitations were performed without RNA (lane 2) or in the presence of the indicated, *in vitro* transcribed, m^{7}G-capped UsnRNAs (lanes 3–7). (**B**) Hydroxyl radical footprint of the U4 snRNA–15.5kD protein complex as analysed by primer extension. Riboses protected from cleavage at increasing 15.5kD protein concentration (0.3–134 pmol) are marked with black bars and designated FP1 and FP2. Since hydroxyl radical attack of a nucleotide results in its complete removal from the RNA chain, reverse transcriptase will stop one nucleotide before the site of attack. The first and the last lanes (labelled '–') contain unmodified U4 snRNA to control for spontaneous stops by the reverse transcriptase. C, U, A and G refer to dideoxysequencing reactions and correspond to the sequence of endogenous HeLa U4 snRNA. 0 indicates a control primer extension with unmodified U4 snRNA where no ddNTPs were added to the reaction. The position of every tenth nucleotide of the U4 snRNA is indicated on the left. (**C**) Secondary structure model of the human U4/U6 snRNA duplex according to Bringmann *et al.* (1984) and Brow and Guthrie (1988). The positions of the footprints, FP1 and FP2, are indicated by black dots. Large dots indicate strong and smaller dots indicate weaker protection of riboses from hydroxyl radicals. The nucleotide positions of the U4 and U6 snRNAs are indicated by numbers.

investigated with a battery of chemically synthesized U4SL mutant RNA constructs. From the combined results, we can classify the seven nucleotides into at least two groups as summarized in Figure 3C. At positions 29 and 30, the requirement for a particular nucleotide is less stringent, in that the adenosines can be replaced individually by guanosines without any loss of 15.5kD protein-binding activity (Figure 3B, lanes 3 and 4). If both positions contain guanosines, protein binding is reduced slightly but reproducibly (lane 5). In contrast, if the two adenosines are replaced simultaneously by cytidines, protein-binding activity is lost (data not shown). Furthermore, deletion of one adenosine residue has a deleterious effect on the capability of the resulting mutant to interact with the 15.5kD protein (lane 7). These data indicate a preference for purines at positions 29 and 30 of the loop.

In good agreement with their phylogenetic conservation (see above), the identity of nucleotides U31, G32, A33, G43 and A44 is crucial for protein-binding activity *in vitro* as analysed by bandshift assays. A summary of the results obtained with all the U4SL mutants tested is shown in Figure 3C and selected examples of the mobility shift experiments are shown in Figure 3B (lanes 6, 9, 10 and 11). In all cases, both transitions and transversions of these nucleotides led to the complete loss of binding activity. Finally, we show that the polarity of the base pair closing the loop is not crucial for recognition by the 15.5kD protein: a U4SL mutant with an inverted base pair exhibits only a slightly lower degree of binding activity (Figure 3B, lane 8).

In support of the RNA structural requirements for recognition by the 15.5kD protein described so far, chemically synthesized mini stem–loops of the human U4atac and yeast U4 snRNAs also bound the 15.5kD protein (Figure 3B, lanes 12 and 13; see also Figure 3A for sequence comparison). Interestingly, as shown by competition assays (Figure 3D), the 15.5kD protein bound to the U4atac stem–loop oligonucleotide even more efficiently than to the human U4SL oligonucleotide. An ~5-fold higher concentration of U4SL, as compared with U4atac oligonucleotide, was required to completely inhibit binding of the 15.5kD protein to U4 snRNA (Figure 3D, compare lanes 3–6 with 11–14). The inhibition of U4 snRNA–15.5kD protein complex formation by both 5′ stem–loop oligonucleotides was specific, since equivalent concentrations of a mutant U4SL oligonucleotide did not compete for 15.5kD protein binding (Figure 3D, lanes 7–10).

The interaction of the 15.5kD protein with the U4atac SL was significant for two reasons. First, this indicates that the 15.5kD protein is most likely to be shared by the major and minor tri-snRNP particle. Second, the fact that the putative G34–C42 and G35–C41 base pairs in human U4 snRNA are changed to G39·U47 and C40–G46 in human U4atac snRNA without loss of 15.5kD protein binding suggests that the base pair formation between these nucleotides and not their identity may be of prime importance for proper protein–RNA complex formation *in vitro*. Consistent with this idea, replacement of the putative G39·U47 base pair in the U4atac snRNA 5′ stem–loop with a U39–A47 base pair had no effect on 15.5kD protein binding (Figure 3B, lane 16). Interestingly, a single point mutation changing U47 to A completely abolished complex formation, while mutation of G39 to U still allowed efficient 15.5kD protein binding (lanes 15 and 18, respectively). These results indicate that non-canonical G·U and U·U, but not G·A base pairs at this position of the U4atac 5′ stem–loop are compatible with recognition by the 15.5kD protein. Finally, 15.5kD protein binding was completely lost when the putative C40–G46 base pair was disrupted by single base mutations such as C40A (Figure 3B, lane 17), experimentally supporting the importance of this base pair in U4atac SL for protein binding.

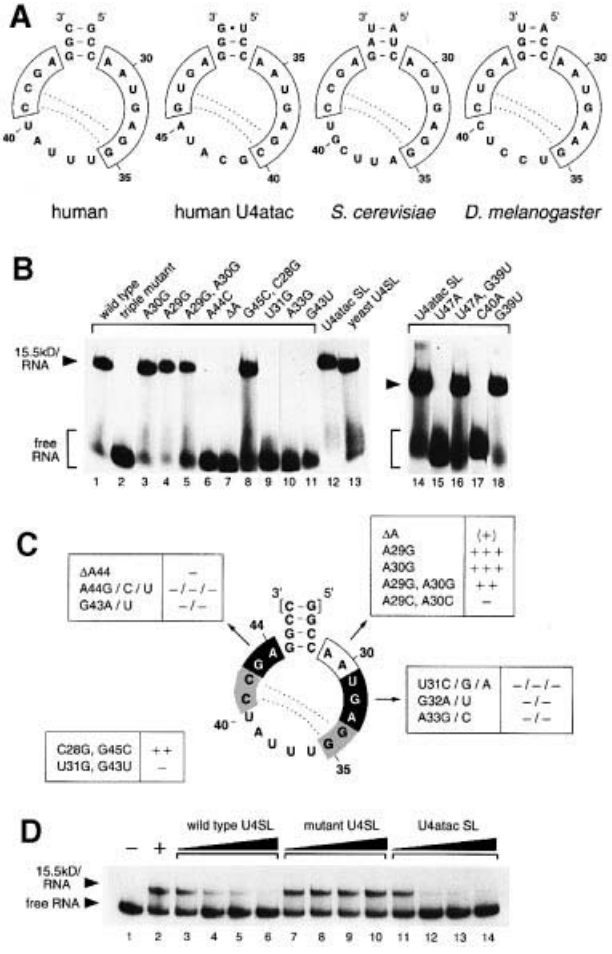

Fig. 3. The 15.5kD protein recognizes phylogenetically conserved nucleotides of the U4 snRNA 5′ stem–loop. (**A**) 5′ Stem–loop sequences of human, *S.cerevisiae* and *D.melanogaster* (D00043) U4 snRNA and human U4atac snRNA. Nucleotide positions are indicated by numbers. Phylogenetically conserved nucleotides are boxed, and potential base pairs are indicated by dashed lines. (**B**) EMSA of the interaction between the 15.5kD protein and U4 snRNA 5′ stem–loop oligonucleotides. Recombinant 15.5kD protein was reconstituted *in vitro* with 5′-end-labelled RNA oligonucleotides and the protein–RNA complexes were resolved on a 10% native polyacrylamide gel. The position of the protein–RNA complex is indicated on the left. Lanes 1–11: bandshifts of wild-type and mutant RNA oligonucleotides corresponding to positions 26–47 of the human U4 snRNA 5′ stem–loop; the triple mutant RNA oligonucleotide shown in lane 2 contains the following substitutions: A30C, U31G, G32U (according to Vankan *et al.*, 1992). Lane 12: RNA oligonucleotide corresponding to positions 31–52 of human U4atac snRNA. Lane 13: RNA oligonucleotide corresponding to positions 26–48 of *S.cerevisiae* U4 snRNA. Lanes 14–18: bandshifts of mutant RNA oligonucleotides corresponding to position 30–51 of the human U4atac snRNA 5′ stem–loop. All RNA oligonucleotides contain an additional G–C base pair at the top of the stem. (**C**) Substitution and deletion mutants of the human U4 snRNA 5′ stem–loop oligonucleotide (human U4SL). The efficiency of 15.5kD protein binding for the mutant RNA oligonucleotides, as determined by the bandshift analysis, is indicated by '+' and '–'. Essential nucleotides are boxed in black and nucleotides involved in potential base pairing are shaded grey. The G–C base pair which is not found in the human wild-type U4 snRNA sequence is shown in brackets. (**D**) The 5′ stem–loop suffices to compete for the 15.5kD protein–U4 snRNA interaction. Inhibition of the U4 snRNA–15.5kD protein complex formation was assayed by EMSA (see Materials and methods for details). The competitors used were human U4SL (lanes 3–6), a triple mutant U4SL (A30C, U31G, G32U, see above; lanes 7–10) and U4atac SL (lanes 11–14). No competition was observed at equivalent concentrations of mutant U4SL. Lane 1: U4 snRNA without 15.5kD protein. Lane 2: U4 snRNA incubated with 15.5kD protein at a final protein concentration of 1 μM. The competitor concentrations assayed were 5 (lanes 3, 7 and 11), 50 (lanes 4, 8 and 12), 100 (lanes 5, 9 and 13) and 250 (6, 10 and 14) pmol. The position of the U4 snRNA–15.5kD protein complex and free U4 snRNA is indicated on the left.

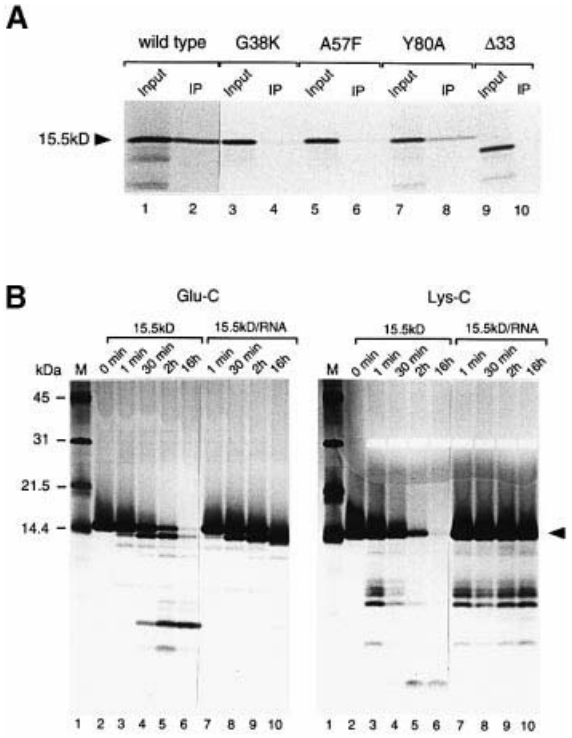

Fig. 4. Characterization of the 15.5kD RNA recognition motif. (**A**) Binding of *in vitro* translated, ^{35}S-labelled wild-type and mutant 15.5kD proteins to human U4 snRNA. The protein–RNA complex was reconstituted *in vitro* and co-immunoprecipitation of wild-type and mutant 15.5kD protein with protein A–Sepharose-coupled monoclonal antibody H20 was monitored by SDS–PAGE. Input: 10% input of *in vitro* translated 15.5kD protein. IP: co-immunoprecipitated 15.5kD protein bound to U4 snRNA. (**B**) Proteolytic digestion of the *in vitro* reconstituted protein–RNA complex. Free recombinant 15.5kD protein and 15.5kD protein bound to the U4atac snRNA 5′ stem–loop oligonucleotide (U4atac SL) were incubated with endoproteases Glu-C (left panel) and Lys-C (right panel) for 0 min, 1 min, 30 min, 2 h or 16 h. Protease digestion products subsequently were separated by SDS–PAGE and visualized by silver staining. The position of the full-length 15.5kD protein is indicated on the right by an arrowhead. The negatively stained double band at 31kD in the right panel is the enzyme Lys-C.

Taken together, our results indicate that the 15.5kD protein-binding site consists of an asymmetric internal purine-rich loop flanked by the long stem of the 5′ stem–loop and a short stem of two base pairs.

The central region of the 15.5kD protein is essential but not sufficient for U4 snRNA binding

A conspicuous feature of the 15.5kD protein is a 56 amino acid central region which exhibits significant homology with several other proteins that are all components of RNP complexes (Figure 1B, see also Henras *et al.*, 1998). Having identified the 15.5kD protein as a novel U4 snRNA-binding protein, we were interested to learn more about the contribution of the central and also the flanking regions of the 15.5kD protein to its RNA-binding activity. To this end, three highly conserved amino acids within the central region of the protein were mutated (see Figure 1B) and the ability of the *in vitro* translated, ^{35}S-labelled mutant proteins to bind U4 snRNA was assayed by co-immunoprecipitation analyses with an anti-cap antibody. As shown in Figure 4A, mutation of Gly38 to lysine (G38K), as well as the exchange of Ala57 with phenylalanine (A57F), completely abolished binding of the 15.5kD protein to U4 snRNA (lanes 4 and 6). This demonstrates that both amino acids, which are 100% conserved among those proteins containing this central homologous region, are essential for 15.5kD RNA-binding activity. Mutation of Tyr80 to alanine (Y80A) strongly reduced U4 snRNA binding (by ~50%) but did not abolish it. Thus, an aromatic amino acid residue at this position is clearly favourable but not essential, consistent with the fact that the human ribosomal protein S12 contains a leucine residue at the equivalent position (see Figure 1B). To investigate the possible contribution of the C-terminal region of the 15.5kD protein to RNA-binding activity, we constructed a deletion mutant of the protein lacking the C-terminal 33 amino acids. As shown in Figure 4A, deletion of this region completely abolished binding to the U4 snRNA (lane 10), indicating that C-terminal residues are also essential for binding activity.

These data further suggest that the evolutionarily conserved central region of the 15.5kD protein is necessary, but not sufficient for RNA binding. This idea is also consistent with the results obtained from partial proteolytic digestion of the protein in the absence or presence of the U4atac SL RNA oligonucleotide. As shown in Figure 4B, binding of U4atac SL to the 15.5kD protein has a striking effect on the stability of the protein towards proteases. While the free protein was hydrolysed rapidly by either Glu-C or Lys-C proteases (>70–80% after 2 h and 100% after 16 h, lanes 2–6 of both panels), the RNA-bound protein was completely protected from protease cleavage by Lys-C even after incubating for 16 h (lanes 7–10, right panel). Similar results were obtained with trypsin and chymotrypsin (data not shown) and Glu-C, except that the 15.5kD protein was cleaved by Glu-C and chymotrypsin close to the C- or N-terminus even when it was complexed with the U4atac SL RNA oligonucleotide (Figure 4B, left panel and data not shown). Importantly, however, the slightly shorter fragment remained stable over the entire 16 h incubation period (lanes 7–10). The dramatic protection of the 15.5kD protein against protease hydrolysis was strictly dependent on U4 snRNA SL–15.5kD protein complex formation. In the presence of a mutant U4SL, inactive in 15.5kD protein binding as determined by a gel mobility shift assay (see Figure 3B, lane 2), protection of the 15.5kD protein from rapid protease digestion was not observed (not shown). Taken together, our combined results suggest that the 15.5kD protein in its entirety may be required for specific interaction with the U4 snRNA 5′ stem–loop.

The U4 snRNA stem–loop oligonucleotide inhibits the first step of pre-mRNA splicing

Previous mutagenesis analyses *in vivo* and *in vitro* demonstrated that the 5′ stem–loop of U4 snRNA is essential for pre-mRNA splicing (see Introduction). However, it was not clear from these studies whether or not the observed deleterious effects were due to the disruption of the interaction of U4 snRNA 5′ stem–loop-binding proteins. We were therefore interested in investigating the role of the 15.5kD protein in pre-mRNA splicing *in vitro* more directly. For this purpose, we added the human, wild-type U4SL RNA oligonucleotide to an *in vitro* pre-mRNA splicing reaction, reasoning that U4SL might

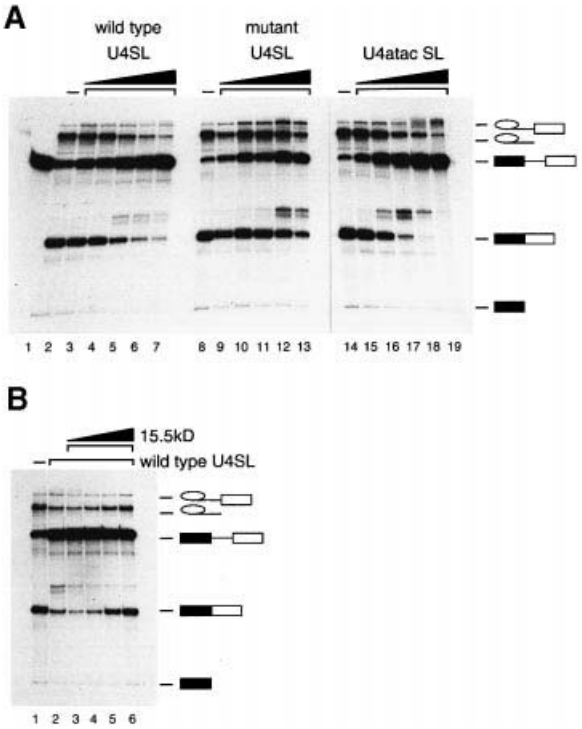

Fig. 5. An excess of wild-type U4 and U4atac SL inhibits pre-mRNA splicing *in vitro*. (**A**) Standard *in vitro* splicing assays were performed with HeLa cell nuclear extract in the presence of increasing amounts of wild-type (human U4SL, lanes 3–7), mutant (U4SL triple mutant, lanes 9–13) and U4atac SL (lanes 15–19) 5′ stem–loop oligonucleotides (20, 40, 80, 160 and 240 pmol, respectively). Lane 1: standard splicing assay performed at 0°C. Lanes 2 and 8: standard splicing assays without RNA oligonucleotide. (**B**) The inhibitory effect of human U4SL can be reversed by the addition of recombinant 15.5kD protein. Lane 1: standard splicing assay. Lane 2: standard splicing assay in the presence of human U4SL (225 pmol). Lanes 3–6: standard splicing assays in the presence of human U4SL (225 pmol) and increasing amounts of recombinant 15.5kD protein (100, 225, 450 and 700 pmol). The reaction mixtures were incubated at 30°C for 60 min. The RNA subsequently was recovered by phenol extraction and ethanol precipitation and separated on a denaturing 14% polyacrylamide–8 M urea gel. The positions of the pre-mRNA, splicing intermediates and product are indicated on the right. From top to bottom: lariat–exon 2, excised intron–lariat, pre-mRNA, spliced mRNA, exon 1.

sequester the 15.5kD protein at some stage during the assembly of the [U4/U6·U5] tri-snRNP into mature spliceosomes. Indeed, human U4SL effectively inhibited pre-mRNA splicing, blocking the first catalytic step of the reaction (Figure 5A, lanes 3–7). In contrast, when a mutant U4SL oligonucleotide that does not bind to the 15.5kD protein (see Figure 3B, lane 2, triple mutant A30C, U31G, G34U) was added to the *in vitro* pre-mRNA splicing reaction, only minor effects on splicing were observed at the highest concentration of oligonucleotide (Figure 5A, lanes 9–13). Similar results were obtained with other U4SL mutants that were shown not to bind the 15.5kD protein (data not shown). Interestingly, U4atac SL inhibited pre-mRNA splicing even more efficiently than U4SL (Figure 5A, compare lanes 5–7 with lanes 17–19). Since U4atac SL binds the 15.5kD protein more efficiently than U4SL (see Figure 3D), these results demonstrate that the ability of the U4 snRNA 5′ stem–loop to inhibit pre-mRNA splicing correlates well with its 15.5kD protein-binding activity. To determine whether the human U4SL

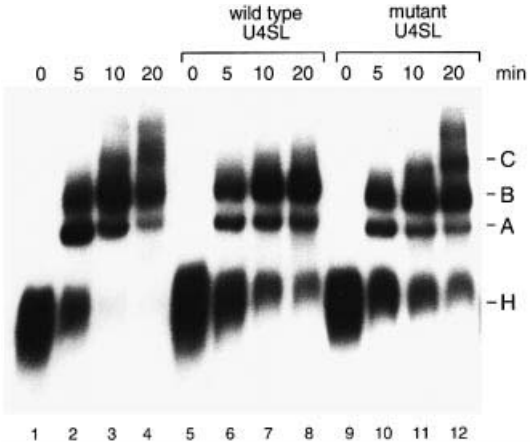

Fig. 6. Wild-type U4SL inhibits the formation of spliceosomal complex C, but not B. Standard splicing reactions were performed at 30°C for 0, 5, 10 and 20 min in the absence of exogenously added RNA oligonucleotide (lanes 1–4) or in the presence of either 250 pmol of wild-type (human U4SL, lanes 5–8) or mutant RNA oligonucleotide (U4SL triple mutant, lanes 9–12). The spliceosomal complexes were resolved by native gel electrophoresis and visualized by autoradiography.

RNA oligonucleotide inhibits pre-mRNA splicing by specifically sequestering the 15.5kD protein, we investigated the effect of wild-type U4SL in the presence of increasing amounts of recombinant 15.5kD protein. Significantly, the 15.5kD protein was able to restore pre-mRNA splicing efficiently (Figure 5B). Addition of the 15.5kD protein to a standard splicing reaction in the absence of U4SL oligonucleotide did not affect the efficiency of the splicing reaction (data not shown). Taken together, our results indicate that the 15.5kD protein plays an important role in the splicing process.

The U4 snRNA stem–loop oligonucleotide inhibits pre-mRNA splicing after spliceosomal complex B formation

To determine at which step in the spliceosomal assembly pathway the human U4SL oligonucleotide exerted its inhibitory effect, we analysed the formation of spliceosomal complexes by native gel electrophoresis. The kinetics of formation of pre-mRNA complexes containing hnRNP proteins (H complex), U1 and U2 snRNPs (A complex), U2 snRNP and the tri-snRNP (B complex) and those that have undergone the first catalytic step of splicing (C complex) is shown in Figure 6 (lanes 1–4). Interestingly, addition of wild-type U4SL, at a concentration which inhibits pre-mRNA splicing by ~90% (see Figure 5A, lane 7), leads to the inhibition of C complex formation and an accumulation of B complexes (Figure 6, lanes 7 and 8). In contrast, mutant U4SL oligonucleotides (i.e. U4SL triple mutant, Figure 6, lanes 9–12), which are incapable of binding the 15.5kD protein, did not inhibit C complex formation, nor did they lead to an accumulation of complex B. This finding is consistent with the observed failure of this U4SL mutant to inhibit pre-mRNA splicing (Figure 5A, lanes 9–13).

The observed accumulation of spliceosomal B complexes in the presence of U4SL (Figure 6) indicated that the oligonucleotide did not interfere with the early steps of [U4/U6·U5] tri-snRNP integration into spliceosomes.

In agreement with this idea, incubation of purified [U4/U6·U5] tri-snRNPs with equivalent concentrations of U4SL oligonucleotide, as used in Figure 6, did not lead to dissociation of the tri-snRNP into U4/U6 and U5 snRNPs as investigated by co-immunoprecipitation experiments with snRNP-specific antibodies (data not shown). Our results therefore suggest that the 15.5kD protein may become accessible to sequestration by the U4SL oligonucleotide only after the [U4/U6·U5] tri-snRNP has been integrated into the spliceosome, but prior to the first catalytic step of the pre-mRNA splicing reaction.

This idea is supported further by the results of immunoprecipitation experiments carried out with a rabbit antibody specific for the recombinant 15.5kD protein (see above). Anti-15.5kD antibodies efficiently precipitated native 15.5kD protein prepared by *in vitro* translation (Figure 7A, compare lanes 2 and 3). However, the antibody failed to precipitate isolated [U4/U6·U5] tri-snRNPs above the background values observed with the non-immune serum control (Figure 7B, lanes 2 and 3), suggesting that the 15.5kD protein probably interacts with other components of the tri-snRNP, rendering it inaccessible to the antibody.

We next investigated whether the 15.5kD protein becomes accessible after the [U4/U6·U5] tri-snRNP has been integrated into spliceosomes. For this purpose, we have analysed the kinetics of the pre-mRNA splicing reaction by denaturing polyacrylamide gel electrophoresis (Figure 7C, lanes 1–7). At each time point, a second aliquot of the splicing reaction was subjected to immunoprecipitation with anti-15.5kD antibodies and the precipitated radiolabelled pre-mRNA molecules were analysed (Figure 7C, lanes 8–14). Interestingly, significant co-precipitation of pre-mRNA by the 15.5kD-specific antibody was first observed after 10 min of incubation (Figure 7C, lane 11). At this time point, efficient spliceosomal B complex formation has already occurred (compare Figure 6), but not catalysis of the pre-mRNA splicing reaction (Figure 7C, lane 4). Non-immune serum, on the other hand, did not precipitate any pre-mRNA above background values at any time point (data not shown). Also, at later time points (20 min), when pre-mRNA splicing occurs, predominantly unspliced pre-mRNA was co-precipitated by the anti-15.5kD antibody (Figure 7C, lanes 5 and 12). In summary, these results suggest that after its integration into the spliceosome, but prior to the first step of splicing, the tri-snRNP complex undergoes a conformational change such that the 15.5kD protein becomes accessible to reaction with antibodies. Similarly, it is also reasonable to assume that the 15.5kD protein may become accessible to sequestration by U4SL during this stage of spliceosome assembly, thus leading to the inhibition of pre-mRNA splicing (Figures 5 and 6).

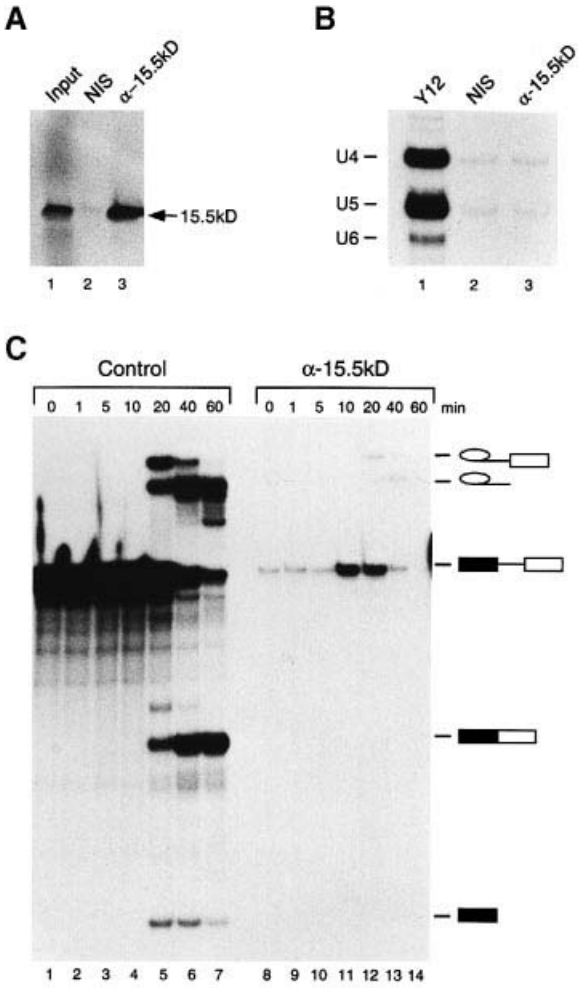

Fig. 7. Differential accessibility of the 15.5kD protein towards anti-15.5kD antibodies in purified [U4/U6·U5] tri-snRNP and spliceosomal B complexes. (**A**) Anti-15.5kD antibodies react specifically with the native 15.5kD protein prepared by translation *in vitro*. ^{35}S-labelled translates of the 15.5kD protein were precipitated with non-immune serum (lane 2) or 15.5kD-specific antiserum (lane 3). *In vitro* translated 15.5kD protein shown in lane 1 is equivalent to 10% of the amount used in the immunoprecipitation assays. Proteins were fractionated on a 13% SDS–polyacrylamide gel and visualized by fluorography. (**B**) Anti-15.5kD antibodies do not precipitate purified HeLa [U4/U6·U5] tri-snRNPs. [U4/U6·U5] tri-snRNPs purified from HeLa nuclear extracts by immunoaffinity chromatography and glycerol gradient centrifugation were used for immunoprecipitation experiments. Immunoprecipitation was performed at 150 mM NaCl with the monoclonal antibody Y12 (Lerner *et al*., 1981) to monitor the total amount of tri-snRNPs present (lane 1), with non-immune serum as a negative control (lane 2) and with antiserum specific for the 15.5kD protein (lane 3). The identity of the precipitated snRNAs is indicated on the left. Note that U6 snRNA is labelled inefficiently at the 3′ end with pCp. (**C**) Anti-15.5kD antibodies precipitate spliceosomal B complexes. Standard *in vitro* splicing assays were performed as described in Materials and methods. Reaction mixtures were incubated for the indicated times, and RNA subsequently was recovered and separated on a 14% polyacrylamide–8 M urea gel (lanes 1–7). At each time point, an aliquot of the total reaction was subjected to immunoprecipitation with anti-15.5kD antibodies, and co-precipitated RNA (lanes 8–14) was analysed in parallel with the control splicing reactions. The positions of the pre-mRNA, splicing intermediates and products are indicated on the right. From top to bottom: lariat–exon 2, excised intron–lariat, pre-mRNA, spliced mRNA, exon 1.

Discussion

The 15.5kD tri-snRNP protein binds to a conserved structural element of the U4 snRNA 5′ stem–loop

The 5′ stem–loop is one of the most conserved regions of U4 snRNA (Guthrie and Patterson, 1988) and plays an essential role in the function of U4 snRNA during pre-mRNA splicing (see Introduction). It previously had been shown that the loop region of the U4 snRNA 5′

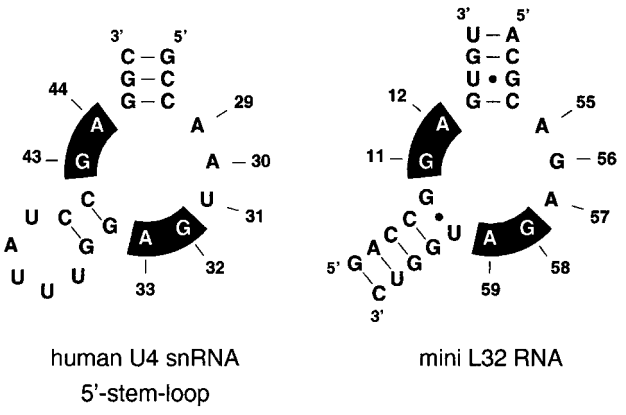

Fig. 8. RNA secondary structures of the 15.5kD and L32 protein-binding sites. The proposed secondary structure of the human U4 snRNA 5′ stem–loop is shown on the left and the structure of the L32 RNA-binding site according to Li *et al.* (1995) is shown on the right. Nucleotides conserved in both structures are boxed in black and nucleotide positions are indicated by numbers.

stem–loop is accessible to oligonucleotide-directed RNase H hydrolysis in the free U4 snRNA, but not in HeLa cell nuclear or yeast cellular extracts (Black and Steitz, 1986; Xu *et al*., 1990), suggesting that this part of U4 might be tightly associated with proteins in these extracts. Here we demonstrate that the 15.5kD protein of the human [U4/U6·U5] tri-snRNP binds directly and specifically to the U4 snRNA 5′ stem–loop. The 5′ stem–loop appears to be the sole binding site of the 15.5kD protein on the U4 snRNA, as indicated by our observations that the recombinant protein formed a binary RNA–protein complex with similar efficiency irrespective of whether the U4/U6 snRNA hybrid, free U4 snRNA or a chemically synthesized RNA oligonucleotide comprising exclusively the 5′ stem–loop sequence was used as RNA substrate (Figures 2 and 3). The results of these binding studies are consistent with those of our hydroxyl radical footprinting experiments which showed that in the U4 snRNA–15.5kD protein complex only nucleotides of the 5′ stem–loop were protected against hydroxyl radical cleavage (Figure 2). Interestingly, the free U4 snRNA can also be folded into a conserved secondary structure in which the 5′ stem–loop structure remains unchanged (Myslinski and Branlant, 1991). This raises the possibility that the interaction between the 15.5kD protein and the 5′ stem–loop of U4 snRNA must not necessarily be disrupted during the unwinding of the U4/U6 snRNA duplex which is required for activation of the spliceosome (see Introduction and below).

With the goal of identifying loop sequence elements and secondary structural features important for 15.5kD protein binding, we have carried out an extensive mutational analysis of the 5′ stem–loop. The results of these experiments, combined with phylogenetic sequence comparisons, indicate that the 15.5kD protein-binding site consists of an internal asymmetric loop which is closed by the long stem of the 5′ stem–loop and a short stem (see Figure 8). The short stem is comprised of two base pairs formed between nucleotides 34/42 and 35/41 (numbering according to the human U4 snRNA sequence). The requirement for the short stem for 15.5kD protein binding was verified experimentally by mutational analysis of the corresponding nucleotides in the U4atac snRNA stem–loop (Figure 3). In addition, the presence of the two base pairs forming the short stem is supported by the fact that compensatory base changes have occurred during evolution at the corresponding positions of the 5′ stem–loop in U4 snRNAs of diverse species (Guthrie and Patterson, 1988; see Figure 3A). Our results further indicate that the stem as such, but not the nature of the base pairs, is of prime importance for 15.5kD protein binding. Likewise, the identity of the base pairs in the long 5′ stem did not appear to be important for complex formation with the 15.5kD protein (Figure 3).

In contrast to the situation with the two stems, the 15.5kD protein-binding site is strictly dependent on the precise sequence of the internal loop, i.e. five out of seven positions, namely U31, G32 and A33 on the 5′ side, and G43 and A44 on the 3′ side of the internal loop, cannot be changed or deleted without abolishing protein-binding activity (Figure 3B). Consistent with their central role in providing a protein-binding site, these nucleotides are evolutionarily extremely conserved among U4 snRNAs of different species and are also conserved between the human U4 and U4atac snRNAs (Figure 3A, Guthrie and Patterson, 1988). The sequence requirement at positions 29 and 30 on the 5′ side of the internal loop is less stringent, but purines, as opposed to pyrimidines, are required for efficient 15.5kD protein binding (Figure 3). Deletion of one or both adenosine residues at positions 29 and 30 is deleterious for protein binding (Figure 3B and data not shown), demonstrating that the 5′ side of the internal loop requires five nucleotides for maintaining an active protein-binding site.

Due to the high conservation of individual bases in the internal loop, it is difficult to distinguish which nucleotides are important for direct protein recognition and which for maintaining an RNA structure active in 15.5kD protein binding. Similarily, it is difficult to verify possible base pairing interactions within the loop. For example, a secondary structural model has been proposed for the free, human U4 snRNA 5′ stem–loop, which would involve a non-canonical base pairing interaction between U31 and G43 (Sahasrabudhe *et al*., 1997). For this reason, we tested an RNA mutant which contained a compensatory base change at these positions (U31G and G43U). This RNA stem–loop oligonucleotide, which potentially could form an inverted G·U base pair, did not possess any protein-binding activity (data not shown). Taken together, these results strongly support the idea that the identity of the nucleotides at positions 31–33, 43 and 44 is crucial for maintaining a three-dimensional structure that is recognized specifically by the 15.5kD protein. Alternatively, it is also possible that the loop may adopt different structures in the naked as compared with the protein-bound state. Clearly, the answer to these questions will have to await the determination of the three-dimensional structure of this RNA–protein complex by X-ray crystallography and/or NMR studies.

Similarities between the 15.5kD protein and the yeast L32 ribosomal protein-binding sites on U4 snRNA and L32 pre-mRNA

The ribosomal protein L32 of *S.cerevisiae* binds to the 5′ end of the L32 transcript thereby autoregulating the

processing and translation of its own mRNA (Eng and Warner, 1991; Vilardell and Warner, 1997). The binding site for L32 consists of an asymmetric (5+2), purine-rich internal loop closed by two stems (Li *et al*., 1995; White and Li, 1996; Li and White, 1997). As the 15.5kD protein and the L32 protein share a homologous 56 amino acid central region (see Figure 1 and below), it is interesting to note that the cognate RNA-binding sites of the two proteins exhibit striking similarities (Figure 8). Significantly, several of the nucleotides in the U4 snRNA internal loop identified in this study to be essential for 15.5kD protein binding, in particular the two GA dinucleotides (boxed in black in Figure 8), are also conserved in the L32-binding site and cannot be replaced by any other nucleotide (see Li and White, 1997, and references therein). Moreover, both RNAs contain purines at positions 29/30 and 55/56, respectively (Figure 8). A major difference between the two RNA-binding sites, however, is that U31 of U4 snRNA, which was shown to be essential for efficient binding of human 15.5kD protein, is replaced by an adenosine residue (A57) in the L32 RNA (Figure 7). An additional distinguishing feature is the non-canonical G·U base pair that closes the internal loop of the L32 RNA and which is critical for strong L32 protein binding (Li and White, 1996). Despite these differences, it is tempting to suggest that the two RNA-binding sites may fold into three-dimensional conformations whose key features are similar in both RNAs. Yet, by varying a few nucleotides that are probably not important for maintaining the crucial framework of the three-dimensional fold, protein binding specificity is brought about. It will be interesting to see whether other members of the protein family sharing the 56 amino acid homology region (see Figure 1 and below) also bind to internal stem–loop stem structures which exhibit sequence similarities with the 15.5kD protein and L32 protein-binding sites.

The 15.5kD protein belongs to a family of proteins with a novel RNA recognition motif

The human 15.5kD protein is evolutionarily highly conserved and orthologues were identified in plants, insects, nematodes and yeasts. In view of the phylogenetic conservation of both the 15.5kD protein and its binding site on U4 snRNA (see Figures 1A and 3A), it is very likely that the orthologues of the human 15.5kD protein will also bind U4 RNA in their respective organisms. Consistent with this idea is the finding that the 15.5kD orthologue from *S.cerevisiae*, termed Snu13p, is present in purified yeast [U4/U6·U5] tri-snRNPs (Gottschalk *et al*., 1999; Stevens and Abelson, 1999).

The human 15.5kD protein (as well as its orthologues) is rich in hydrophobic amino acids (53 out of 128 positions), particularly in its central region where several blocks of hydrophobic amino acids are conserved (Figure 1A). Based on sequence comparisons, the 15.5kD protein does not exhibit obvious structural similarities with members of well-established families of RNA-binding proteins (see Introduction). However, it shares the homologous central region of 56 amino acids with several proteins from a variety of species which have one feature in common: they are all associated with RNP particles. The N-terminal 25 amino acids of the extended homology region (Figure 1B) were first noted by Koonin *et al*. (1994) to contain a consensus sequence of a putative RNA-binding motif. Our data establish the 15.5kD protein as a new member of this protein family that can interact directly with RNA. Moreover, by introducing point mutations into the 15.5kD protein, we show that two highly conserved amino acids of the central 56 amino acid homology region (G38 and A57) contribute to its RNA-binding activity; mutation of these residues abolishes binding (Figure 4A). In addition, mutation of residue Y80 within the homology region also significantly reduces RNA-binding activity, but does not abolish it (Figure 4A). In summary, the results discussed above suggest that the 56 amino acid region shared by the proteins listed in Figure 1B may indeed be a novel RNA recognition motif. Although the RNA-binding motif appears to be essential, it is not sufficient for the U4 snRNA-binding activity of the 15.5kD protein, i.e. a deletion mutant of the 15.5kD protein lacking 33 amino acids at its C-terminus was no longer capable of binding the U4 snRNA (Figure 4A). Moreover, the entire 15.5kD protein was strongly protected from proteolytic cleavage only when it was bound to RNA, strengthening the idea that not only the conserved RNA recognition motif, but also its flanking N- and C-terminal regions contribute to U4 snRNA-binding activity. The results of the limited proteolysis studies may also indicate that the 15.5kD protein undergoes a conformational change upon RNA binding. As unbound 15.5kD protein shows a high protease susceptibility, it is possible that the structure of the free protein might be partially or entirely disordered, but becomes stably folded when it is bound specifically to RNA, resulting in increased resistance to proteases. This kind of RNA-induced protein folding has been described for several protein–RNA complexes, including the bacteriophage λN-protein–boxB RNA complex (Zheng and Gierasch, 1997) and the Ffh–4.5S RNA interaction (Mogridge *et al*., 1998).

Functional importance of the U4 snRNA–15.5kD protein interaction for splicing

Our finding that the 15.5kD protein interacts directly with the U4 snRNA 5′ stem–loop is not only interesting from a structural point of view, but also provides a key to understanding further the function of U4 snRNA in pre-mRNA splicing. The essential role of the U4 snRNA 5′ stem–loop in pre-mRNA splicing, as previously demonstrated by mutagenesis experiments (see Introduction), can now be attributed to the pivotal role that it plays in 15.5kD protein binding. The requirement for this stem–loop for 15.5kD protein binding not only explains the inhibitory effects observed with U4 snRNA mutants lacking the entire 5′ stem–loop (Bordonné *et al*., 1990; Wersig and Bindereif, 1992), but also those obtained with certain U4 snRNA point mutants. For example, it has been shown that the 5′ stem–loop triple mutation A30C, U31G and G32U completely abolishes pre-mRNA splicing in a splicing complementation assay carried out in *Xenopus* oocytes (Vankan *et al*., 1992). Moreover, pre-mRNA splicing was also inhibited *in vivo* and *in vitro* when the yeast U4 snRNA contained a point mutation at position 45 (A45C; Hu *et al*., 1995). We demonstrate here that mutation of the corresponding nucleotides within the human U4 snRNA 5′ stem–loop abolishes binding of the 15.5kD protein (Figure 3, see U4SL A44C and U4SL

triple mutant). Curiously, a U4 snRNA 5′ stem–loop mutant where the loop sequence G45C46C47 has been replaced by U45A46A47 still supported splicing in *X.laevis* oocytes (Vankan *et al*., 1992). Consistent with the results obtained with other mutants shown in Figure 3C, a U4SL oligonucleotide containing the U45A46A47 sequence was indeed inactive in 15.5kD protein binding (data not shown). Additional studies are thus needed to clarify whether under splicing conditions the integration of the 15.5kD protein into U4 snRNPs is stabilized by protein–protein interactions with other U4 snRNP or tri-snRNP proteins when the strength of interaction between the 15.5kD protein and its cognate U4 snRNA-binding site is significantly lowered or abolished.

In this study we have investigated the functional importance of the 15.5kD–U4 snRNA 5′ stem–loop interaction in pre-mRNA splicing using an alternative approach. Addition of an RNA oligonucleotide comprising the human, wild-type U4 snRNA 5′ stem–loop (human U4SL) to HeLa cell nuclear extracts blocked the first step of pre-mRNA splicing. Moreover, our observation that the U4atac stem–loop oligonucleotide which binds the 15.5kD protein more efficiently than the U4SL oligonucleotide also inhibits pre-mRNA splicing more efficiently than U4SL (Figure 5) indicates that the U4 stem–loop oligonucleotides specifically sequester the 15.5kD protein. This idea is supported further by the finding that splicing inhibition could be rescued by the addition of purified 15.5kD protein (Figure 5). Analysis of spliceosome assembly revealed that C complex formation was inhibited specifically, while B complexes accumulated under the same conditions (Figure 6). Our results thus suggest that the U4SL oligonucleotide interferes with the function of the 15.5kD protein at a late stage of spliceosome assembly, after the [U4/U6·U5] tri-snRNP has been integrated into the spliceosome, but prior to the first catalytic step. Consistent with this interpretation, the addition of human U4SL to purified [U4/U6·U5] tri-snRNPs did not disrupt the integrity of the tri-snRNP (data not shown). It is therefore reasonable to assume that upon integration of the tri-snRNP complex into the spliceosome, conformational rearrangements occur, rendering the 15.5kD protein accessible to displacement by the U4SL oligonucleotide. This notion is corroborated strongly by our finding that the 15.5kD protein becomes accessible to anti-15.5kD antibodies after integration of the tri-snRNP complex into the spliceosome, while it remains inaccessible in the purified tri-snRNP complex (Figure 7). Similarly, mutation of A45 to C in the 5′ stem–loop of the yeast U4 snRNA, which is critical for 15.5kD protein binding in the human system (see above), also led to the accumulation of spliceosomal B complexes *in vitro*, accompanied by an inhibition of the first step of splicing (Hu *et al*., 1995).

The activation of the spliceosome for the first step of splicing requires the unwinding of the U4/U6 RNA duplex (see Introduction). One possible explanation of our results is that this unwinding step is inhibited due to the sequestering of the 15.5kD protein by the U4SL RNA oligonucleotide. However, we cannot distinguish between a direct or indirect role of the 15.5kD protein in the B to C complex transition. In particular, the observed inhibition of C complex formation may also be due to co-sequestering of additional U4 snRNP proteins by human U4SL which might be directly associated with the 15.5kD protein. Candidates for such proteins are the 20/60/90kD protein heteromer which is associated with the purified U4/U6 snRNP complex and for which no direct RNA interaction has yet been demonstrated (see Introduction). Interestingly, a mutation in one WD40 repeat of the U4 snRNP-associated yeast protein Prp4p (counterpart of the human 60kD protein, see Introduction) has been shown to lead to an accumulation of spliceosomal B complexes and to the inhibition of the U4/U6 snRNA unwinding step (Ayadi *et al*., 1997). It is thus possible that the 15.5kD protein might act as a nucleation factor for the assembly of additional U4 snRNP proteins, i.e. the 20/60/90kD heteromer. In summary, complex formation of the 15.5kD protein and additional U4 snRNP proteins with the stem–loop of U4 snRNA may be important to maintain a U4 snRNA conformation that is active in the concerted protein-mediated structural changes of the U4 and U6 snRNAs prior to splicing catalysis. It will be very interesting to define in future studies the exact composition of the spliceosomal complex B which accumulates in the presence of the human U4SL RNA oligonucleotide *in vitro*.

Future studies will also provide further insights into the protein composition of the U4atac/U6atac RNP complex. The results presented here indicate that the 15.5kD protein is probably also present in the [U4atac/U6atac·U5] tri-snRNP. If the 15.5kD protein acts as a nucleation factor for the assembly of additional U4 snRNP proteins, it is likely that the 20/60/90kD proteins of the major U4/U6 snRNP complex are also shared by the U4atac/U6atac snRNP.

Materials and methods

Isolation and cDNA cloning of the 15.5kD protein

Nuclear extracts were prepared from HeLa cells (Computer Cell Culture, Mons, Belgium) according to Dignam *et al*. (1983) and UsnRNP particles were purified by affinity chromatography using the monoclonal antibody H20 as described (Lauber *et al.*, 1996). Preparative amounts of 25S [U4/U6·U5] tri-snRNP were isolated by centrifugation of the snRNP mixture on a 10–30% glycerol gradient according to Laggerbauer *et al.* (1996). Isolation of the 15.5kD [U4/U6·U5] tri-snRNP protein for microsequencing was performed as described (Lauber *et al*., 1996). Partial amino acid sequences of tryptic peptides of the 15.5kD protein were determined by microsequencing on an ABI 477A protein sequencer (Harvard Microchemistry Facility, Cambridge, USA). The peptide sequences obtained were: (i) ADVN; and (ii) ACGVSRPVIACSVTI. An EST derived from adult human breast tissue (DDBJ/EMBL/GenBank accession No. R72749) whose ORF contained both partial peptide sequences was obtained commercially. A HeLa cell cDNA library was constructed using a cDNA synthesis kit as described by the manufacturer (Stratagene). A DNA fragment containing the coding region of EST R72749 was generated by PCR using gene-specific primers (15.5kD FOR and 15.5 REV, see below). Recombinant phages were screened on duplicate filters with the ^{32}P-labelled coding region of EST R72749 by standard methods (Sambrook *et al.*, 1989) and positive phages were purified by two additional rounds of screening. The sequence analysis of the isolated 1.5 kb HeLa cDNA insert was performed as described previously (Lauber *et al.*, 1996). The sequence of the human 15.5kD protein has been submitted to the DDBJ/EMBL/GenBank database under accession No. AF155235.

Antibody production and Western blot analysis

*Bam*HI and *Sal*I restriction sites were added by PCR to the ORF of EST R72749 using the gene-specific primers 15.5kD FOR and 15.5kD REV (see below). The fragment was then subcloned into pGEX 4T-2 (Pharmacia) and recombinant GST–15.5kD fusion protein was purified from *Escherichia coli* BL21 cells under native conditions using glutathione–Sepharose as described by the manufacturer (Pharmacia).

Antibodies against GST–15.5kD were raised by immunization of New Zealand white rabbits as described (Lauber *et al.*, 1996). For Western blot analysis, proteins were separated on a 13% SDS–polyacrylamide gel, transferred to nitrocellulose and subsequently immunostained using the ECL-Detektion Kit as described by the manufacturer (Amersham).

GST–15.5kD pull-downs

A 50 µg aliquot of either GST or GST–15.5kD fusion protein was coupled to 30 µl of pre-swollen glutathione–Sepharose beads (Pharmacia) in 500 µl of buffer A [20 mM HEPES–KOH (pH 7.9), 150 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.1% Triton X-100] for 2 h at 4°C. The beads were then washed twice with buffer A, resuspended in 500 µl of the same buffer and incubated for 2 h at 4°C with 2.5 µg of the spliceosomal snRNA mixture in the presence of 50 µg of *E.coli* tRNA (Boehringer Mannheim) and 50 µg of heparin (Sigma). The snRNAs were obtained from immunoaffinity-purified snRNP particles by phenol extraction at 0°C as described previously (Bringmann *et al.*, 1984). For the experiment in Figure 2A, lane 4, the UsnRNA mixture was denatured by incubation for 1 min at 90°C and subsequently placed on ice prior to use. After the incubation, the beads were washed three times with buffer A and resuspended in 200 µl of the same buffer. UsnRNAs were extracted with phenol and precipitated from the aqueous phase with ethanol. The precipitated snRNAs subsequently were 3′-end-labelled with [5′-^{32}P]pCp (3000 Ci/mmol; Amersham), separated on a 10% polyacrylamide–8 M urea gel and visualized by autoradiography.

In vitro transcription and translation of the 15.5kD protein

The plasmid containing the cDNA encoding the human 15.5kD protein was linearized with *Xho*I (New England Biolabs) and transcribed with T3 RNA polymerase (Promega) in a final volume of 50 µl. After phenol extraction and ethanol precipitation, 1 µg of *in vitro* transcribed mRNA was translated with wheat germ extract (Promega) in the presence of [^{35}S]methionine (1000 Ci/mmol; Amersham) in a total volume of 75 µl according to the manufacturer (Promega).

Immunoprecipitations

For the co-immunoprecipitations shown in Figures 2A (lower panel) and 4A, 250 µg of monoclonal antibody H20 (Bringmann *et al.*, 1983) was coupled to 30 µl of pre-swollen protein A–Sepharose beads (Pharmacia) in 400 µl of phosphate-buffered saline (PBS; pH 8.0) at 4°C overnight. The beads were washed three times with IPP buffer [20 mM Tris–HCl pH 8.0, 150 mM NaCl, 0.05% (v/v) NP-40] and resuspended in 300 µl of the same buffer. A 15 µl aliquot of standard *in vitro* translation reactions (see above) was incubated with 420 ng of m^{7}G-capped, *in vitro* transcribed UsnRNAs in 100 µl of buffer A for 1 h at 4°C. The *in vitro* reconstitution mixture was then incubated with the protein A–Sepharose–antibody conjugate for 1 h at 4°C. The beads were washed three times with IPP buffer, dried and heated for 5 min at 90°C in 30 µl of SDS–PAGE loading buffer. Co-immunoprecipitated protein was separated on a 13% SDS–polyacrylamide gel and subsequently visualized by autoradiography.

For the immunoprecipitation experiments shown in Figure 7, protein A–Sepharose–antibody conjugates were first prepared as described (Lauber *et at.*, 1996). A 20 µl aliquot of protein A–Sepharose and either 20 µl of anti-15.5kD antiserum, 20 µl of control non-immune serum or 5 µl of Y12 antibody was used per reaction. Immunoprecipitation was then performed by incubating the various samples is a total volume of 400 µl of PBS pH 8.0 for 2 h at 4°C with head-over-tail rotation. Subsequently, the beads were washed four times with IPP buffer and after the last washing step beads were processed as described below. For the experiment in Figure 7A, protein A–Sepharose–antibody conjugates were incubated with 10 µl of a standard *in vitro* translation reaction (see above). The immunoprecipitated, ^{35}S-labelled 15.5kD protein was released from the beads by heating for 5 min at 90°C in 30 µl of SDS–PAGE loading buffer, separated on a 13% SDS–polyacrylamide gel and visualized by autoradiography. For immunoprecipitation of tri-snRNP (Figure 7B), 5 µg of gradient-purified HeLa [U4/U6·U5] tri-snRNP was incubated with the various protein A–Sepharose–antibody conjugates. UsnRNAs were released from the beads by phenol extraction, 3′-end-labelled and analysed as above. For immunoprecipitation of spliceosomes (Figure 7C), protein A–Sepharose–antibody conjugates were added to the splicing reaction aliquots (see below). ^{32}P-Labelled RNA was released from the beads by phenol extraction.

Purification of recombinant 15.5kD protein

The concentration of recombinant GST–15.5kD protein purified from *E.coli* BL21 cells was adjusted to 2 mg/ml to prevent precipitation of the protein during proteolytic cleavage. Cleavage of the GST–15.5kD fusion protein was performed at room temperature for 12 h using 10 U of thrombin (Pharmacia) per 1 mg of protein. To separate GST and thrombin from the 15.5kD protein, the reaction mixture was applied directly onto a gel filtration column (Superdex 75, 26/60, Pharmacia) equilibrated with buffer B [20 mM HEPES–KOH pH 7.6, 120 mM NaCl, 2 mM dithiothreitol (DTT)]. The elution of the protein was monitored at 215 nm. Peak fractions were pooled and protein was concentrated by ultrafiltration (Amicon, 10kD cut-off) to ~10 mg/ml. With a calculated mol. wt of 14 174 Da for the 15.5kD protein, this stock solution had a concentration of ~0.7 mM.

Hydroxyl radical footprinting

In vitro transcribed, ^{32}P-labelled U4 snRNA (0.6 pmol) was incubated for 1 h at 4°C with increasing amounts of recombinant 15.5kD protein (0.3, 0.6, 3, 6, 33 and 134 pmol) in a final volume of 10 µl of buffer A. Ribose cleavage of the RNA backbone was initiated by hydroxyl radicals generated from H_2O_2 and free Fe(II)-EDTA complexes as previously described (Hartmuth *et al.*, 1999). The cleavage products were analysed by primer extension using a ^{32}P-labelled oligonucleotide complementary to nucleotides 65–82 of human U4 snRNA (Oligo B1, see below). The primer extension analysis was performed as described (Hartmuth *et al.*, 1999).

Electrophoretic mobility shift assay (EMSA) and competition studies

RNA oligonucleotides were obtained commercially from Eurogentech (Belgium) or from the microchemistry laboratory at the IMT (M. Krause, Marburg). All RNA oligonucleotides were gel purified before use. The RNA oligonucleotides were 5′-end-labelled using [γ-^{32}P]ATP (5000 Ci/mmol; Amersham). Recombinant 15.5kD protein was incubated at a final concentration of 14 µM with 0.5 pmol of RNA oligonucleotide for 1 h at 4°C in the presence of 10 µg of *E.coli* tRNA (Boehringer) and in a final volume of 10 µl of buffer A. RNA and RNA–protein complexes subsequently were resolved on a native 10% (80:1) polyacrylamide gel containing 0.5× TBE and visualized by autoradiography. The amount of shifted RNA was quantified by Phosphorimager analysis. For the competition studies, recombinant 15.5kD protein (10 pmol) was pre-incubated with increasing amounts of RNA oligonucleotides (5, 50, 100 and 250 pmol) for 20 min on ice before it was incubated for 45 min at 4°C with U4 snRNA (40 fmol) in a final volume of 10 µl of buffer A and in the presence of 10 µg of *E.coli* tRNA (Boehringer Mannheim). The protein–RNA complexes were analysed by EMSA as above except that a 6% polyacrylamide gel was used. The U4 snRNA used was *in vitro* transcribed in the presence of [α-^{32}P]UTP (3000 Ci/mmol; Amersham) and gel purified.

Site-directed mutagenesis

The QuickChange site-directed mutagenesis method (Stratagene) was used to introduce the point mutations G38K (GGA to AAA), A57F (GCT to TTT) and Y80A (TAC to GTT) into the 15.5kD protein, as well as to delete the C-terminal 33 amino acids (Δ33). Two oligonucleotides primers, each complementary to opposite strands of the 15.5kD protein and containing the desired mutation, were designed (oligos 1A–4B, see below) and mutagenesis was performed as described by the manufacturer. The PCR cycling parameters were as follows: one cycle (30 s at 95°C), 16 cycles (30 s at 95°C, 1 min at 55°C and 9 min at 68°C), using native *Pfu* DNA polymerase (New England Biolabs). The complete coding sequence of all mutant clones was verified by DNA sequencing.

Proteolytic digestion experiments

In vitro reconstitution of the 15.5kD–RNA oligonucleotide complex was carried out with 3 µg of purified recombinant protein and a 5-fold excess of RNA oligonucleotide (U4atac SL and U4SL triple mutant, respectively) in a final volume of 10 µl of buffer A for 1 h at 4°C. A 3 µg aliquot of either purified 15.5kD protein or *in vitro* reconstituted 15.5kD–RNA complex was incubated at 37°C with endoprotease Lys-C (sequencing grade, Roche, EC 2.4.21.50), Glu-C (sequencing grade, Roche, EC 3.4.21.19), chymotrypsin (sequencing grade, Roche, EC 3.4.21.1) or trypsin (sequencing grade, Promega, EC 3.4.21.4) with an enzyme to substrate ratio of 1:20. The reactions were stopped after 1 min, 30 min, 2 h and 16 h by adding 3 vols of SDS–PAGE sample buffer and heating for 5 min at 95°C. Aliquots of the reactions were loaded onto a 16.6% Tris–Tricine gel (Schägger and von Jagow, 1987) and protein fragments were visualized by silver staining.

In vitro splicing

Splicing was performed with ^{32}P-labelled pre-mRNA derived from the adenovirus major late transcription unit (pMINX; Zillmann *et al.*, 1988). The pre-mRNA was transcribed *in vitro* with [α-^{32}P]UTP (3000 Ci/mmol; Amersham) as described (Will *et al.*, 1996). Standard splicing assays contained 5 µl of HeLa cell nuclear extract (40% final concentration) in buffer D (Dignam *et al.*, 1983), 40 mM KCl, 3 mM $MgCl_2$, 2 mM ATP, 20 mM creatine phosphate and ^{32}P-labelled pre-mRNA (3×10^4 c.p.m., 25 fmol) in a final volume of 12.5 µl. For the experiment in Figure 5, the splicing reaction mixtures were incubated for 60 min at 30°C in the presence of increasing amounts of RNA oligonucleotides (20, 40, 80, 160 and 240 pmol). Recombinant 15.5kD protein was added to the *in vitro* splicing reaction to 8, 18, 36 or 56 µM final concentration. For the experiment in Figure 7C, a 300 µl standard splicing reaction mixture was prepared and incubated at 30°C. After 0, 1, 5, 10, 20, 40 and 60 min incubation, three 12.5 µl aliquots were withdrawn, mixed with 1.25 µl of heparin (5 mg/ml), incubated for an additional 5 min at 30°C and subsequently placed on ice. After completion of the time course, ^{32}P-labelled RNA was recovered from one aliquot of each time point by phenol extraction and ethanol precipitation. Protein A–Sepharose–antibody conjugates containing anti-15.5kD antiserum or non-immune antiserum were added to the remaining aliquots. Co-immunoprecipitation and RNA recovery were performed as described above. All RNA samples were then analysed on a 14% polyacrylamide–8 M urea gel.

Analysis of splicing complex formation

Standard splicing reactions were carried out as described above in a final volume of 50 µl with or without 250 pmol of RNA oligonucleotide (see figure legend for detail). After 0, 5, 10 and 20 min incubation at 30°C, a 10 µl aliquot of the splicing reaction was mixed with 2 µl of heparin (5 mg/ml) and 2 µl of 87% glycerol and placed on ice. Splicing complexes were separated on a 3.75% composite gel as described (Behrens *et al.*, 1993) and visualized by autoradiography.

DNA oligonucleotides used in this work

Primer 15.5kD FOR, 5′-GGGGATCCATGACTCAGGCTGATGT-GA-3′; Primer 15.5kD REV, 5′-TTGTCGACGGCAAGGGAAGC-AACTTGGC-3′. The *Bam*HI and *Sal*I sites are underlined. Oligo 1A (G38K), 5′-GGTGGCCTCATTGGCTTTTTTCCGAAGCTGCTT-3′; Oligo 1B (G38K), 5′-AAGCAGCTTCGGAAAAAAGCCAATGAGG-CCACC-3′; Oligo 2A (A57F), 5′-GAGTTCATCGTGAT- ATTTGC-AGACGCCGAGCCA-3′; Oligo 2B (A57F), 5′-TGGCTCGGCGT-CTGCAAACATCACGATGAACTC-3′; Oligo 3A (Y80A), 5′-GACAA-GAATGTGCCCGCCGTGTTTGTGCGCTCC-3′; Oligo 3B (Y80A), 5′-GGAGCGCACAAACACGGCGGGCACATTCTTGTC-3′; Oligo 4A (del33), 5′-TTACACTAAGAGCCTTCTCCCCAGGGCCTG CTT-GG-3′; Oligo 4B (del33), 5′-CCAAGCAGGCCCTGGGGAGAAGGCT-CTTAGTGTAA-3′; Oligo B1, 5′-GGTATTGGGAAAAGTTT-3′.

Acknowledgements

We thank D.Meyer and I.Öchsner for excellent technical assistance and J.Lauber, C.Schneider and N.J.Watkins for helpful advice and discussions. We are indebted to Cindy L.Will for her thoughtful comments on the manuscript. This work was supported by the Gottfried Wilhelm Leibniz Program and grants from the Deutsche Forschungsgemeinschaft (SFB 286 and SFB 397) and the Fonds der Chemischen Industrie.

References

Anthony,J.G., Weidenhammer,E.M. and Woolford,J.L.,Jr (1997) The yeast Prp3 protein is a U4/U6 snRNP protein necessary for integrity of the U4/U6 snRNP and the U4/U6.U5 tri-snRNP. *RNA*, **3**, 1143–1152.

Ayadi,L., Callebaut,I., Saguez,C., Villa,T., Mornon,J.P. and Banroques,J. (1998) Functional and structural characterization of the Prp3 binding domain of the yeast Prp4 splicing factor. *J. Mol. Biol.*, **284**, 673–687.

Banroques,J. and Abelson,J.N. (1989) PRP4: a protein of the yeast U4/U6 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.*, **9**, 3710–3719.

Behrens,S.E., Galisson,F., Legrain,P. and Lührmann,R. (1993) Evidence that the 60-kD protein of 17S U2 small nuclear ribonucleoprotein is immunologically and functionally related to the yeast PRP9 splicing factor and is required for the efficient formation of prespliceosomes. *Proc. Natl Acad. Sci. USA*, **90**, 8229–8233.

Black,D.L. and Steitz,J.A. (1986) Pre-mRNA splicing *in vitro* requires intact U4/U6 small nuclear ribonucleoprotein. *Cell*, **46**, 697–704.

Bordonné,R., Banroques,J., Abelson,J. and Guthrie,C. (1990) Domains of yeast U4 spliceosomal RNA required for PRP4 protein binding, snRNP–snRNP interactions and pre-mRNA splicing *in vivo*. *Genes Dev.*, **4**, 1185–1196.

Bringmann,P., Rinke,J., Appel,B., Reuter,R. and Lührmann,R. (1983) Purification of snRNPs U1, U2, U4, U5 and U6 with 2,2,7-trimethylguanosine-specific antibody and definition of their constituent proteins reacting with anti-Sm and anti-(U1)RNP antisera. *EMBO J.*, **2**, 1129–1135.

Bringmann,P., Appel,B., Rinke,J., Reuter,R., Theissen,H. and Lührmann,R. (1984) Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J.*, **3**, 1357–1363.

Brow,D.A. and Guthrie,C. (1988) Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature*, **334**, 213–218.

Brow,D.A. and Guthrie,C. (1989) Splicing a spliceosomal RNA. *Nature*, 337, 14–15.

Burd,C.G. and Dreyfuss,G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.

Burge,C.B., Tuschl,T. and Sharp,P.A. (1999) Splicing of precursors to mRNAs by the spliceosome. In Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds), *The RNA World.* 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 525–560.

Celander,D.W. and Cech,T.C. (1990) Iron (II)-ethylene-diamine-tetraacetic acid catalysed cleavage of RNA and DNA oligonucleotides: similar reactivity towards single- and double-stranded forms. *Biochemistry*, **29**, 1355–1361.

Cusack,S. (1999) RNA–protein complexes. *Curr. Opin. Struct. Biol.*, **9**, 53–61.

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

Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.

Eng,F.J. and Warner,J.R. (1991) Structural basis for the regulation of splicing of a yeast messenger RNA. *Cell*, **31**, 797–804.

Fabrizio,P. and Abelson,J. (1990) Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing. *Science*, **250**, 404–409.

Gottschalk,A., Neubauer,G., Banroques,J., Mann,M., Lührmann,R. and Fabrizio,P. (1999) Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6·U5] tri-snRNP. *EMBO J.*, **18**, 4535–4548.

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

Hartmuth,K., Raker,V.A., Huber,J., Branlant,C. and Lührmann,R. (1999) An unusual chemical reactivity of Sm site adenosines strongly correlates with proper assembly of core U snRNP particles. *J. Mol. Biol.*, **285**, 133–147.

Hashimoto,C. and Steitz,J.A. (1984) U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res.*, **12**, 3283–3293.

Henras,A., Henry,Y., Bousquet-Antonelli,C., Noaillac-Depeyre,J., Gélugne,J.P. and Caizergues-Ferrer,M. (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J.*, **17**, 7078–7090.

Higgins,D.G. and Sharp,P.M. (1988) Clustal: a package for performing multiple sequence alignment on a microcomputer. *Gene*, **73**, 237–244.

Horowitz,D.S., Kobayashi,R. and Krainer,A.R. (1997) A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *RNA*, **3**, 1374–1387.

Hu,J., Xu,D., Schappert,K., Xu,Y. and Friesen,J.D. (1995) Mutational analysis of *Saccharomyces cerevisiae* U4 small nuclear RNA identifies functionally important domains. *Mol. Cell. Biol.*, **15**, 1274–1285.

Hüttenhofer,A. and Noller,H.F. (1992) Hydroxyl radical cleavage of tRNA in the ribosomal P site. *Proc. Natl Acad. Sci. USA*, **89**, 7851–7855.

Incorvaia,R. and Padgett,R.A. (1998) Base pairing with U6atac snRNA is required for 5′ splice site activation of U12-dependent introns *in vivo*. *RNA*, **4**, 709–718.

Kandels-Lewis,S. and Séraphin,B. (1993) Involvement of U6 snRNA in 5′ splice site selection. *Science*, **262**, 2035–2039.

Koonin,E.V., Bork,P. and Sander,C. (1994) A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal

proteins and a ribosome modification enzyme? *Nucleic Acids Res.*, **22**, 2166–2167.
Laggerbauer,B., Lauber,J. and Lührmann,R. (1996) Identification of an RNA-dependent ATPase activity in mammalian U5 snRNPs. *Nucleic Acids Res.*, **24**, 868–875.
Lamond,A.I., Konarska,M.M., Grabowski,P.J. and Sharp,P.A. (1988) Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. *Proc. Natl Acad. Sci. USA*, **85**, 411–415.
Lauber,J., Fabrizio,P., Teigelkamp,S., Lane,W.S., Hartmann,E. and Lührmann,R. (1996) The HeLa 200kD U5 snRNP-specific protein and its homologue in *Saccharomyces cerevisiae* are members of the DEXH-box protein family of putative RNA helicases. *EMBO J.*, **15**, 4001–4015.
Lauber,J., Plessel,G., Prehn,S., Will,C.L., Fabrizio,P., Gröning,K., Lane,W.S. and Lührmann,R. (1997) The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. *RNA*, **3**, 926–941.
Lerner,M.R., Boyle,J.A., Hardin,J.A. and Steitz,J.A. (1981) Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. *Science*, **211**, 400–402.
Lesser,C.F. and Guthrie,C. (1993) Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science*, **262**, 1982–1988.
Li,H. and White,S.A. (1997) RNA apatamers for yeast ribosomal protein L32 have a conserved purine-rich internal loop. *RNA*, **3**, 245–254.
Li,H., Dalal,S., Kohler,J., Vilardell,J. and White,S.A. (1995) Characterization of the pre-mRNA binding site for yeast ribosomal protein L32: the importance of a purine-rich internal loop. *J. Mol. Biol.*, **250**, 447–459.
Madhani,H.D. and Guthrie,C. (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell*, **71**, 803–817.
Madhani,H.D. and Guthrie,C. (1994) Dynamic RNA–RNA interactions in the spliceosome. *Annu. Rev. Genet.*, **28**, 1–26.
Michel,F. and Costa,M. (1998) Inferring RNA structure by phylogentic and genetic analyses. In Simons,R.W. and Grunberg-Manago,M. (eds), *RNA Structure and Function*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 175–202.
Mogridge,J., Legault,P., Li,J., Van Oene,M.D., Kay,L.E. and Greenblatt,J. (1998) Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage λN protein. *Mol. Cell*, **1**, 265–275.
Moore,M.J., Query,C.C. and Sharp,P.A. (1993) Splicing of precursors to mRNA by the spliceosome. In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.
Myslinski,E. and Branlant,C. (1991) A phylogenetic study of U4 snRNA reveals the existence of an evolutionarily conserved secondary structure corresponding to 'free' U4 snRNA. *Biochimie*, **73**, 17–28.
Nilsen,T.W. (1998) RNA–RNA interactions in nuclear pre-mRNA splicing. In Simons,R.W. and Grunberg-Manago,M. (eds), *RNA Structure and Function*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 279–307.
Peterson-Bjørn,S.P., Soltyk,A., Beggs,J.D. and Friesen,J.D. (1989) PRP4 (RNA4) from *Saccharomyces cerevisiae*: its gene product is associated with the U4/U6 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.*, **9**, 3698–3709.
Rinke,J., Appel,B., Digweed,M. and Lührmann,R. (1985) Localization of a base-paired interaction between small nuclear RNAs U4 and U6 in intact U4/U6 ribonucleoprotein particles by psoralen cross-linking. *J. Mol. Biol.*, **185**, 721–731.
Sahasrabudhe,P.V., Sun,J. and Gmeiner,W.H. (1997) The 5′ stem–loop from human U4 snRNA adopts a novel conformation stabilized by G–C and G–U base pairs. *J. Biomol. Struct. Dyn.*, **14**, 567–577.
Saito,H., Fujiwara,T., Shin,S., Okui,K. and Nakamura,Y. (1996) Cloning and mapping of a human novel cDNA (NHP2L1) that encodes a protein highly homologous to yeast nuclear protein NHP2. *Cytogenet. Cell. Genet.*, **72**, 191–193.
Sambrook,J., Fritsch,E. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sawa,H. and Abelson,J. (1992) Evidence for a base-pairing interaction between U6 small nuclear RNA and 5′ splice site during the splicing reaction in yeast. *Proc. Natl Acad. Sci. USA*, **89**, 11269–11273.
Schägger,H. and von Jagow,G. (1987) Tricine–sodium dodecyl sulfate–polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100kD. *Anal. Biochem.*, **166**, 368–379.
Sontheimer,E.J. and Steitz,J.A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262**, 1989–1996.
Staley,J.P. and Guthrie,C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs and things. *Cell*, **92**, 315–326.
Stevens,S.W. and Abelson,J. (1999) Purification of the yeast U4/U6·U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc. Natl Acad. Sci. USA*, **96**, 7226–7231.
Sun,J.S. and Manley,J.L. (1995) A novel U2–U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev*, **9**, 843–854.
Tarn,W.Y. and Steitz,J.A. (1996) Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT–AC introns. *Science*, **273**, 1824–1832.
Teigelkamp,S., Achsel,T., Mundt,C., Gothel,S.F., Cronshagen,U., Lane,W.S., Marahiel,M. and Lührmann,R. (1998) The 20kD protein of human [U4/U6·U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *RNA*, **4**, 127–141.
Vankan,P., McGuigan,C. and Mattaj,I.W. (1992) Roles of U4 and U6 snRNAs in the assembly of splicing complexes. *EMBO J.*, **11**, 335–343.
Vilardell,J. and Warner,J.R. (1997) Ribosomal protein L32 of *Saccharomyces cerevisiae* influences both the splicing of its own transcript and the processing of rRNA. *Mol. Cell. Biol.*, **17**, 1959–1965.
Watkins,N.J., Gottschalk,A., Neubauer,G., Kastner,B., Fabrizio,P., Mann,M. and Lührmann,R. (1998) Cbf5p, a potential pseudouridine synthase and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA*, **4**, 1549–1568.
Wersig,C. and Bindereif,A. (1992) Reconstitution of functional mammalian U4 small nuclear ribonucleoprotein: Sm protein binding is not essential for splicing *in vitro*. *Mol. Cell. Biol.*, **12**, 1460–1468.
White,S.A. and Li,H. (1996) Yeast ribosomal protein L32 recognises an RNA G:U juxtaposition. *RNA*, **2**, 226–234.
Will,C.L. and Lührmann,R. (1997) snRNP structure and function. In Krainer,A.R. (ed.), *Eukaryotic mRNA Processing*. IRL Press at Oxford University Press, Oxford, UK, pp. 130–173.
Will,C.L., Rümpler,S., Klein Gunnewiek,J., van Venrooij,W.J. and Lührmann R. (1996) *In vitro* reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res.*, **24**, 4614–4623.
Wu,J.A. and Manley,J.L. (1991) Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA. *Nature*, **352**, 818–821.
Xu,Y., Petersen-Bjørn,S. and Friesen,J.D. (1990) The PRP4 (RNA4) protein of *Saccharomyces cerevisiae* is associated with the 5′ portion of the U4 small nuclear RNA. *Mol. Cell. Biol.*, **10**, 1217–1225.
Yean,S.L. and Lin,R.J. (1991) U4 small nuclear RNA dissociates from a yeast spliceosome and does not participate in the subsequent splicing reaction. *Mol. Cell. Biol.*, **11**, 5571–5577.
Yu,Y.T. and Steitz,J.A. (1997) Site-specific crosslinking of mammalian U11 and U6atac to the 5′ splice site of an AT–AC intron. *Proc. Natl Acad. Sci. USA*, **94**, 6030–6035.
Zheng,N. and Gierasch,L.M. (1997) Domain interactions in *E.coli* SRP: stabilization of M domain by RNA is required for effective signal sequence modulation of NG domain. *Mol. Cell*, **1**, 79–87.
Zillmann,M., Zapp,M.L. and Berget,S.M. (1988) Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Mol. Cell. Biol.*, **8**, 814–821.

Received June 10, 1999; revised and accepted September 7, 1999