RNA (1997), 3:926-941. Cambridge University Press. Printed in the USA.

The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p

JÜRGEN LAUBER,[1,4] GABRIELE PLESSEL,[1] SIEGFRIED PREHN,[2] CINDY L. WILL,[1] PATRIZIA FABRIZIO,[1] KARSTEN GRÖNING,[1] WILLIAM S. LANE,[3] and REINHARD LÜHRMANN[1]

[1]Institut für Molekularbiologie und Tumorforschung, Emil-Mannkopff-Straße 2, 35037 Marburg, Germany
[2]Institute for Biochemistry, Humboldt-University Berlin, Hessische Straße 3-4, 10115 Berlin, Germany
[3]Harvard Microchemistry Facility, 16 Divinity Avenue, Cambridge Massachusetts 02138, USA

ABSTRACT

Immunoaffinity-purified human 25S [U4/U6.U5] tri-snRNPs harbor a set of polypeptides, termed the tri-snRNP proteins, that are not present in Mono Q-purified 20S U5 snRNPs or 10S U4/U6 snRNPs and that are important for tri-snRNP complex formation (Behrens SE, Lührmann R, 1991, *Genes & Dev 5*:1439–1452). Biochemical and immunological characterization of HeLa [U4/U6.U5] tri-snRNPs led to the identification of two novel proteins with molecular weights of 61 and 63kD that are distinct from the previously described 15.5, 20, 27, 60, and 90kD tri-snRNP proteins. For the initial characterization of tri-snRNP proteins that interact directly with U4/U6 snRNPs, immunoaffinity chromatography with an antibody directed against the 60kD protein was performed. We demonstrate that the 60 and 90kD tri-snRNP proteins specifically associate with the U4/U6 snRNP at salt concentrations where the tri-snRNP complex has dissociated. The primary structures of the 60kD and 90kD proteins were determined by cloning and sequencing their respective cDNAs. The U4/U6-60kD protein possesses a C-terminal WD domain that contains seven WD repeats and thus belongs to the WD-protein family, whose best-characterized members include the Gβ subunits of heterotrimeric G proteins. A database homology search revealed a significant degree of overall homology (57.8% similarity, 33.9% identity) between the human 60kD protein and the *Saccharomyces cerevisiae* U4/U6 snRNP protein Prp4p. Two additional, previously undetected WD repeats (with seven in total) were also identified in Prp4p, consistent with the possibility that 60kD/Prp4p, like β-transducin, may adopt a propeller-like structure. The U4/U6-90kD protein was shown to exhibit significant homology, particularly in its C-terminal half, with the *S. cerevisiae* splicing factor Prp3p, which also associates with the yeast U4/U6 snRNP. Interestingly, U4/U6-90kD shares short regions of homology with *E. coli* RNase III, including a region encompassing its double-stranded RNA binding domain. Based on their structural similarity with essential splicing factors in yeast, the human U4/U6-60kD and 90kD proteins are likely also to play important roles in the mammalian splicing process.

Keywords: pre-mRNA splicing; snRNP protein; WD repeat

INTRODUCTION

Pre-mRNA splicing occurs via a two-step mechanism involving first 5′ splice site cleavage and covalent linkage of the first nucleotide of the intron to an adenosine residue (i.e., the branch point) upstream of the 3′ splice site, followed by 3′ splice site cleavage and exon ligation (reviewed by Moore et al., 1993; Krämer, 1995). The splicing reaction is catalyzed by the spliceosome, which is formed by the ordered interaction of the U1, U2, U4/U6, and U5 snRNPs, as well as an undefined number of non-snRNP splicing factors, with the pre-mRNA. The spliceosomal snRNPs contain either one (U1, U2, U5) or two (U4/U6) snRNAs, complexed with proteins that are either common to all snRNP species or specific for a given particle (for review, see Will & Lührmann, 1997a). The U1 and U2 snRNPs are required for the formation of the earliest spliceosomal complexes E and A, respectively, in which, among others, functional pairing of the 5′ and 3′ splice sites occurs. Mature spliceosomes are formed by the sub-

Reprint requests to: Reinhard Lührmann, Institut für Molekularbiologie und Tumorforschung, Emil-Mannkopff-Str. 2, D-35037 Marburg, Germany; e-mail: luehrmann@imt.uni-marburg.de.
[4]Present address: Qiagen GmbH, Max-Volmer-Strasse 4, 40724 Hilden, Germany.

sequent association of the U4/U6 and U5 snRNPs, in the form of a 25S [U4/U6.U5] tri-snRNP complex (reviewed by Moore et al., 1993).

During splicing, a complex RNA network consisting of several snRNA–pre-mRNA and snRNA–snRNA interactions is formed (reviewed by Madhani & Guthrie, 1994; Nilsen, 1994). For example, during the early stages of splicing complex formation, the U1 and U2 snRNPs recognize and base pair with the 5′ splice site and the branch site, respectively. In the subsequent steps of spliceosome assembly, a number of RNA conformational changes, involving predominantly the snRNAs of the [U4/U6.U5] tri-snRNP, are known to occur. Within the tri-snRNP complex, the U4 and U6 snRNAs base pair to form a Y-shaped structure with two intermolecular helices (Bringmann et al., 1984; Hashimoto & Steitz, 1984; Rinke et al., 1985; Brow & Guthrie, 1988). After the integration of this complex into the spliceosome, the U4/U6 helices are completely unwound and the U6 snRNA interacts with the 5′ end of the U2 snRNA (Datta & Weiner, 1991; Wu & Manley, 1991; Madhani & Guthrie, 1992; Sun & Manley, 1995), as well as intron sequences at the 5′ splice site (Sawa & Abelson, 1992; Sawa & Shimura, 1992; Wassarman & Steitz, 1992; Kandels-Lewis & Séraphin, 1993; Lesser & Guthrie, 1993; Sontheimer & Steitz, 1993). The U4 snRNP, on the other hand, only loosely interacts or even dissociates from the spliceosome (Lamond et al., 1988; Yean & Lin, 1991). In addition to novel U6 snRNA interactions, the U5 snRNA interacts with exon nucleotides at the 5′ splice site in the pre-catalytic spliceosome and, after the first catalytic step of splicing, additionally with exon nucleotides at the 3′ splice site (Newman & Norman, 1991, 1992; Wyatt et al., 1992; Cortes et al., 1993; Sontheimer & Steitz, 1993). It is thought that the U4/U6 interaction domain re-forms, and that the [U4/U6.U5] tri-snRNP complex reassembles after each round of splicing.

The proteins involved in mediating these various RNA conformational changes have not been identified fully yet. A number of spliceosomal proteins have been identified as putative RNA helicases and, accordingly, have been proposed to mediate structural rearrangements within the spliceosome (reviewed by Guthrie, 1991; Beggs, 1995; Will & Lührmann, 1997b). In yeast, for example, the non-snRNP splicing factor Prp28p, which is a member of the DEAD-box family of putative RNA helicases, has been proposed to destabilize the U4/U6 helix (Strauss & Guthrie, 1991). Because the U4 and U6 snRNAs undergo dramatic conformational changes, it is also likely that intrinsic proteins of the tri-snRNP complex, in particular those interacting with U4 and U6, play a major role in facilitating both the formation and unwinding of the U4/U6 helices. Interestingly, at least one component of the [U4/U6.U5] tri-snRNP complex, the U5-200kD protein, has been identified as a putative RNA helicase both in humans and yeast (Lauber at al., 1996; Lin & Rossi, 1996; Noble & Guthrie, 1996; Xu et al., 1996). However, the precise target of this putative helicase is not clear presently.

The biochemical composition of the tri-snRNP complex has been investigated intensively in HeLa cells. In *Saccharomyces cerevisiae*, the initial biochemical characterization of the yeast [U4/U6.U5] tri-snRNP indicates that the complexity of its protein composition is similar to that of the human complex (Fabrizio et al., 1994). Purified HeLa tri-snRNPs contain two sets of the Sm proteins B′, B, D3, D2, D1, E, F, and G, which are known to associate with every species of m_3G-capped snRNA, and at least 14 additional proteins (Behrens & Lührmann, 1991). Nine of these, with molecular weights of 15, 40, 52, 100, 102, 110, 116, 200, and 220kD, are operationally defined as U5-specific proteins because they can be co-isolated with 20S U5 snRNPs under stringent conditions such as ion exchange chromatography (Bach et al., 1989). U4/U6 snRNPs isolated by Mono Q chromatography, on the other hand, contain only Sm proteins (Bach et al., 1990). However, the inability to detect U4/U6-specific proteins in purified particles may simply be a consequence of the harsh isolation conditions employed. The tri-snRNP complex, for example, is salt sensitive and it, along with several tri-snRNP-specific proteins, dissociates at ionic strengths above 300 mM. Proteins that are found only in tri-snRNP complexes, but not in purified U5 or U4/U6 snRNPs, namely those with molecular weights of 15.5, 20, 27, 60, and 90kD, have been defined as tri-snRNP-specific (Behrens & Lührmann, 1991), although, under physiological conditions, several of these proteins may actually bind directly to the U4 and U6 snRNAs. Indeed, evidence indicating that two spliceosomal-associated proteins (SAPs), with molecular weights of 60 and 90kD, are associated with the U4/U6 snRNP has been provided by immunoprecipitation studies with an autoimmune patient serum (Gozani et al., 1994).

In *S. cerevisiae*, two U4/U6-specific proteins, namely Prp4p and Prp3p, have been identified genetically (Banroques & Abelson, 1989; Petersen-Bjørn et al., 1989; J.G. Anthony & J.L. Woolford, pers. comm.). Prp4p has been shown to be required for the association of the U4/U6 snRNP with the U5 snRNP and to interact within the U4/U6 particle, either directly or indirectly, with the conserved 5′ stem-loop of the U4 snRNA (Banroques & Abelson, 1989; Petersen-Bjorn et al., 1989; Bordonné et al., 1990; Xu et al., 1990). Interestingly, Prp4p contains several WD repeats, initially found in the Gβ subunit of the heterotrimeric G protein transducin, which have been shown to be essential for its function during splicing (Dalrymple et al., 1989; Hu et al., 1994; Ayadi et al., 1997). Prp3p was originally proposed to be U4/U6-specific based on the observation that overexpression of its gene can compensate for a defective *PRP4* gene (Last et al., 1987). Immunopre-

cipitation studies have subsequently confirmed that it associates with the U4/U6 snRNP (J.G. Anthony & J.L. Woolford, pers. comm.).

In an effort to identify proteins interacting with the HeLa U4/U6 snRNP, immunoaffinity chromatography of total snRNPs was performed at various salt concentrations with antibodies directed against the [U4/U6.U5] tri-snRNP 60kD protein. At salt concentrations higher than 350 mM, U4/U6 snRNPs containing the common Sm proteins and the 60 and 90kD protein could be isolated by elution with a competing 60kD protein peptide, thereby demonstrating that the 60 and 90kD proteins interact preferentially with the U4/U6 particle. Based on partial peptide sequence data derived by microsequencing of gel-purified tri-snRNP proteins, cDNAs coding for the 60 and 90kD proteins were isolated. Sequence alignments revealed that the 60kD protein is the mammalian homologue of *S. cerevisiae* Prp4p, whereas the the 90kD protein shares homology with Prp3p. Both Prp4p and the 60kD protein appear to contain seven WD repeats, and thus potentially form a seven-bladed propeller structure characteristic of the Gβ subunits of heteromeric G proteins (Wall et al., 1995; Lambright et al., 1996; Sondek et al., 1996).

RESULTS

Protein composition of gradient-purified 20S U5 snRNPs and 25S [U4/U6.U5] tri-snRNPs

As a first step toward the identification of proteins potentially interacting with U4/U6 snRNPs, the protein composition of 20S U5 snRNPs and 25S [U4/U6.U5] tri-snRNPs was analyzed in more detail. To this end, U1, U2, U4/U6, and U5 snRNPs were isolated from HeLa nuclear extract by anti-m_3G immunoaffinity chromatography under low-salt conditions (250 mM NaCl). 20S U5 snRNPs and 25S tri-snRNPs were purified from this snRNP mixture by glycerol gradient centrifugation and their protein compositions analyzed by SDS-PAGE (Fig. 1). As reported previously (Bach et al., 1989), in addition to the common Sm proteins (B′ to G), proteins with apparent molecular weights of 220, 200, 116, 110, 102, 100, 52, 40, and 15kD were found associated with 20S U5 snRNPs. Surprisingly, an additional protein of approximately 65kD is reproducibly observed in glycerol gradient-purified U5 snRNP preparations. Because Mono Q-purified U5 particles generally lack this protein, it likely represents a loosely associated component of the U5 snRNP. Apart from the aforementioned U5 snRNP proteins, seven additional proteins with apparent molecular weights of 90, 63, 61, 60, 27, 20, and 15.5kD were detected as components of gradient-purified 25S [U4/U6.U5] tri-snRNP complexes. In contrast to previous studies, where it was not possible to distinguish between one or more proteins in the 60kD range (Behrens & Lührmann, 1991), at least three tri-snRNP proteins in this size range could clearly be detected.

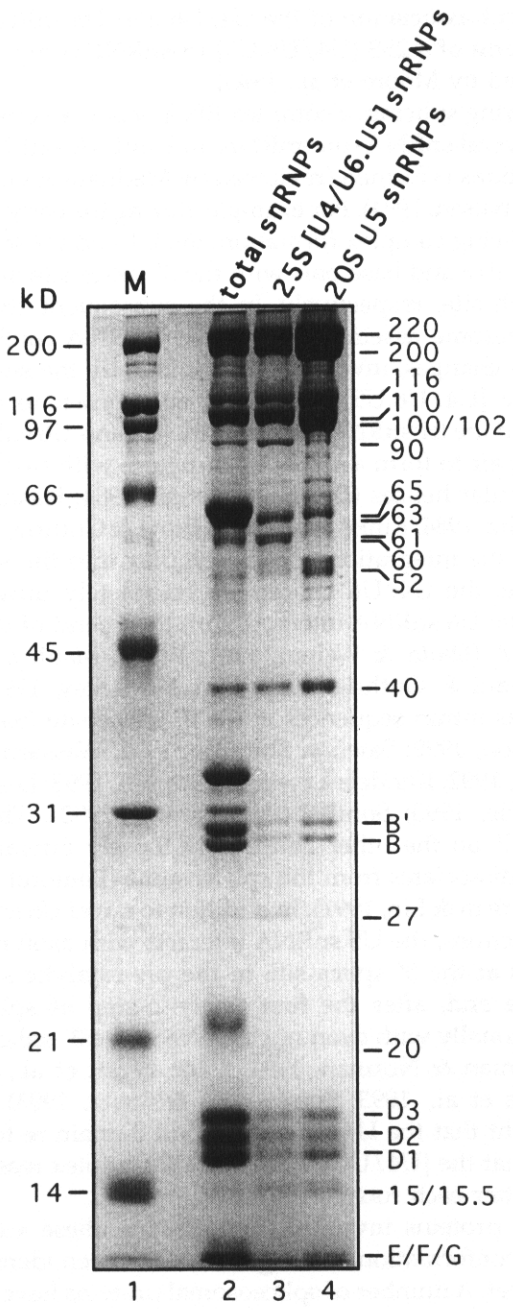

FIGURE 1. Protein composition of the 20S U5 snRNP and 25S [U4/U6.U5] tri-snRNP complex from HeLa cells. Proteins isolated from 20 μg of gradient-purified 20S U5 snRNPs or 25S (U4/U6.U5) tri-snRNPs were separated on a 12% SDS-polyacrylamide gel and visualized by staining with Coomassie blue. Lane 1, molecular weight markers; lane 2, total spliceosomal snRNP proteins; lane 3, 25S [U4/U6.U5] tri-snRNP proteins; lane 4, 20S U5 snRNP proteins. The molecular masses (in kD) of protein standards (lane 1) are indicated on the left and proteins present in the [U4/U6.U5] tri-snRNP complex, on the right.

Characterization of antibodies raised against the 60, 61, 63, and 90kD tri-snRNP proteins

For the initial characterization of a subset of the tri-snRNP proteins, the 60, 61, 63, and 90kD proteins were gel purified and partial peptide sequences were obtained by microsequencing. Polyclonal antibodies were raised against synthetic peptides of these proteins and their specificity tested by immunoblotting. As shown in Figure 2A, anti-60kD and anti-90kD sera specifically recognized the 60 and 90kD proteins, respectively, on immunoblots of gradient-purified HeLa tri-snRNP proteins (compare lanes 2 and 5 with lane 1); the corresponding non-immune sera (lanes 6 and 9) did not react with any of the tri-snRNP proteins. Anti-61kD and anti-63kD sera recognized the 61 and 63kD proteins, respectively (Fig. 2A, compare lanes 3 and 4 with lane 1), but cross-reacted with the 40kD protein (in the case of anti-61kD) or one or more proteins in the 100kD range (in the case of anti-63kD); both cross-reactions appeared to be due to antibodies present initially in the corresponding non-immune serum (lanes 7 and 8). The lack of cross-reactivity of these anti-sera with the 60, 61, and 63kD proteins provided further support for the designation of these proteins as unique polypeptides. Additionally, the fact that the 65kD protein was not recognized by any of these anti-sera when immunoblotting was performed with gradient-purified 20S U5 snRNPs containing significant amounts of the 65kD protein (Fig. 2B) suggested that it might be distinct from the three tri-snRNP proteins of similar molecular weight and could, thus, represent a novel U5 snRNP protein. Alternatively, the 65kD protein could represent a post-translationally modified version of the 60, 61, or 63kD protein which does not react with the corresponding anti-serum. The latter experiment also confirmed that the 60, 61, 63, and 90kD proteins are largely absent from purified 20S U5 snRNPs (Fig. 2B). To what extent the 65kD protein is also present in 25S [U4/U6.U5] tri-snRNPs is currently unclear. Isolation of partial cDNA clones coding for the 60, 61, and 63kD proteins provided independent confirmation that they are unique polypeptides (see below; data not shown).

The ability of the various anti-peptide antibodies to specifically recognize native proteins present in the isolated [U4/U6.U5] tri-snRNP complex was assayed subsequently by immunoprecipitation. For this purpose, antibodies were bound to Protein A–Sepharose and incubated with a mixture of immunoaffinity-purified snRNPs (U1, U2, U4/U6, and U5). The ability of each antibody to react with a given snRNP was determined by assaying for co-precipitation of the corresponding snRNA. Anti-90kD, anti-63kD, and anti-61kD antibodies failed to precipitate any snRNAs (data not shown), suggesting that they recognize only the denatured form of these proteins and/or that their cognate epitope is not accessible in intact tri-snRNPs. In contrast, anti-60kD antibodies recognized [U4/U6.U5] tri-snRNPs, as evidenced by the coprecipitation of predominantly U4, U5, and U6 snRNAs (data not shown, see below). Precipitation of the latter could be competed for by the addition of an excess of the

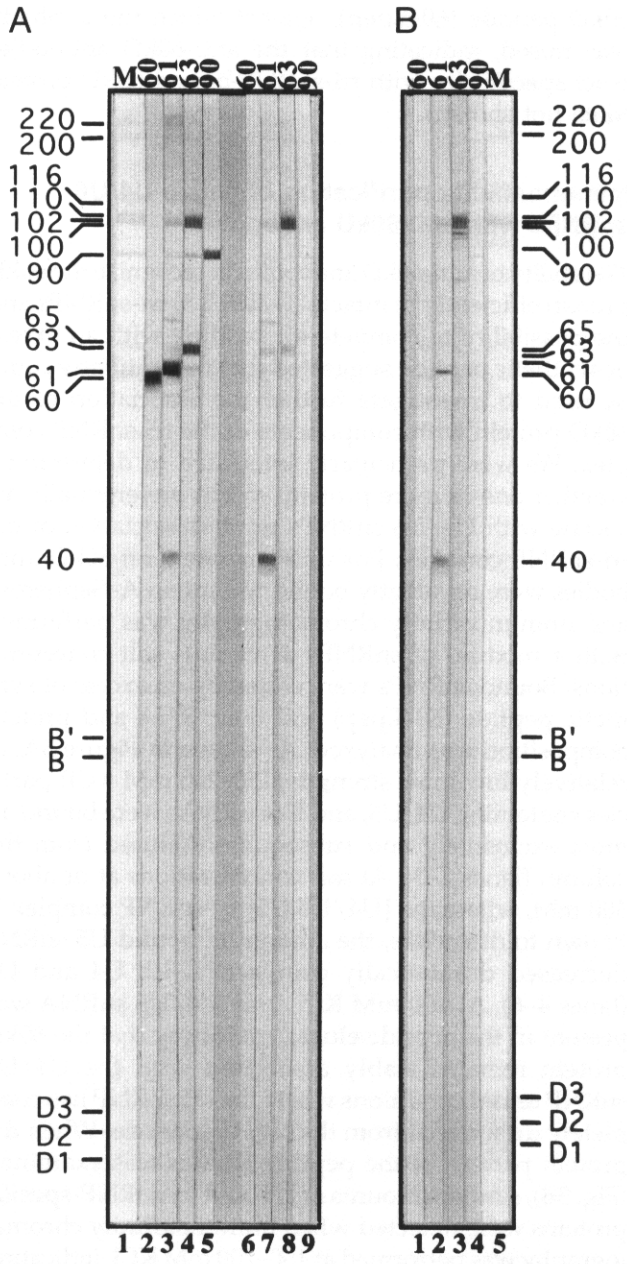

FIGURE 2. Characterization by western blotting of antibodies directed against the high molecular weight tri-snRNP-specific proteins. SnRNP proteins isolated from gradient-purified 25S [U4/U6.U5] tri-snRNPs (**A**) or 20S U5 snRNPs (**B**) were blotted onto nitrocellulose and immunostained as described in Materials and Methods. Lane 1 in panel A: tri-snRNP proteins stained with Ponceau S. Lane 5 in panel B: U5 snRNP proteins stained with Ponceau S. Lanes 2–5 in panel A and 1–4 in B were immunostained with the following antibodies: anti-60kD diluted 1:10,000, anti-61kD diluted 1:200, anti-63kD diluted 1:200, and anti-90kD diluted 1:1,000. Non-immune sera (lanes 6–9 in panel A) were diluted to the same extent as their respective immune sera. The positions of proteins present in the [U4/U6.U5] tri-snRNP complex are labeled on the left or right of each panel.

60kD peptide (60-4.pep), against which the antibody was raised, indicating that the anti-60kD antibodies react specifically with tri-snRNP-bound 60kD protein (data not shown).

Immunoaffinity purification of native U4/U6 snRNPs with anti-60kD antibodies

The ability of anti-60kD antibodies to recognize the 60kD protein efficiently in intact [U4/U6.U5] tri-snRNPs and the possibility to compete for binding with an excess of synthetic peptide suggested that this antibody could be used to investigate further the association of the 60kD protein with components of the tri-snRNP complex. We were particularly interested in determining whether one or more proteins might preferentially associate with U4/U6 snRNPs upon dissociation of the tri-snRNP complex. For this purpose, anti-60kD antibodies were covalently bound to Protein A–Sepharose and immunoaffinity chromatography was performed with a mixture of snRNPs at various salt concentrations. Bound snRNPs were eluted by an excess of synthetic peptide (60-4.pep) and their RNA and protein composition was analyzed. As shown in Figure 3A, at relatively low ionic strength (150–200 mM KCl) particles containing U4, U5, and U6 snRNAs were bound almost exclusively and subsequently eluted from the column (lanes 2–3). At salt concentrations at or above 300 mM, where the [U4/U6.U5] tri-snRNP complex is known to dissociate, the amount of bound U5 snRNP decreased dramatically compared with U4 and U6 (lanes 4–6). At 400 mM KCl, only U4/U6 snRNA was present in the peptide eluate, indicating that the 60kD protein remains stably associated with the U4/U6 snRNP under conditions where the U5 snRNP has completely dissociated from the U4/U6 particle. When the protein pattern of the peptide eluates was examined (Fig. 3B), similar amounts of U5- and tri-snRNP-specific proteins were detected when immunoaffinity chromatography was performed at 150–200 mM KCl, indicating that predominantly [U4/U6.U5] tri-snRNP complexes were bound and eluted. Because of the high amount of synthetic peptide used for elution, the presence of low molecular weight proteins was obscured. In addition, a number of proteins are poorly resolved on this particular gel, including the 63 and 65kD proteins, as well as those in the 100 and 200kD range. Consistent with the reduced amount of U5 snRNA eluted, a corresponding reduction in U5 snRNP-specific proteins in the eluates was observed when snRNPs were applied to the column in buffer containing 300–400 mM KCl (Fig. 3B, lanes 3–5). A similar reduction in the amount of the 61kD and, presumably, the 63kD tri-snRNP protein was also observed as the salt concentration was increased, suggesting that they also dissociate from the U4/U6 snRNP at ionic strengths above 300 mM KCl. In contrast, significant amounts of the 60 and 90kD protein were still detected when immunoaffinity chromatography was performed at 400 mM KCl (Fig. 3B, lane 5). Because relatively pure U4 and U6 snRNAs were also eluted under these conditions (Fig. 3A, lane 6), the 60 and 90kD proteins appear to associate preferentially with these snRNAs and thus should be referred to as U4/U6 snRNP-specific proteins.

Cloning of the U4/U6 snRNP-specific 60kD protein

As a prerequisite for the detailed structural and functional analysis of the 60 and 90kD U4/U6 snRNP proteins, we set out to clone and sequence their respective cDNAs. A database search for corresponding open reading frames was thus performed using the partial amino acid sequences obtained by microsequencing of purified 60 and 90kD protein. Because, at that time, no sequences matching those obtained from the 60kD protein were found, a HeLa cDNA library was screened with degenerate oligonucleotide primers designed from the sequence of the 60kD peptide (60-3.pep). This led to the isolation of a 1.87-kb cDNA clone that contained a truncated open reading frame of 241 amino acids. To obtain a cDNA encoding the entire 60kD protein, PCR amplification of reverse-transcribed HeLa mRNA with gene-specific primers was performed, resulting in the isolation of 2.8-kb full-length cDNA clone. The open reading frame of this cDNA codes for a protein with a predicted molecular mass of 58.5kD and a calculated isoelectric point of 6.99, which is in good agreement with both the apparent size and experimentally determined isoelectric point (pI 6.7) of the 60kD protein (Fig. 4 and data not shown). All of the peptide sequences obtained from microsequencing of the purified 60kD protein are present in the deduced amino acid sequence; only two amino acids in 60-1.pep and one in 60-4.pep are different (Fig. 4). Consistent with it being the bona fide 60kD cDNA, peptide sequences obtained from the purified protein are located at both the C- and N-terminus of the cDNA-encoded protein. An AUG codon (nt 94–96), whose surrounding sequence conforms to the consensus sequence for eukaryotic translation initiation (Kozak, 1987), is located 93 nucleotides from the 5′ end of the cDNA clone. An inframe stop codon (nt 43–45) is found upstream of this initiator codon, confirming that it codes for the first methionine. A polyadenylation signal (AUUAAA, nt 2749–2754) is located 1,091 nt downstream of the stop codon (nt 1658–1660) and 23 nt upstream of the poly A tail.

The N-terminal part of the 60kD protein does not contain any known motifs that might provide insight into its probable function. In contrast, at least six WD repeats are found in the C-terminal portion of the protein. The WD repeat (also referred to as the WD-40 or β-transducin repeat) was first observed in the β-subunit

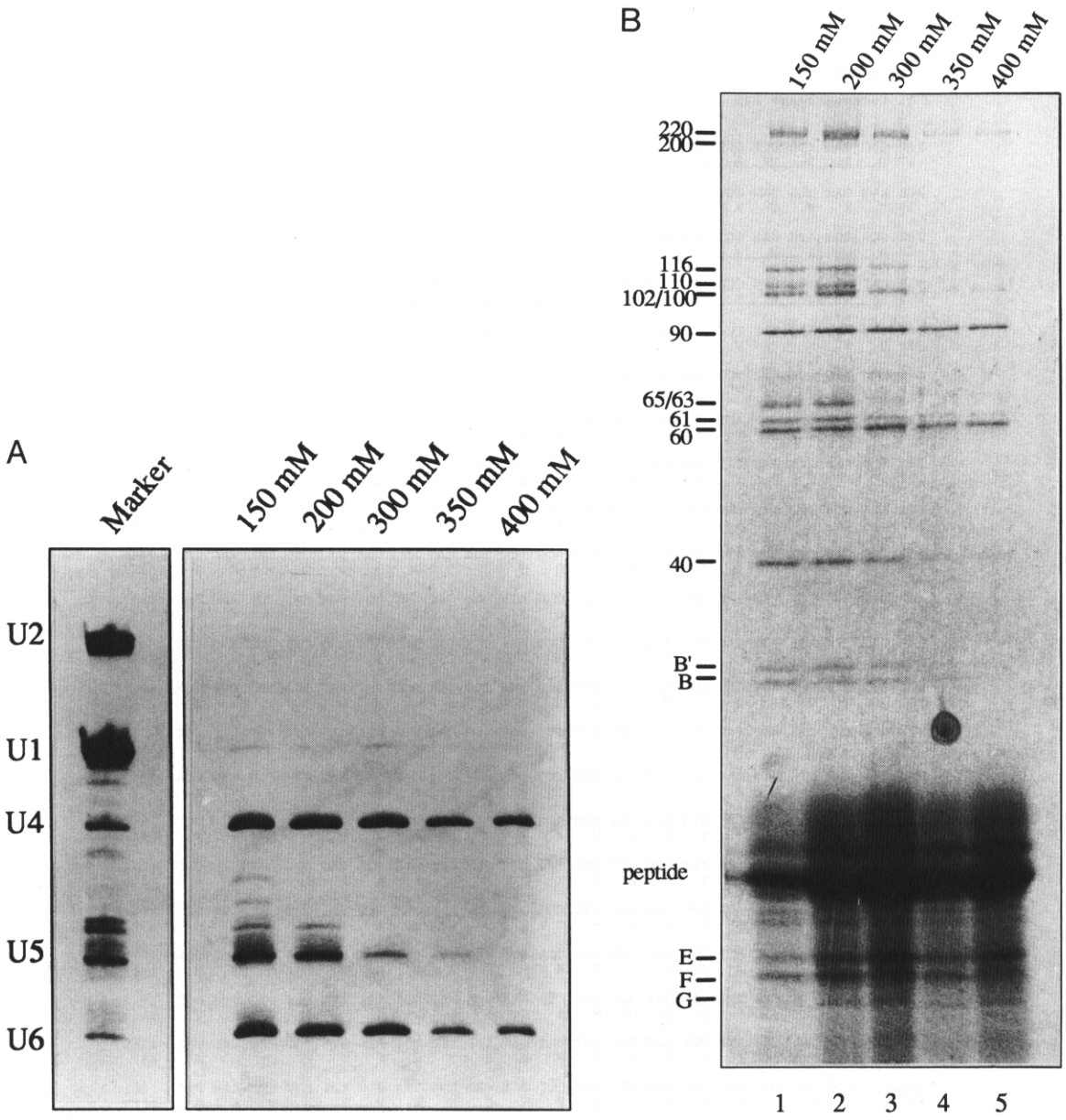

FIGURE 3. Immunoaffinity purification of U4/U6 snRNPs. Anti-m_3G-purified HeLa snRNPs were passed over an anti-60kD affinity column at various salt (KCl) concentrations as indicated above each lane. Bound particles were eluted with competing peptide (60-4.pep) as described in Materials and Methods, and the RNA and protein composition of the eluates was subsequently analyzed. **A:** snRNAs isolated from the anti-60kD column eluates were fractionated on a 10% polyacrylamide–7 M urea gel and visualized by staining with silver. The identity of the snRNAs is indicated on the left. **B:** Proteins isolated from the eluates were fractionated on a 10% SDS-polyacrylamide gel and stained with Coomassie blue. The identity of the eluted proteins is shown on the left.

of heteromeric GTP-binding (G) proteins involved in signal transduction and is normally present four to eight times within each polypeptide (reviewed by Neer et al., 1994). The conserved core of the WD repeat consists of a region 23–41 amino acids in length that is typically bracketed by GH (Gly-His) and WD (Trp-Asp): [GH-X_{23-41}-WD]. This conserved region is proceeded by a variable region that can range from 6 to 94 amino acids in length. The first WD repeat of the 60kD protein is located from amino acid 270 to 308, the sixth from 489 to 519 (Figs. 4, 7). A seventh potential WD repeat, which matches relatively poorly the consensus sequence defined by Neer et al. (1994), is located from amino acid 213 to 236. Altogether, these seven WD repeats encompass the entire C-terminal half of the U4/U6-60kD protein.

Cloning of the human U4/U6 snRNP-specific 90kD protein

A database search performed with the partial amino acid sequences obtained by microsequencing of puri-

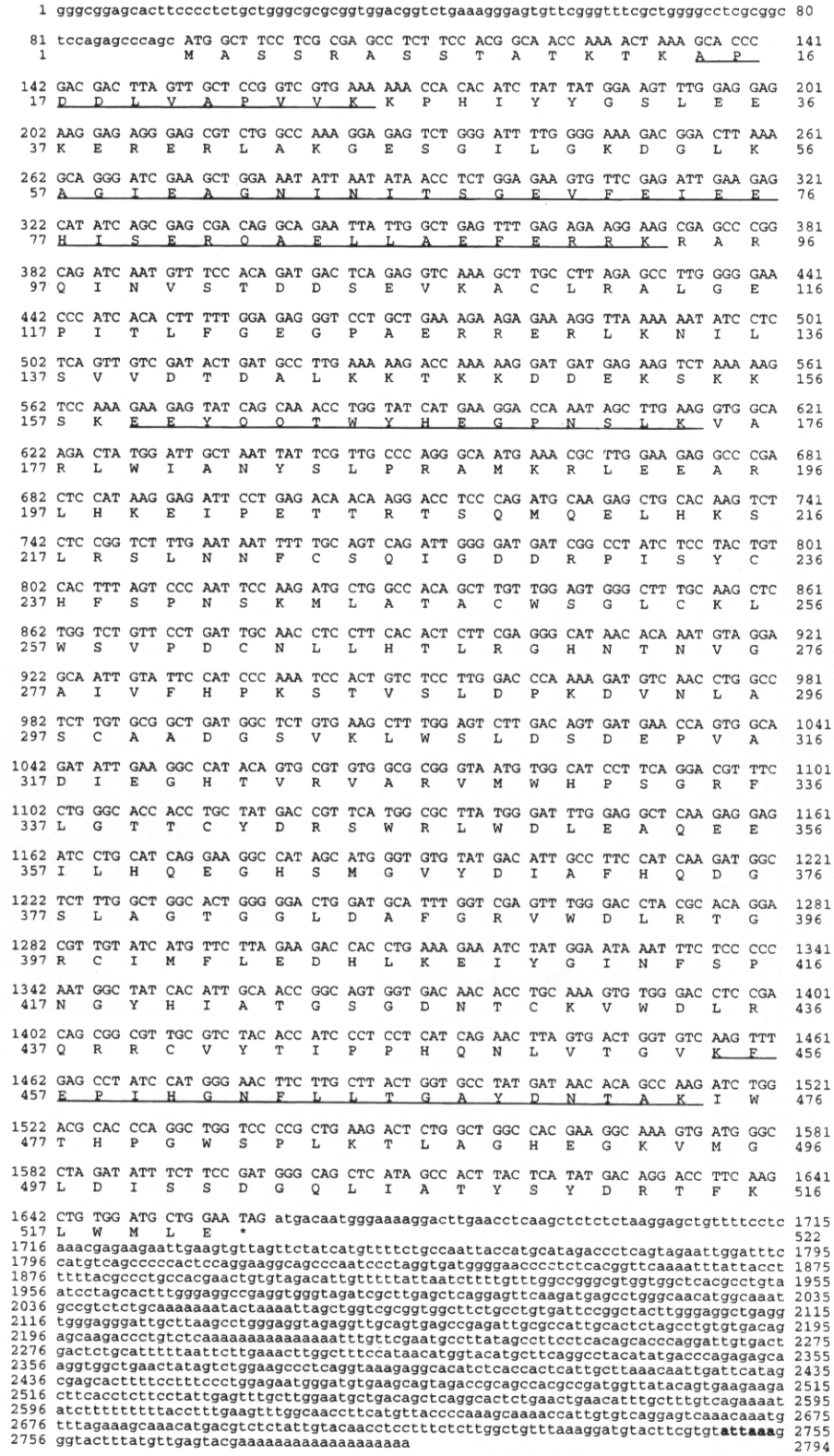

FIGURE 4. Nucleotide and predicted amino acid sequence of the U4/U6-60kD protein. The sequence of the 60kD cDNA isolated from HeLa cells is shown (Genbank accession no. AF001687). Peptide sequences obtained by microsequencing of the native protein are underlined. The polyadenylation signal in the 3′ untranslated region is shown in bold type and the termination codon is marked with an asterisk.

fied 90kD protein led to the identification of an expressed sequence tag (EST), derived from a human neuroepithelium cDNA library (clone 646542, accession no. AA205409), with an open reading frame matching one of the 90kD peptide sequences. This EST was sequenced in its entirety and, based on its sequence, appeared to encode the full length 90kD protein. The open reading frame of this cDNA encodes a protein with a predicted molecular weight of 77.5kD and a calculated isoelectric point of 9.5. All peptide sequences obtained by microsequencing of the purified 90kD protein are found in the predicted amino acid sequence of the cDNA clone, both in N- and C-terminal regions of the sequence (Fig. 5; a single amino acid difference was observed with 90-1.pep). A putative AUG initiation codon is located 72 bp from the 5′ end of this cDNA. An inframe stop codon is found directly upstream (nucleotides 66-68) of this initiator codon, confirming that it codes for the first methionine. A putative polyadenylation signal (AUUAAA, nucleotides 2320–2325) is present 197 nucleotides downstream of the stop codon (nucleotides 2121–2123), and a poly A tail is present 12 nucleotides downstream of the AUUAAA sequence, indicating that the cDNA contains the complete carboxy-terminus of the protein.

The predicted amino acid sequence of the U4/U6-90kD protein contains a biased amino acid composition with a large overall percentage of basic residues (18% arginine and lysine), a proline-rich region between amino acids 120 and 190, and a serine-rich amino-terminus which contains regions similar to the mixed charged clusters often found in SR proteins. At least three regions which are homologous to one another (amino acids 65–102, 374–394, and 506–544), and which share a central ELK motif, are also present. In addition, two copies (amino acids 5–21 and 518–534) of a nuclear localization signal (NLS) (Dingwall & Laskey, 1991) can be detected. Consistent with the proposed consensus sequence, these two NLSs are comprised of two basic amino acids followed by a 10–11 amino acids long spacer region and a basic cluster in which three out of five amino acids are basic (NLS1, KREL DELKPWIEKTVKR; NLS2, RKLTAEQRKVKKIKKLK). However, whether or not these NLSs are functional is presently not clear. Finally, a database BLOCKS search (Henikoff & Henikoff, 1994) revealed a region with some similarity to a double-stranded RNA binding motif (see below).

Isolated cDNAs encode proteins with apparent molecular weights of 60 and 90kD

To provide evidence that the isolated cDNAs indeed code for for full-length 60kD or 90kD protein, we analyzed the electrophoretic mobility of the cDNA-derived proteins. The 60 and 90kD cDNAs were thus transcribed and translated in vitro, and the migration behavior of the in vitro-translated proteins compared with that of the native proteins. As shown in Figure 6, in vitro-translated 60kD protein possessed an apparent molecular weight similar, if not identical, to that of native U4/U6-60kD protein (compare lanes 2 and 3). Additional evidence that the 60kD band generated by in vitro translation corresponds to the U4/U6-60kD protein was provided by immunoprecipitation studies with polyclonal anti-60kD antibodies that specifically react with the 60kD, but not the 61 or 63kD, protein (data not shown). In vitro translation of the putative U4/U6-90kD cDNA yielded a protein product that comigrates on SDS-polyacrylamide gels with the native U4/U6-90kD protein (Fig. 6, compare lanes 2 and 4). The basis for the difference between the predicted (77.5) and observed molecular weight of the U4/U6-90kD protein is not clear, but could indicate that the 90kD protein undergoes some form of posttranslational modification. Alternatively, its aberrant migration behavior could be caused by stretches of acidic amino acids present in its N-terminal half. In sum, we conclude from these results that both cDNAs encode full-length proteins.

U4/U6-60kD protein is homologous to Prp4p from yeast

A database search with the amino acid sequence of the 60kD protein revealed that it is homologous to the *S. cerevisiae* U4/U6 snRNP protein, Prp4p. An alignment of the Prp4p and 60kD sequences is shown in Figure 7. Compared with Prp4p (465 amino acids), U4/U6-60kD (521 amino acids) contains several insertions in its N-terminal domain. An overall identity of 33.9% (57.8% similarity) is observed between the two proteins and regions of homology are not restricted to a particular domain, but rather are found throughout the entire sequence. The greatest degree of homology is found in the C-terminal half of these proteins where Prp4p, like the 60kD protein, contains at least six WD repeats (Fig. 7B,C). A seventh putative WD repeat can be identified at positions 165–205 in Prp4p. Based on its significant degree of homology, we conclude that the U4/U6-60kD protein is the mammalian counterpart of Prp4p.

An alignment of the WD repeats of human U4/U6-60kD, yeast Prp4p, and bovine Gβ-transducin, the prototype member of the WD-repeat protein family, is shown in Figure 7C. The overall conservation of these motifs is striking. Members of this protein family function in biochemical pathways as diverse as signal transduction, cell cycle control, and RNA processing. The exact function of the WD repeat has not been determined, but a common feature of all members of the family is their assembly into multiprotein complexes, suggesting that it may mediate protein–protein interactions (Neer et al., 1994).

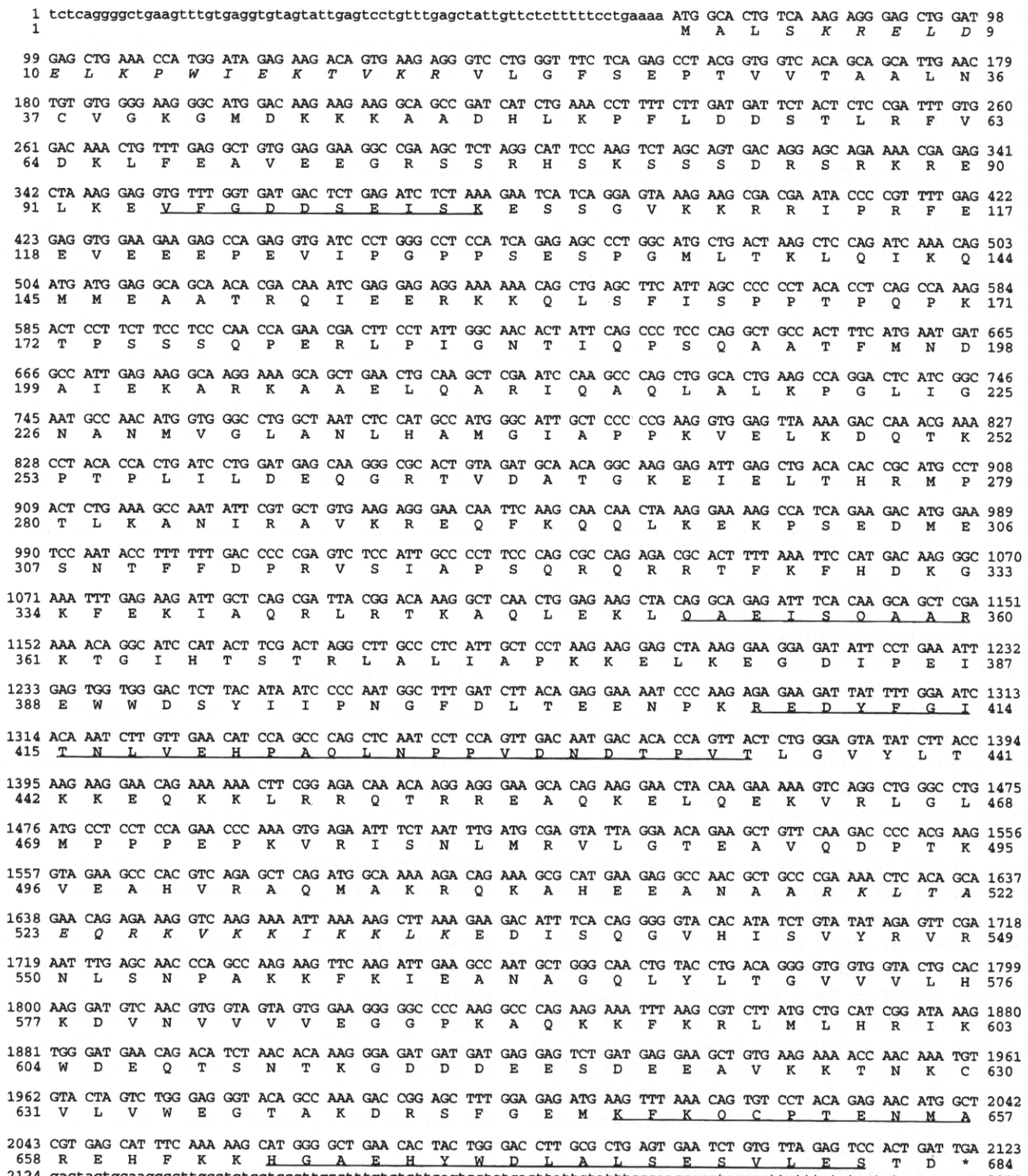

FIGURE 5. Nucleotide and predicted amino acid sequence of the U4/U6-90kD protein. The sequence of the 90kD cDNA isolated from human neuroepithelial cells is shown. Peptide sequences obtained by microsequencing of the native protein are underlined. The polyadenylation signal in the 3′ untranslated region is shown in bold type and the termination codon is marked with an asterisk. Putative nuclear localization signals are italicized.

U4/U6-90kD protein is homologous to Prp3p from *S. cerevisiae*

A database search using the amino acid sequence of the U4/U6-90kD protein led to the identification of probable homologues in *Caenorhabditis elegans* (M03C11.7, accession no. Z49128), *Schizosaccharomyces pombe* (g1044928, accession no. Z66525), and *S. cerevisiae* (g927760, accession no. U33050). The clones from

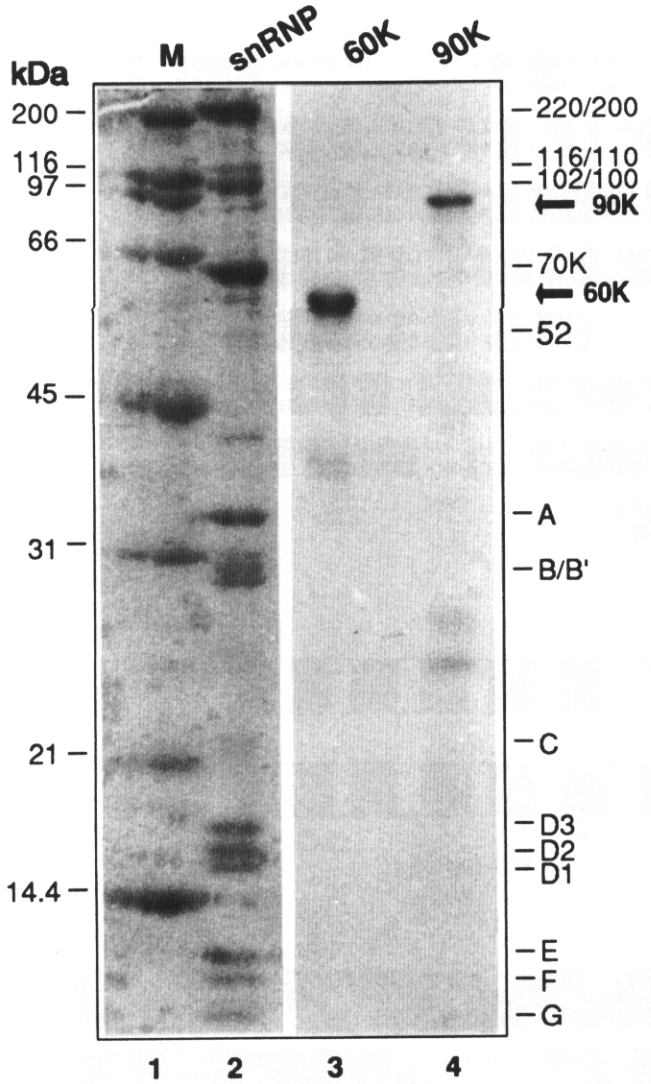

FIGURE 6. The cDNA-encoded 60kD and 90kD proteins exhibit a gel migration behavior similar to their native counterparts. Isolated cDNAs were used for in vitro translation of the 60kD (lane 3) and 90kD protein (lane 4) as described in Materials and Methods. A 5-μL aliquot of each assay was fractionated on a 12% SDS-polyacrylamide gel and visualized by fluorography. Molecular weight markers (lane 1) and total HeLa snRNP-proteins (lane 2) were fractionated on the same gel and visualized by staining with Coomassie blue. The molecular masses (in kD) of marker proteins are shown on the left and the positions of the major snRNP proteins are indicated on the right.

C. elegans and *S. pombe* potentially code for proteins with predicted molecular weights of 71.2 and 62.6kD, respectively. These polypeptides display 35.8% and 31.8% sequence identity, respectively, with the human 90kD protein. The clone from *S. cerevisiae* codes for a 469 amino acid-long protein with a predicted molecular weight of 55.8kD. The *S. cerevisiae* protein and human U4/U6-90kD share 19.8% identity (40% similarity) in a 121 amino acid-long region at their N-termini, and 30% identity (52% similarity) in a region encompassing approximately 260 amino acids at the C-terminal part of both proteins (residues 418–682 and 218–469 of the human and yeast protein, respectively) (Fig. 8A). The human protein is 218 amino acids longer than its yeast counterpart and thus several large insertions are observed in its central region. A sequence alignment of all four 90kD homologues (data not shown) reveals a high degree of similarity among their C-terminal halves, suggesting that this domain is the most functionally significant region of the protein. Interestingly, the homologous protein from *S. cerevisiae* has been identified as Prp3p, a U4/U6 snRNP-associated protein detected previously by genetic means (J.L. Woolford, pers. comm.). Based on their shared homology, we conclude that the U4/U6-90kD protein from neuroepithelial cells is the human counterpart of the yeast Prp3p.

Several short regions of homology between U4/U6-90kD and *E. coli* RNase III were also detected (Fig. 8B). Interestingly, one of these regions (amino acids 149–222 and 554–626 in RNase III and the 90kD protein, respectively) encompasses the double-stranded (ds) RNA binding domain of RNase III (Kharrat et al., 1995). Similarity to the dsRNA binding domains of other proteins, e.g., domain 3 of the *Drosophila* staufen protein, was also detected (St. Johnston et al., 1992; Bass et al., 1994; Bycroft et al., 1995). Significantly, amino acids known to be essential for dsRNA binding activity, such as the phenylalanine at position 32 and the lysine at position 50 of domain 3 of the *Drosophila* staufen protein, are also conserved in the U4/U6-90kD protein (Fig. 8B; Bycroft et al., 1995). However, whether the 90kD protein exhibits double-stranded, or even single-stranded, RNA binding activity is presently an open question.

DISCUSSION

In the present work, we demonstrate by immunoaffinity chromatography that components of the 25S [U4/U6.U5] tri-snRNP, namely the 60 and 90kD proteins, associate preferentially with the U4/U6 snRNP (Fig. 3). The designation of these proteins as U4/U6-specific was further supported by the subsequent demonstration that they are structural homologues of the *S. cerevisiae* splicing factors, Prp4p and Prp3p, known components of the yeast U4/U6 snRNP (Figs. 7, 8) (Banroques & Abelson, 1989; Petersen-Bjørn et al., 1989; J.G. Anthony & J.L. Woolford, pers. comm.). These results are also consistent with previous immunoprecipitation studies identifying two spliceosomal associated proteins (SAPs), with molecular weights of 60 and 90kD, as components of the HeLa U4/U6 snRNP (Gozani et al., 1994). Based on their identical apparent molecular weights, SAP60 and SAP90 most likely correspond to the 60 and 90kD proteins that we have described and characterized here.

By amino acid sequence comparisons, we have identified the U4/U6-60 and 90kD proteins as mammalian

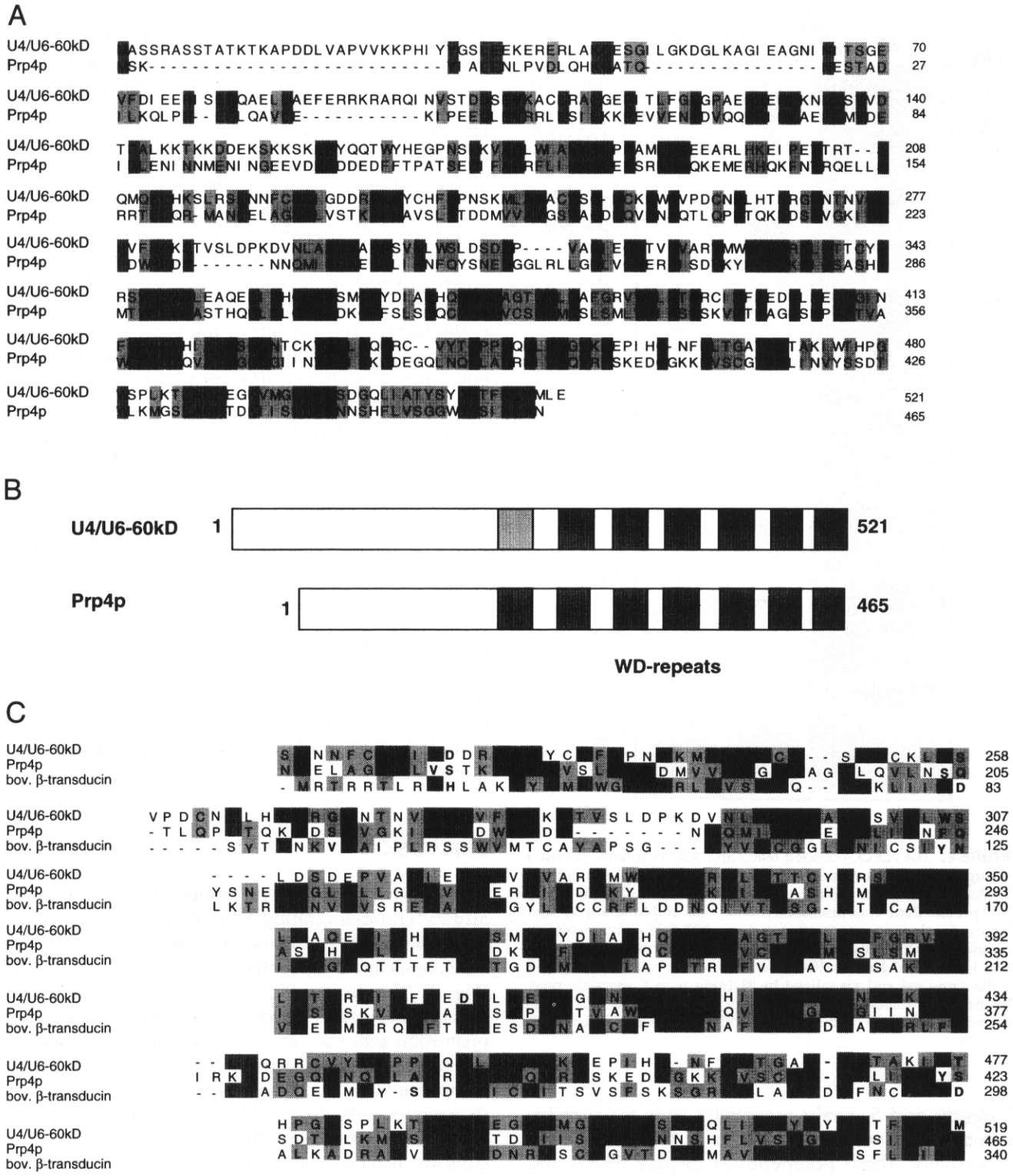

FIGURE 7. Primary structure comparison of the human 60kD and *S. cerevisiae* Prp4p proteins and alignment of their WD repeats with bovine β-transducin. **A:** Amino acid alignment of the HeLa U4/U6-60kD protein and yeast Prp4p (Genbank accession no. M28518) using the ClustalV program. Identical residues are indicated by dark grey boxes. Conserved residues are marked by light grey boxes and are grouped as follows: (LVIMAPFW), (YGCNQST), (KRH), and (ED). **B:** Schematic diagram of the primary structure of the 60kD and Prp4p. The positions of conserved WD repeats (GH-X_{27-35}-WD) within each protein are boxed; all spacing is according to scale. WD repeats of the U4/U6-60kD protein are located at the following amino acid positions: 270–308, 320–350, 362–392, 404–434, 446–477, and 489–519. A seventh potential WD repeat is located from amino acid positions 213 to 236 (light gray box). **C:** Multiple alignment of the WD repeats of the human U4/U6-60kD protein, Prp4p, and bovine β-transducin. Amino acid sequences were aligned using the ClustalV program. Conserved residues (see above) are indicated by light gray boxes and identical residues by dark gray boxes. Amino acid positions are indicated on the right.

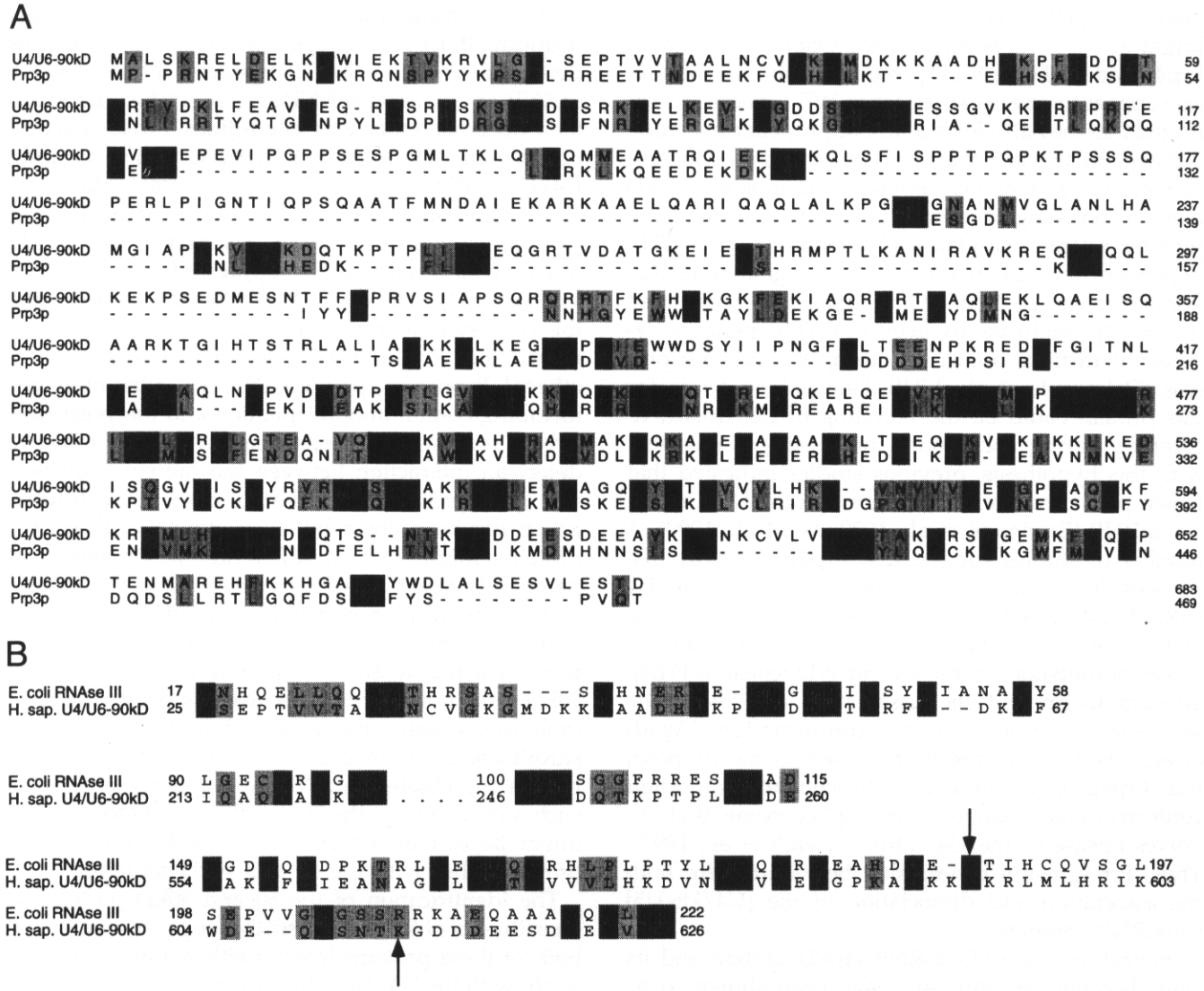

FIGURE 8. **A:** Sequence alignment of the human U4/U6-90kD and *S. cerevisiae* Prp3p proteins. Amino acids of the human 90kD protein and yeast Prp3p (Genbank accession no. U33050) were aligned using the ClustalV program. Identical residues are indicated by dark gray boxes and similar residues (grouped as described above) are marked by light gray boxes. Amino acid positions are indicated on the right. **B:** Sequence alignment between regions of the human U4/U6-90kD protein and *E. coli* RNase III (Genbank accession no. U36841). Amino acids were aligned and shaded as described above. Amino acids known to be essential for dsRNA binding activity in the *Drosophila* staufen protein (Bycroft et al., 1995), which are also conserved in the U4/U6-90kD protein, are indicated by arrowheads. RNase III and the 90kD protein exhibited 50% similarity in region 1 (amino acids 17–58 and 25–67, respectively), 56% similarity in region 2 (amino acids 90–115 and 213–260, respectively), and 40% similarity in the double-stranded binding domain of RNase III (amino acids 149–222 and 554–626, respectively). Amino acid positions are indicated at the left and right.

homologues of the yeast essential splicing factors Prp4p and Prp3p, respectively. A growing number of snRNP proteins have been shown to be both structurally and functionally evolutionarily conserved between yeast and man (for review see Beggs, 1995; Krämer, 1995). The shared structural homology of the human 60 and 90kD proteins with essential splicing factors in yeast suggests that they, too, may carry out functions essential to the mammalian splicing process. Prp4p was shown to be required early in the spliceosome assembly pathway, as well as just prior to the first catalytic step of splicing (Banroques & Abelson, 1989; Petersen-Bjorn et al., 1989; Ayadi et al., 1997). By comparison, relatively little is known about Prp3p other than that it is an essential splicing factor. The presence of at least six, perhaps even seven, WD repeats in both Prp4p and U4/U6-60kD (Fig. 7C) suggests that the function of these proteins is mediated, at least in part, by this evolutionarily conserved domain. Indeed, a detailed mutational analysis of Prp4p demonstrated that its WD repeats are required for its function in splicing (Hu et al., 1994). The structure of the prototypical WD repeat protein, namely the transducin Gβ subunit, has been solved recently and shown to form a seven-

bladed propeller-like structure (Wall et al., 1995; Lambright et al., 1996; Sondek et al., 1996). Because both Prp4p and the HeLa 60kD protein, like the Gβ subunit, also appear to possess six or seven WD repeats, it is expected that they also form a similar propeller structure.

Based on the fact that the Gβ subunits of heterotrimeric G proteins also possess a homologous WD domain consisting of seven WD repeats, and that these proteins interact with Gα GTPase subunits in a GTP-dependent, cyclical manner during signal transduction (reviewed by Neer, 1995), it has been speculated that Prp4p might be involved in GTP-dependent cyclical interactions occurring during the splicing process (Dalrymple et al., 1989). One potential role for this protein could be the recycling of the tri-snRNP complex during each round of splicing. Consistent with this hypothesis, previous studies indicated that Prp4p promotes interactions between the U4/U6 and U5 snRNP (Bordonné et al., 1990; Xu et al., 1990). A number of studies suggest that Prp4p normally dissociates from the spliceosome, along with the U4 snRNP, just prior to the first catalytic step of splicing (Tarn et al., 1993; Ayadi et al., 1997). Significantly, a subset of mutations in the second WD repeat of Prp4p appears to prevent its dissociation and block the spliceosome in an inactive conformation (Ayadi et al., 1997). These results have led to the proposal that Prp4p is required for a functionally important conformational change in the spliceosome that involves release of the U4 snRNP (Ayadi et al., 1997). Thus, Prp4p could conceivably be required for both the association and dissociation of the [U4/U6.U5] tri-snRNP complex.

Interestingly, the U5 snRNP 116kD protein and its yeast homologue, Snu114p, have been shown to be structurally homologous to the ribosomal elongation factor EF-2 and to contain a G domain with all of the consensus elements (G1–G5) typical of GTPases (Fabrizio et al., 1997). U5-116kD was additionally shown to bind GTP specifically and, in yeast, to be essential for splicing. By analogy to the heterotrimeric G proteins, the cyclical interaction of the U5 snRNP with U4/U6 to form the [U4/U6.U5] tri-snRNP complex could conceivably be mediated by GTP-dependent interactions between the U5-116kD and U4/U6-60kD proteins. However, it is not clear presently whether splicing is dependent upon the presence of GTP. Further, it is not known whether the latter two proteins do indeed interact with one another, and, if so, whether their interaction is influenced by the binding of GTP to the U5-116kD protein.

Because proteins containing WD repeats function in numerous biological pathways other than signal transduction and are present ubiquitously in many heteromeric protein complexes (Neer et al., 1994), the function of this structural motif may simply be to mediate protein–protein interactions. A number of protein–protein interactions involving Prp4p can be envisaged, including those involving U5-specific proteins (e.g., as described above), other components of the U4/U6 snRNP, or even non-snRNP splicing factors. Based on their documented genetic interactions, Prp4p was proposed initially to interact with Prp3p within the yeast U4/U6 snRNP particle (Last et al., 1987; Hu et al., 1994). Consistent with the involvement of WD repeats in such an interaction, a subset of WD repeat mutations in Prp4p can be suppressed by overexpression of Prp3p (Hu et al., 1994). Recent studies using the yeast two-hybrid system subsequently have confirmed that Prp4p and Prp3p interact with one another (J. Banroques, pers. comm.). A number of assays, including co-immunoprecipitation and the yeast two-hybrid system, have been employed in an effort to detect an interaction between the human U4/U6-60 and 90kD proteins. However, to date, no such interaction has been observed, suggesting that these two proteins might behave differently in human compared with yeast cells. Alternatively, it is conceivable that the association of the 60 and 90kD proteins requires the presence of one or more additional proteins, which play an essential role in bridging a 60/90kD protein interaction. Because the presence of low molecular weight proteins in affinity-purified U4/U6 snRNPs was obscured by the large amounts of peptide in the eluate (Fig. 3B), we cannot exclude that such a protein (e.g., the 20 or 27kD tri-snRNP protein) might be co-eluted from the anti-60kD affinity column together with the 60 and 90kD proteins.

The identification of the 60 and 90kD proteins as constituents of the U4/U6 snRNP suggests that one or both of these proteins interact either directly or indirectly with the U4/U6 RNA. The noted similarity of a region of the 90kD protein with a double-stranded RNA binding motif (Fig. 8B) raises the interesting possibility that it could interact directly with the U4/U6 snRNA. Alternatively, other components of the U4/U6 snRNP, such as the Sm proteins or additional, as yet undetected, U4/U6 snRNP proteins could also act as RNA binding proteins that help to recruit the 60 and 90kD proteins into the U4/U6 particle. Interestingly, yeast antibodies directed against Prp4p were shown to precipitate the 5′ portion of the U4 snRNA, including the conserved 5′ stem-loop structure (Bordonné et al., 1990; Xu et al., 1990). However, it remains unclear whether Prp4p binds directly to U4 or indirectly via protein–protein interactions. A similar interaction between the 60kD protein and the HeLa U4/U6 snRNP has not yet been demonstrated, but, based on the fact that the general structure of the tri-snRNP is conserved from yeast to man (Fabrizio et al., 1994), it is expected that a number of similar protein–RNA and protein–protein interactions will be found. Future in vitro reconstitution experiments with recombinant 60 and 90kD protein, as well as additional purified pro-

teins from the tri-snRNP complex, should help clarify these questions.

MATERIALS AND METHODS

Isolation and microsequencing of the 60 and 90kD proteins

Nuclear extracts were prepared from HeLa cells (Computer Cell Culture Company, Mons, Belgium) according to Dignam et al. (1983). U1, U2, U4/U6, and U5 snRNPs were purified from nuclear extracts at 250 mM NaCl by immunoaffinity chromatography using the mAb H20 as described by Laggerbauer et al. (1996). Preparative amounts of 20S U5 snRNPs or 25S [U4/U6.U5] tri-snRNPs were isolated by centrifugation of the snRNP mixture on a 10–30% (w/w) glycerol gradient according to Laggerbauer et al. (1996). Isolation of snRNP proteins for microsequencing was performed as described previously (Lauber et al., 1996). Partial amino acid sequences of tryptic peptides from the 60 and 90kD proteins were determined by microsequencing as described by Aebersold et al. (1987).

Peptide sequences obtained from the 60kD protein were:

60-1.pep EEYQQTWYHEGPNSLK
60-2.pep APDDLVAPGNK
60-3.pep FEPIHGNFLLTGAYDNTAK
60-4.pep AGIEAGNINITSGEVFEIEEHISERQAEV LAEFERRK.

Peptide sequences obtained from the 90kD protein, where small letters represent an amino acid that could not be determined unequivocally, were as follows:

90-1.pep KFKQkPTENGr
90-2.pep HGAEHYWDLALSESVLES
90-3.pep REDYFGITNLVEHPAQLNPPVDNDTPVT
90-4.pep VFGDDSEISK
90-5.pep QAEIsqAar.

Database search

Database searches were performed on the NIH mailserver using the TBLASTN and BLASTP programs (Altschul et al., 1990). Several expressed sequence tags (ESTs) with an open reading frame matching the partial peptide sequence 90-2.pep of the 90kD protein were retrieved. The longest of these corresponded to the clone 646542 (Genbank accession no. AA205409), which had been derived from a human neuroepithelium cDNA library. An open reading frame (ORF) derived from chromosome IV of *S. cerevisiae* (g927760; Genbank accession no. U33050) codes for Prp3p, which is homologous to the 90kD protein. ORFs derived from chromosome III of *C. elegans* (M03C11.7; Genbank accession no. Z49128) and chromosome I of *S. pombe* (g1044928, Genbank accession no. Z66525) also code for proteins that are homologous to the 90kD protein. No sequences were found matching the partial amino acid sequences obtained from the 60kD protein. However, a search performed with the entire amino acid sequence of the 60kD protein revealed that it is homologous to the *S. cerevisiae* Prp4p (Genbank accession no. M28518).

cDNA cloning and sequencing

90kD Protein

The EST (Genbank accession no. AA205409) containing sequences complementary to the peptide 90-2.pep was sequenced in its entirety in both directions. Sequencing was performed with an automated ABI 373 DNA sequencer (Applied Biosystems) using Taq polymerase and double-stranded templates (PRISM Ready Reaction DyeDeoxy Terminator cycle sequencing kit, Pharmacia).

60kD Protein

Degenerate oligonucleotides were designed from the partial amino acid sequence of the tryptic peptide (60-3.pep). Recombinant phage from a HeLa cDNA library (Lauber et al., 1996) were screened on duplicate filters with ^{32}P-end-labeled oligonucleotides by standard procedures (Sambrook et al., 1989). Positive phage were purified by two additional rounds of screening. The 1.9-kb insert of one positive phage was isolated and subcloned. Sequence analysis of this partial cDNA was performed as described for the 90kD protein. For isolation of 5′ end sequences missing from the partial 60kD protein cDNA, a 5′-RACE with reverse-transcribed HeLa mRNA was performed with the following primers: 60GSP2, 5′-GTC CAG ATC TTG GCT GTG TT-3′; 60GSP3, 5′-GTG ATG AAC CAG TGG CAG ATA TTG AAG G-3′; 60GSP4, 5′-CCC ATG GAT AGG CTC AAA CTT-3′. The resulting 2.7-kb DNA fragment was sequenced as described above.

In vitro transcription and translation

DNA fragments comprising the coding region of the 60kD and the 90kD protein were isolated from the respective full-length cDNA by PCR. The resulting fragments were subcloned into pBLSCRSK$^-$ (60kD) or pGEMEX-1 (90kD) (Stratagene) and linearized with *Bam*H I or *Sal* I (New England Biolabs), respectively. The linearized plasmids were used as templates for transcription with T3 (60kD) or SP6 (90kD) RNA polymerase. One microgram of each in vitro-transcribed mRNA was translated with rabbit reticulocyte lysate in the presence of [^{35}S]-methionine (Amersham) in a total volume of 75 μL as described by the manufacturer (Promega Biotech). A 5-μL aliquot was analyzed by SDS-PAGE on a 12% high-TEMED gel (Lehmeier et al., 1990). The latter was treated subsequently with Amplify (Amersham) and the proteins visualized by fluorography.

Antibody production and immunostaining

Antibodies were raised by immunization of New Zealand White Rabbits with synthetic peptides as described previously (Lauber et al., 1996). Peptides used for immunization included 60-4.pep (60kD) and 90-3.pep (90kD), as well as ones derived from microsequencing of the 61 and 63kD proteins. For western blot analysis, proteins were separated on a 12% SDS-polyacrylamide gel, transferred to nitrocellulose,

and immunostained as described previously (Lehmeier et al., 1990).

Immunoaffinity chromatography with anti-60kD antibodies

An anti-60kD immunoaffinity column was prepared by incubating 1 mL of preswollen Protein A-Sepharose (Pharmacia) with 2 mL of anti-60kD serum. Bound antibodies were coupled irreversibly to the matrix with dimethylpimelimidate as described by Harlow and Lane (1988). The antibody-Protein A-Sepharose conjugate was transferred to a column, washed three times with 10 bed volumes of PBS (20 mM potassium phosphate and 130 mM NaCl, pH 8.0), and equilibrated with 10 bed volumes of buffer W [20 mM HEPES/KOH, pH 7.9, 150 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTE, 0.5 mM PMSF, 5% (v/v) glycerol]. Approximately 250 μg of anti-m_3G immunoaffinity-purfied snRNPs, in buffer W containing 150–400 mM KCl (as indicated), were applied three times to the column at a flow rate of 0.5 mL/h. The column was washed subsequently with 20 bed volumes of buffer W. Bound snRNPs were eluted with 4 bed volumes of buffer containing an excess of competing peptide (2 mg of 60-4.pep dissolved in 50 mL buffer W). The RNA and protein composition of the eluates were analyzed, after phenol/chloroform extraction, by denaturing (7 M urea) or SDS-PAGE, respectively.

ACKNOWLEDGMENTS

We thank Axel Badouin, Peter Kempkes, and Irene Öchsner for excellent technical assistance. We are also grateful to Enno Hartmann for helpful discussions on sequence alignments and homology searches, and to John Woolford and Josette Banroques for communicating results prior to publication. This work was supported by grants from the Deutsche Forschungsgemeinschaft (SFB 272 and SFB 397) and the Fonds der Chemischen Industrie to R.L.

Received April 15, 1997; revised manuscript received April 30, 1997

REFERENCES

Aebersold RH, Leavitt J, Saavedra RA, Hood LE, Kent SBH. 1987. Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc Natl Acad Sci USA 84*:6970–6974.

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

Ayadi L, Miller M, Banroques J. 1997. Mutations within the yeast U4/U6 snRNP protein Prp4 affect a late stage of spliceosome assembly. *RNA 3*:197–209.

Bach M, Bringmann P, Lührmann R. 1990. Purification of small nuclear ribonucleoprotein particles with antibodies against modified nucleosides of small nuclear RNAs. *Methods Enzymol 181*:232–257.

Bach M, Winkelmann G, Lührmann R. 1989. 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proc Natl Acad Sci USA 86*:6038–6042.

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

Bass BL, Hurst SR, Singer JD. 1994. Binding properties of newly identified *Xenopus* proteins containing dsRNA-binding motifs. *Curr Biol 4*:301–314.

Beggs JD. 1995. Yeast splicing factors and genetics strategies for their analysis. In: Lamond AI, ed. *Pre-mRNA processing*. Austin, Texas: R.G. Landes Company. pp 79–95.

Behrens SE, Lührmann R. 1991. Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes & Dev 5*:1439–1452.

Bordonné R, Banroques J, Abelson J, 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, Appel B, Rinke J, Reuter R, Theissen H, 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 DA, Guthrie C. 1988. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature 334*:213–218.

Bycroft M, Grünert S, Murzin AG, Proctor M, St. Johnston D. 1995. NMR solution structure of a dsRNA binding domain from *Drosophila* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J 14*:3563–3571.

Cortes JJ, Sontheimer EJ, Seiwert SD, Steitz JA. 1993. Mutations in the conserved loop of human U5 snRNA generate use of novel cryptic 5′ splice sites in vivo. *EMBO J 12*:5181–5189.

Dalrymple MA, Petersen-Bjørn S, Friesen JD, Beggs JD. 1989. The product of the PRP4 gene of *S. cerevisiae* shows homology to β subunits of G proteins. *Cell 58*:811–812.

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

Dignam JD, Lebovitz RM, Roeder RG. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res 11*:1475–1489.

Dingwall C, Laskey RA. 1991. Nuclear targeting sequences—A consensus? *Trends Biochem Sci 16*:478–481.

Fabrizio P, Esser S, Kastner B, Lührmann R. 1994. Isolation of *S. cerevisiae* snRNPs: Comparison of U1 and U4/U6.U5 to their human counterparts. *Science 264*:261–265.

Fabrizio P, Laggerbauer B, Lauber J, Lane WS, Lührmann R. 1997. An evolutionarily conserved U5 snRNP-specific protein is a GTP binding factor closely related to the ribosomal translocase EF-2. *EMBO J*. Forthcoming.

Gozani O, Patton JG, Reed R. 1994. A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. *EMBO J 13*:3356–3367.

Guthrie C. 1991. Messenger RNA splicing in yeast: Clues to why the spliceosome is a ribonucleoprotein. *Science 253*:157–163.

Harlow E, Lane DP. 1988. *Antibodies. A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

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

Henikoff S, Henikoff JG. 1994. Protein family classification based on searching a database of blocks. *Genomics 19*:97–107.

Hu J, Xu Y, Schappert K, Harrington T, Wang A, Braga R, Mogridge J, Friesen JD. 1994. Mutational analysis of the PRP4 protein of *Saccharomyces cerevisiae* suggests domain structure and snRNP interactions. *Nucleic Acids Res 22*:1724–1734.

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

Kharrat A, Macias MJ, Gibson TJ, Nilges M, Pastore A. 1995. Structure of the dsRNA binding domain of *E. coli* RNase III. *EMBO J 14*:3572–3584.

Kozak M. 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res 15*:8125–8148.

Krämer A. 1995. The biochemistry of pre-mRNA splicing. In: Lamond AI, ed. *Pre-mRNA processing*. Austin, Texas: R.G. Landes Company. pp 35–64.

Laggerbauer B, Lauber J, Lührmann R. 1996. Identification of an RNA-dependent ATPase activity in mammalian U5 snRNPs. *Nucleic Acids Res 24*:868–875.

Lambright DG, Sondek J, Bohm A, Skiba NP, Hamm HE, Sigler PB. 1996. The 2.0 Å crystal structure of a heterotrimeric G protein. *Nature 379*:311–319.

Lamond AI, Konarska MM, Grabowski PJ, Sharp PA. 1988. Spliceosome assembly involves binding and release of U4 small nuclear ribonucleoprotein. *Proc Natl Acad Sci USA 85*:411–415.

Last RL, Maddock JR, Woolford JL. 1987. Evidence for related functions of the RNA genes of *Saccharomyces cerevisiae*. *Genetics 117*:619–631.

Lauber J, Fabrizio P, Teigelkamp S, Lane WS, Hartmann E, Lührmann R. 1996. The HeLa 200 kDa U5 snRNP-specific protein and its homolog in *Saccharomyces cerevisiae* are members of the DEXH-box protein family of putative RNA helicases. *EMBO J 15*:4001–4015.

Lehmeier T, Foulaki K, Lührmann R. 1990. Evidence for three distinct D proteins, which react differentially with anti-Sm autoantibodies, in the cores of the major snRNPs U1, U2, U4/U6 and U5. *Nucleic Acids Res 18*:6475–6484.

Lesser CF, Guthrie C. 1993. Mutations in U6 snRNA that alter splice site specificity: Implications for the active site. *Science 262*:1982–1988.

Lin J, Rossi JJ. 1996. Identification and characterization of yeast mutants that overcome an experimentally introduced block to splicing at the 3′ splice site. *RNA 2*:835–848.

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

Moore MJ, Query CC, Sharp PA. 1993. Splicing of precursors to messenger RNAs by the spliceosome. In: Gesteland RF, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 303–357.

Neer EJ. 1995. Heterotrimeric G proteins: Organizers of transmembrane signals. *Cell 80*:249–257.

Neer EJ, Schmidt CJ, Nambudripad R, Smith TF. 1994. The ancient regulatory-protein family of WD-repeat proteins. *Nature 371*:297–300.

Newman A, Norman C. 1991. Mutations in yeast U5 snRNA alter the specificity of 5′ splice-site cleavage. *Cell 65*:115–123.

Newman A, Norman C. 1992. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell 68*:1–20.

Nilsen TW. 1994. RNA-RNA interactions in the spliceosome: Unraveling the ties that bind. *Cell 78*:1–4.

Noble SM, Guthrie C. 1996. Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics 143*:67–80.

Petersen-Bjørn SA, Soltyk JD, Beggs JD, Friesen JD. 1989. PRP4 (RNA4) from *Saccharomyces cerevisiae*: Its gene product is associated with the U4/U6 small nuclear ribonucleoprotein particle. *Mol Cell Biol 9*:3698–37099.

Rinke J, Appel B, Digweed M, 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.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sawa H, 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.

Sawa H, Shimura Y. 1992. Association of U6 snRNA with the 5′ splice site region of pre-mRNA in the spliceosome. *Genes & Dev 6*:244–254.

Sondek J, Bohm A, Lambright DG, Hamm HE, Sigler PB. 1996. Crystal structure of a G_A protein $\beta\gamma$ dimer at 2.1 Å resolution. *Nature 379*:369–374.

Sontheimer EJ, Steitz JA. 1993. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science 262*:1989–1996.

St. Johnston D, Brown NH, Gall JG, Jantsch M. 1992. A conserved double-stranded RNA-binding domain. *Proc Natl Acad Sci USA 89*:10979–10983.

Strauss EJ, Guthrie C. 1991. A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. *Genes & Dev 5*:629–641.

Sun JS, Manley JL. 1995. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes & Dev 9*:843–854.

Tarn WY, Lee KR, Cheng SC. 1993. The yeast PRP19 protein is not tightly associated with small nuclear RNAs, but appears to associate with the spliceosome after binding of U2 to the pre-messenger RNA and prior to formation of the functional spliceosome. *Mol Cell Biol 13*:1883–1891.

Wall MA, Coleman DE, Lee E, Iniguez-Lluli JA, Posner BA, Gilman AG, Sprang SR. 1995. The structure of the G protein heterotrimer $G i\alpha 1\beta 1\gamma 2$. *Cell 83*:1047–1058

Wassarman DA, Steitz JA. 1992. Interactions of small nuclear RNAs with precursor messenger RNA during in vitro splicing. *Science 257*:1918–1925.

Will CL, Lührmann R. 1997a. SnRNP structure and function. In: Krainer A, ed. *Eukaryotic mRNA processing*. Oxford: IRL Press. Forthcoming.

Will CL, Lührmann R. 1997b. Protein functions in pre-mRNA splicing. *Curr Opin Cell Biol*. Forthcoming.

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

Wyatt JR, Sontheimer EJ, Steitz JA. 1992. Site-specific cross-linking of mammalian U5 snRNP to the 5′ splice site before the first step of pre-mRNA splicing. *Genes & Dev 6*:2542–2553.

Xu D, Nouraini S, Field D, Tang SJ, Friesen JD. 1996. An RNA-dependent ATPase associated with U2/U6 snRNAs in pre-mRNA splicing. *Nature 381*:709–713.

Xu Y, Petersen-Bjørn S, Friesen JD. 1990. The PRP4 (RNA4) protein in *Saccharomyces cerevisiae* is associated with the 5′ portion of the U4 snRNA. *Mol Cell Biol 10*:1217–1225.

Yean SL, Lin RJ. 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.