

Identification and characterization of human genes encoding Hprp3p and Hprp4p, interacting components of the spliceosome

Anan Wang¹, Julie Forman-Kay², Yu Luo⁴, Ming Luo⁴, Yu-Hua Chow¹, Jonathan Plumb¹, James D. Friesen^{3,5}, Lap-Chee Tsui³, Henry H. Q. Heng⁶, John L. Woolford, Jr⁷ and Jim Hu^{1,*}

¹Division of Respiratory Research, Hospital for Sick Children and Department of Paediatrics, ²Division of Biochemistry, Hospital for Sick Children and Department of Biochemistry and ³Department of Genetics, Hospital for Sick Children and Department of Medical Genetics, University of Toronto, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada, ⁴Center for Macromolecular Crystallography, Basic Health Science Building, Room 260, University of Alabama at Birmingham, Birmingham, AL 35294, USA, ⁵Banting and Best Department of Medical Research and Department of Medical Genetics, University of Toronto, 112 College Street, Room 411, Toronto, Ontario M5G 1L6, Canada, ⁶Biology Department, York University, 4700 Keele Street, North York, Ontario M3J 1P3, Canada and ⁷Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA

Received July 3, 1997; Revised and Accepted August 18, 1997

DDBJ/EMBL/GenBank accession nos AF001947, U82756

Nuclear RNA splicing occurs in an RNA–protein complex, termed the spliceosome. U4/U6 snRNP is one of four essential small nuclear ribonucleoprotein (snRNP) particles (U1, U2, U5 and U4/U6) present in the spliceosome. U4/U6 snRNP contains two snRNAs (U4 and U6) and a number of proteins. We report here the identification and characterization of two human genes encoding U4/U6-associated splicing factors, Hprp3p and Hprp4p, respectively. Hprp3p is a 77 kDa protein, which is homologous to the *Saccharomyces cerevisiae* splicing factor Prp3p. Amino acid sequence analysis revealed two putative homologues in *Caenorhabditis elegans* and *Schizosaccharomyces pombe*. Polyclonal antibodies against Hprp3p were generated with His-tagged Hprp3p over-produced in *Escherichia coli*. This splicing factor can co-immunoprecipitate with U4, U6 and U5 snRNAs, suggesting that it is present in the U4/U6-U5 tri-snRNP. Hprp4p is a 58 kDa protein homologous to yeast splicing factor Prp4p. Like yeast Prp4p, the human homologue contains repeats homologous to the β -subunit of G-proteins. These repeats are called WD repeats because there is a highly conserved dipeptide of tryptophan and aspartic acid present at the end of each repeat. The primary amino acid sequence homology between human Hprp4p and yeast Prp4p led to the discovery of two additional WD repeats in yeast Prp4p. Structural homology between these human and yeast splicing factors and the

β -subunit of G-proteins has been identified by sequence-similarity comparison and analysis of the protein folding by threading. Structural models of Hprp4p and Prp4p with a seven-blade β -propeller topology have been generated based on the structure of β -transducin. Hprp3p and Hprp4p have been shown to interact with each other and the first 100 amino acids of Hprp3p are not essential for this interaction. These experiments suggest that both Hprp3p and Hprp4p are components of human spliceosomes.

INTRODUCTION

Most eukaryotic genes contain intervening sequences or introns that have to be excised from primary RNA transcripts to form mature mRNA following transcription. This process of intron removal, or pre-mRNA splicing, occurs on a dynamic RNA–protein complex called the spliceosome, which contains a pre-mRNA, four essential small nuclear ribonucleoprotein (snRNP) particles (U1, U2, U5 and U4/U6), and many auxiliary proteins (1–7). Each snRNP contains one snRNA, except U4/U6 snRNP, and about a dozen proteins. U4/U6 snRNP contains U4 and U6 snRNAs in addition to the protein components. A spliceosome is assembled through a series of RNA–RNA, RNA–protein and protein–protein interactions that involve recognition of highly conserved sequences of the pre-mRNA at the 5' and 3' splice-sites and the branch site. During the early stages of spliceosome assembly, the 5' splice site is recognized by U1 snRNP through base-pair interactions with U1 snRNA and the branch site is recognized by U2 snRNP through base-pair interactions with U2

*To whom correspondence should be addressed. Tel: +1 416 813 6412; Fax: +1 416 813 5771; Email: jhu@sickkids.on.ca

snRNA (8–11). Before entering a spliceosome, U4 and U6 snRNAs interact with each other through complementary base-pair interactions to form the U4/U6 snRNP. Following binding of U1 and U2 to the pre-mRNA, the U4/U6 snRNP enters the spliceosome together with U5 snRNP as a U4/U6-U5 complex (4,5). The U4/U6 snRNP undergoes a series of conformational rearrangements during pre-mRNA splicing. For example, U4 snRNA dissociates from U6 snRNA after the tri-snRNP enters the spliceosome (12–15) so that U2 snRNA can interact with U6 to form the active spliceosome (9,16–18). Currently, little is known about how these conformational rearrangements occur and what triggers them. It is possible that some of the U4/U6-associated proteins may facilitate or regulate these events.

In yeast, several U4/U6 snRNP-specific splicing factors have been identified through genetic screening. These factors include Prp4p (19,20), Prp3p (21), Prp6p (22) and Prp24p (23). Among the known U4/U6 snRNP-associated proteins, yeast Prp4p is very intriguing because not only is it required for maintaining the U4/U6 snRNP stability (24–26) but also for interaction with the U5 snRNP (27). The 54 kDa Prp4p can be divided into three domains: N-terminal, central and C-terminal (26). The C-terminal domain contains WD repeats similar to the β -subunit of G-proteins (28). These repeats contain ~40 amino acids with a number of amino acids conserved, including a Trp-Asp dipeptide (the WD) at the end of each repeat. They are also present in many proteins involved in a variety of cellular functions (29). The structure of the β -subunit of G proteins has been recently determined independently by two groups and the WD repeats of the G- β have been shown to be involved in interactions with the α - and γ -subunits (30,31). Thus, the C-terminal domain of Prp4p may play a role in protein-protein interactions within the U4/U6 snRNP and/or in conformational rearrangements necessary for formation of the active spliceosome.

The mammalian U4/U6 snRNP-specific factors are relatively less well characterized. We report here the identification and characterization of two human splicing factors, Hprp3p and Hprp4p. Hprp3p is a human homologue of the yeast U4/U6-associated splicing factor, Prp3p. We have demonstrated that Hprp3p is present in human U4/U6-U5 tri-snRNP and more tightly associated with U4/U6 snRNP. Hprp4p is homologous to the yeast U4/U6 associated protein, Prp4p. We have shown that Hprp4p co-immunoprecipitates with Hprp3p and interacts with Hprp3p *in vitro*. The sequence similarity between human Hprp4p and yeast Prp4p led to the discovery of two additional WD repeats in yeast Prp4p. We have shown that the WD domains of these two proteins are likely folded into a seven-blade propeller structure (30–32). These observations suggest that G-protein-like interactions may be involved in U4/U6 snRNP conformation rearrangements.

RESULTS

Isolation of full size cDNA clones

Two human cDNA clones encoding U4/U6-associated splicing factors were isolated by cDNA library screening (see Materials and Methods), and the complete nucleotide sequences of the cDNAs determined. One of the clones codes for a protein of 682 amino acids (GenBank accession no. AF001947) with a predicted molecular mass of 77 kDa and the gene is named *HPRP3* because it is homologous to the yeast *PRP3* gene (Fig. 1). In addition to Prp3p of *Saccharomyces cerevisiae*, the amino acid sequences of

two other homologues from *Schizosaccharomyces pombe* and *Caenorhabditis elegans* were accessible in GenBank although the identity and function of these genes were not reported. The alignment of amino acid sequences of the human protein and its homologues in *S.cerevisiae*, *S.pombe*, and *C.elegans* (Fig. 1) shows that the C-terminal portion of the sequence is highly conserved whereas the N-terminal portion is not well conserved. The amino acid sequence of the human protein, when directly compared with each of the other three, shows 64% overall amino acid sequence similarity with the *C.elegans* protein, 62% with the *S.pombe* protein and 55% with *S.cerevisiae* Prp3p when gaps are not taken into consideration. There is a stretch of amino acids between positions 76 and 110 rich in serine and positively charged amino acids, and it is not known whether any of the serine residues in this segment are phosphorylated by SR kinase (9).

The other clone encodes 520 amino acids (GenBank accession no. U82756) with 60% overall similarity to the entire yeast Prp4p (Fig. 2) and its predicted molecular mass, 58 kDa, is consistent with the apparent molecular weight of the protein in HeLa cells detected by antibodies (Fig. 5), and the gene product expressed in *Escherichia coli* (data not shown). Based on its sequence homology to yeast Prp4p and its structural homology to β -transducin proteins (described below), Hprp4p can be divided into three domains as well: N-terminal (aa 1–164), central (aa 165–218) and C-terminal (aa 219–520). Compared with yeast Prp4p, additional amino acids are present in the N-terminal region of human Hprp4p. Since the N-terminal domain of yeast Prp4p is not critical for splicing (26), this difference is not surprising. In addition, an amino acid sequence near the end of the N-terminal domain (aa 145–158) which appears to be a nuclear localization signal (33) is not conserved in yeast Prp4p.

Immunoblot analysis of Hprp3p

Hprp3p over-expressed in *E.coli* was purified to at least 95% homogeneity and used for antibody production in rabbits. We noticed that although the predicted molecular mass of Hprp3p is 77 kDa, the protein produced in *E.coli* migrated like a 90 kDa protein in SDS-PAGE (data not shown). This cannot be attributed to a sequencing error because the whole cDNA sequence, including the poly(A) tail, was determined and the coding sequence does not have the capacity to encode a 90 kDa protein. It is likely that the slow migration in gels of Hprp3p produced in *E.coli* might be due to the positive charge of the protein (its predicted isoelectric point is ~9.9). Since human proteins with molecular masses of 60 and 90 kDa have been reported to be associated with U4/U6 snRNP, it is possible that Hprp3p is the 90 kDa protein detected in human spliceosomes (34,35). However, it is equally plausible that Hprp3p may have posttranslational modifications in mammalian cells and migrate quite differently in SDS gels from the protein expressed in *E.coli*. Since we have generated antibodies against Hprp3p, this question can be resolved by western blot analysis.

To examine the molecular masses of Hprp3p expressed in *E.coli* and in human cells, we used western blot with anti-Hprp3p antibodies to probe proteins in extracts from HeLa cells and from *E.coli* cells containing pETHR3-d or pETHR3. To reduce the background in western blots, acetone powder (from rat liver) was prepared according to a standard protocol (36) to absorb non-specific antibodies present in anti-Hprp3p serum and in pre-immune serum. As shown in Figure 3, a protein band close

Hu	MALSKRELDLKPWIEKTVKRVLGFSEPTVVTAALNCVKGMDKKAADH	50
Sp	MDRNKRRL-----	9
Ce	MG---KLEDI---CDQYLSRKKE--QDKVIEAVERCLKKGYDYKLLKDR	41
Sc	MP-PRNTYKGNPKRQNSPYKPSFLRRBETTNDEE---KFQGHGLKTE	45
	*	
Hu	LKPFLLDDSTLRFVVKLFEAVEEGR-SSRHSKSSSDRSRRELKEVFGDSS	99
Sp	-----SSQLRESSSPNS	21
Ce	VQEAISDPQK--ARKVLSAVCDHY-POLRKRTRTRDR-----GEDR	79
Sc	LHSALKSSNLNLRRTYQTGENPYLSDPHDRGSSSRFNRRY-----ER	48
Hu	EISKESGKRRIRPFEEVEEPEVPGPPSESGMLTKLQTKQMMEEA	149
Sp	QNKPNAMIEIKRR-----RLQLEQLAQ-	44
Ce	DRKRDRTRDRRKPDAVPAVAEPPKPAEQSIE----ERLRKEMMYKA	124
Sc	GLKFYQKGEISKRIA-----QERTLQK	110
Hu	TRQIEERKKQLSFI SPPTPQPKTPSSSQPERL--PIGNTIQ--PSQAATF	195
Sp	-----QAQVPWEKRSEN----GNNAGMETIQNR--	68
Ce	QQEIEETKRKMA-IAAGTLQPKAALQAKTQMVAPGGKSLSLGRDEATML	173
Sc	QQEEEELRKRLK-----QEEDKDKRKLIES-----	136
Hu	MNDAI EKARKAAELQARIQAQLALKPGLIGNANMVGLANLHAMGIAPPKV	245
Sp	I SELKEKTAK-----RFNANIPKSEGLFRDDSNAGKGLK-VGIHP--V	109
Ce	MTHSMDKSRMEELKARL-----ARTTVMKIDAM-----	203
Sc	-----GDLPNLELHEDKF	149
Hu	ELKDQTKPTPLILDEQGRVDTATGKEISVTHRMPTLKANIRAVKREQFKQ	295
Sp	LLDGNIQNTILT PENRKRRTASFSTKGVLSLQHLKPPAI--TEQNPF--	155
Ce	-----VGNAGAGV----MAPLPEAVQQAQK-KVQQ	228
Sc	LLD-----LSKFK-	157
Hu	QLKEKPEDEMESNTFFDPRVSIAPSQRORRTFKPHDKGKFEKIAQRLRTK	345
Sp	-LDTAPT RLEDSPFYDPRTKSRKTRGSRNLHLNESGKFI EEAQARRQ	204
Ce	QITKESKEPEKLEIYLDPRIQARTADRRRRGFNFHEKGEFKLANKQRAM	278
Sc	-----IYYD-----NNHGYEWDATYLDKKEG-----	179
	*	
Hu	AQLEKLQAEISQAARKTGHTSTRALALIAPKKELK-EGDIPETEWDYSYI	394
Sp	ARLEDLKKRIALHSHKAGIEDEL--ITSKSIGR--DTIPNIEWDLPF	249
Ce	AKLERLQNEVSSAAQSTGISSAVKLVPTGTAKMENGVPDI EWWDMLV	328
Sc	-----LMEKYDMNG-----TSPAEEKLAEDIDEVD-----	204
	*	
Hu	IPN-GFDLTEENPK--REDYF-GITNLVEHPAQLNPPVDNDTPVTLGVYL	440
Sp	IK--DYNDYGDENNWLIDGPQSIINSAIQHPVPVLPYAKNQPSHSHVFL	297
Ce	LDKVNYDEI PAEND--MERYSQTVSELVEHPISMRAPTEPLTQQYLRVYL	376
Sc	-----DDDDDEHPSIR-----YVAHPLPEKI--NEAKVSIKAYL	236
	**	
Hu	TKKEQKLRRTREARAEKQKELQKQVRLGLMPPPEPKVRI SNLMRVLGTEA-	489
Sp	TKKEQKIRRTREARAEKQKQDRQLLGI EPPPEPKVLSNLMHVLGDDA-	346
Ce	TTKEKKIRRNQKREVLKEKTEKIRLGLKAPPEPKVKSINLMRVLGNEA-	425
Sc	TQHERKLRNRNRKMAREARETIKILGLLFPKPEPKVLSNMMSVFNENDQ	286
	* * * * *	
Hu	VQDPTKVEAHVRAQMAKQKAHEEANAARKLTAEQRKVKKIKKLEDISQ	539
Sp	IKDPTKIEAEVRKQVEERRLRHERENEERKLTPEERKEKAFRKDEDSAA	396
Ce	IQDPTKMEAQVRKQMAERLKHETLNAERKLTEDQKRAKTKKLS EDTST	475
Sc	ITDPTAWEKVVKDQVDRKRKH-LEENERRHEDAIKRRKEAVNMNVEKPT	335
	. * * * * .	
Hu	GVHISVYRVRNLSNPAKFKIEANAGQLYLTVGV--VLHKDVNVVVVEG-	586
Sp	GLRCLVFRKYL AHRPHRLKIDLNAGQWATGVC--ILNANFNLVIFEAG	444
Ce	AVNVSVYRVKSLAHPKFKVETNAKQLQMSGAI--MMHKAQNVIVVEGG	523
Sc	VYHCKVFQKLNQPKIRFKLKMNSKELSLKGLCLRIRDDGPGI IIVVGN	385
	* . . . * . . . * . . . *	
Hu	PKAQKFKRLMLHRIKWDE--QTSNTKGGDDDEESEAVKKT-NKCVLV	632
Sp	QKAIKLLKRLMLERIDWTDTSRNSIIAQGNKLVDTEGRELNYTENTCNLV	494
Ce	PKQKFKYKNLMINRIKWSD--EIQGKKDAEKDAPGE----RNLCEMI	565
Sc	EKSKFYENLVMKRIKWNEDEFELHTNTGDIKMDMHN-SISKT-----	427
	* * * * *	
Human	WEGTAKDRSPGEMKFKQCPTENMAREHFKKHG-AEHYWDLALSESVLE--	679
S Pombe	WEGEIGRRAFRYWFRSCPSENDAKSYLBEQGGAEHPWMLAKSWS----	539
C E	WEGQVRRNRFRDFTVHTATLEKQAREFFEKKG-VAQYWDLCYSTTVLLEG	614
Sc	WEGYLQDCKFKGWFMKVNDQDSLRLRLGQFDSEHFYSVPV-----	467
	*** *	
Hu	----STD	682
Sp	----ENV	542
Ce	QDILPTA	621
Sc	----QT	469

Figure 1. Sequence alignment of Hprp3p and its homologues. The alignment was carried out with the computer program CLUSTAL V. The origins of the proteins are indicated on the left: Hu, human; Ce, *C. elegans*; Sp, *S. pombe*; and Sc, *S. cerevisiae*. The numerical position of the last amino acid in each row is indicated on the right. Gaps are represented by dashes. The identical and conserved amino acid residues are marked by asterisks and periods, respectively, under the corresponding positions. The underlined amino acid residues correspond to the positions of PCR primers.

Hprp4p	MASSRASSTATKTKAPDDLAPVVKKPHIYVGSLEEKERERLAKGESGIL	50
Prp4p	M-----SKYIALENLFPVLDQHKGATQNESTADILKQLPNERLQ----AVL	41
	* * * * * * * * * * * * * *	
Hprp4p	GKDGLKAGIEAGNINITSGEVFEIEEISASDRQKYWLSLREGSEPGSRSMF	100
Prp4p	EK-----I	44
	* * * * * * * * * * * * *	
Hprp4p	PPDDSEVKACLRALGEPITLFGEGPAERRERLRNLI LSVVGTDALKKTKKD	150
Prp4p	PEEDLEVRRLLSILKKPEVVENEDVQQRRI RLAEIILMVDEIDLLENINME	94
	* * * * * * * * * * * * *	
Hprp4p	DEKSKSKSEYQQTWYHEGPNLSK-ARLWIANYSLPRAMKRLLEEARELHKE	199
Prp4p	NINGEEVDEDDDFFTPATSELI FARRFLINYSLEERSRKRRLQKEMERHQ	144
	* * * * * * * * * * * * *	
Hprp4p	IPETTRT--SQMQELHKSRLSLNFFCSQIGDDRPISYCHFSFNSKMLATA	247
Prp4p	KFNTRQELLSRRELQ--MANLLEAGSGLVSTKPI SAVSLSTDMVVATG	193
	* * * * * * * * * * * * *	
Hprp4p	CWSG-LCKLWVSDCNLHLTLRGHNTKKEQLHSIPKSTVSLDPKDVNLAS	296
Prp4p	SWAGDLQVLSLQTLQFLQKLDHSVKGIGAI DWHFDS-----NNQMIS	236
	* * * * * * * * * * * * *	
Hprp4p	CAADGSVKLWSDSDEP---VADIEGHTVVRVARVMWHPHSGRFLGTTTCYD	342
Prp4p	CAEDGLIKNFQYSNEEGGLRLLGLDVGHERRI SDVKYHPSGKFIGSASHD	286
	* * * * * * * * * * * * *	
Hprp4p	RSWRLWDLAQEEIILHQEGHSMGVYDIAFHQDGLSAGTGGDLAFGRVWDL	392
Prp4p	MTWRLWDASTHQELLQEGHDKGVFSLSFQCDGSLVCSGMDLSMLWDI	336
	* * * * * * * * * * * * *	
Hprp4p	RTGRCIMFLEGLHKEIYGINFSPNGYHIATGSGDNTCKVWDLRQRR--V	440
Prp4p	RSGSKVMTLAGHSPKIYTVAWSPNGYQVATGGDGI INVWDIRKRDEGQL	386
	* * * * * * * * * * * * *	
Hprp4p	YTI PAHQNLVTGVKFEPIHG-NFLLLTGAYDNTAKI WTHPGWSPKTLAGH	489
Prp4p	NQILAHRNITVQVRFSEKDGKGLVSCGYDNLINVYSSDTWLKMGSLAGH	436
	* * * * * * * * * * * * *	
Hprp4p	EGKVMGLDISSDGQLIATCSYDRTFKLMMAE	520
Prp4p	TDKIISLDISNNSHFLVSGWDRSIKLWN	465
	* * * * * * * * * * * * *	

Figure 2. Amino acid sequence alignment of human Hprp4p and yeast Prp4p. The amino acid sequences of Hprp4p and Prp4p were aligned with computer program, CLUSTAL V, with default settings. The identical and conserved amino acid residues are marked by asterisks and periods, respectively, under the corresponding positions. The numerical position of the last amino acid in each row are indicated on the right. Gaps are represented by dashes. The putative nuclear localization sequences in human Hprp4p are underlined.

to 90 kDa was detected in HeLa nuclear extracts by the anti-Hprp3p serum (lane 1 of the left panel); a similar size protein band was also detected in extracts from *E. coli* cells expressing the full size Hprp3p protein (lane 3 of the left panel). As expected, a slightly smaller protein band was detected when a deleted version of Hprp3p was expressed in *E. coli* (lane 2 of the left panel). Multiple protein bands below the 68 kDa marker in lane 2 are likely due to protein degradation. The acetone liver powder absorbed pre-immune serum did not detect any specific protein bands (Fig. 3, right panel). These observations suggest that Hprp3p present in HeLa cells has a gel mobility similar to the protein produced in *E. coli* and its estimated molecular mass in SDS gels is larger than its predicted molecular mass of 77 kDa.

Association of Hprp3p with snRNAs

To determine whether Hprp3p is associated with U4/U6 snRNP or U4/U6-U5 tri-snRNP, immunoprecipitation was performed with anti-Hprp3p antiserum. The snRNAs co-immunoprecipitated by anti-Hprp3p antibodies were then analyzed by primer extension. We included a U1 primer in the primer extension analysis because normally free U1 snRNP does not associate with U4/U6 snRNP and it can serve as an internal, negative control for specificity of the immunoprecipitation. The U4 primer was used to probe for the presence of U4 and U6 snRNAs because normally U4 snRNAs are associated with U6 snRNAs. To verify that the primers used can yield extension products with expected sizes, (80 bases for U1, 98 for U5 and 109 for U4), total nuclear RNAs were isolated and tested

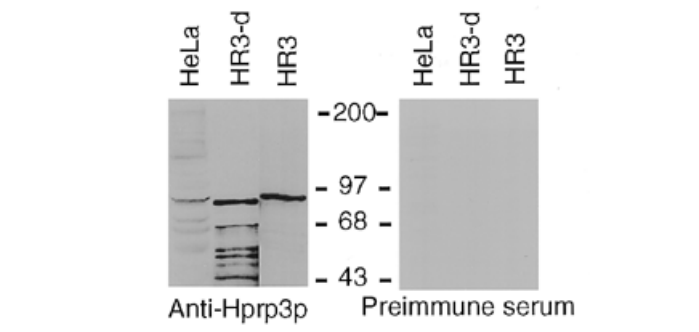


Figure 3. Immunodetection of Hprp3p in HeLa cell nuclear extract or expressed in *E. coli* cells. Cell extracts were separated in 8% SDS-PAGE and immunoblotted with polyclonal rabbit antibodies against Hprp3p. The source of the proteins is indicated on top of each lane: HeLa, 1 µl of HeLa cell nuclear extract; HR3-d, cell pellet from 20 µl *E. coli* expressing a version of His-tagged Hprp3p with its first 100 amino acids deleted; HR3, cell pellet from 20 µl *E. coli* expressing a full size Hprp3p with a His-tag. The molecular weight markers are indicated on the right.

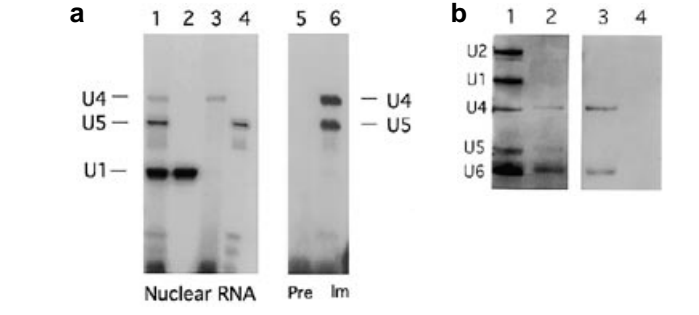


Figure 4. Co-immunoprecipitation of Hprp3p with snRNAs. (a) Primer extension analysis of the snRNAs present in the immunoprecipitates. Immunoprecipitation was performed as described in Materials and Methods. Total nuclear RNAs (left panel) or RNAs extracted from the immunoprecipitates (right panel) were subject to primer extension analysis. A combination of U1, U4 and U5 primers (lanes 1, 5 and 6) or individual primers for U1 (lane 2), U4 (lane 3) and U5 (lane 4) were used in the analysis. Pre, pre-immune serum; Im, anti-Hprp3p serum. Because the annealing conditions for different primers are likely to be different, the primer extension described here can only be used as a qualitative, rather than quantitative measurement. (b) Northern blot analysis of the snRNA present in the immunoprecipitates. Total nuclear RNA (lane 1) or RNAs recovered from immunoprecipitation with anti-Hprp3p antibodies under normal conditions (lane 2), or in the presence of high salt (500 mM NaCl) and high detergent (0.5% NP-40) concentrations (lane 3), or with preimmune serum (lane 4), were subject to northern blot analysis with DIG-labeled probes of U1, U2, U4, U5 and U6 snRNAs (see Materials and Methods for details). The low intensity of U4 and U5 snRNAs (lanes 1 and 2) is due to uneven probe concentrations in the hybridization mixture. The migration positions of U1 (164 nt), U2 (187 nt), U4 (145 nt), U5 (116 nt) and U6 (106 nt) are indicated on the left.

with primer extension. As shown in the left panel of Figure 4a, a combination of U1, U4 and U5 primers (lane 1) or individual primers (lanes 2-4) yielded correct primer extension products. The snRNAs from the immunoprecipitates were then analyzed with triple primers (U1, U4 and U5) by primer extension analysis. As shown in the right panel, U4 and U5 snRNA, but not U1 snRNA, were co-immunoprecipitated with Hprp3p by anti-Hprp3p serum (lane 6). Pre-immune serum from the same rabbit did not precipitate any one of the three snRNAs (lane 5).

Because our available U6 primer suitable for primer extension yields a similar size extension product as the U5 primer, we performed a northern blot analysis on the RNAs present in the

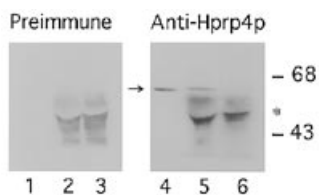


Figure 5. Immunodetection of Hprp4p in the Hprp3p complex. Immunoprecipitation was performed with polyclonal antibodies to Hprp3p (lanes 2 and 5) or preimmune serum from the same rabbit (lanes 3 and 6). The immunoprecipitates were fractionated on 7% SDS-PAGE. As a control, HeLa nuclear extracts (lanes 1 and 4) were loaded in the same gel. Proteins were transferred to a nitrocellulose filter and probed for the presence of Hprp4p with chicken antibodies against the protein (right) or preimmune IgYs from the same hen (left). The molecular markers are indicated on the right. The arrow points to Hprp4p detected by chicken antibodies against the protein, and the non-specific protein band indicated by an asterisk was contributed by the rabbit sera.

immunoprecipitates and confirmed that U4, U5 and U6, but not U1 and U2, snRNAs are present in the same complex precipitated by anti-Hprp3p antibodies (Fig. 4b, lane 2). U5 snRNA was absent in the complex when more stringent washing conditions were used (Fig. 4b, lane 3) and pre-immune serum did not precipitate any snRNA. These observations are consistent with the results of our primer extension analysis (Fig. 4a) and suggest that Hprp3p is present in U4/U6-U5 tri-snRNP and more tightly associated with U4/U6 snRNP.

Association of Hprp4p with Hprp3p

Since yeast Prp4p is an integral part of the U4/U6 snRNP (19,20), the human homologue, Hprp4p, is also expected to be associated with U4 and U6 snRNAs. To determine whether Hprp4p is associated with U4/U6 snRNP, we precipitated the Hprp3p complex with polyclonal antibodies against Hprp3p and probed the immunocomplex for the presence of Hprp4p with chicken antibodies to Hprp4p. As shown in Figure 5, a protein of ~58 kDa was detected with anti-Hprp4p antibodies in HeLa nuclear extracts (lane 4), and in immunocomplexes precipitated with antibodies against Hprp3p (lane 5), but not with pre-immune serum from the same rabbit (lane 6). This 58 kDa protein was not detected in HeLa nuclear extracts (lane 1) or immunocomplexes (lanes 2 and 3) with pre-immune chicken IgYs. The strong non-specific signal in lanes 2, 3, 5 and 6 was contributed by rabbit sera. These observations suggest that Hprp4p is associated with Hprp3p as a component of human U4/U6-U5 tri-snRNP and they are consistent with our finding that Hprp3p produced in *E.coli* interacts with Hprp4p in HeLa nuclear extracts (see below).

Hprp3p interacts with Hprp4p present in HeLa cell nuclear extracts

As a first step to probe the interactions in human U4/U6 snRNP, we examined whether Hprp3p produced in *E.coli* can interact with Hprp4p present in HeLa cell nuclear extracts. Hprp3p affinity beads were generated (see Materials and Methods) by binding His-tagged Hprp3p or Hprp3p-d in *E.coli* extracts to Ni-NTA resin (Qiagen Inc.). As a control, one aliquot of Ni-NTA resin was also mixed with extracts from *E.coli* cells harboring the vector plasmid, pET28a, and washed in the same way. The protein-affinity beads, or control beads, were then mixed with HeLa cell nuclear extracts and washed. As shown in Figure 6,

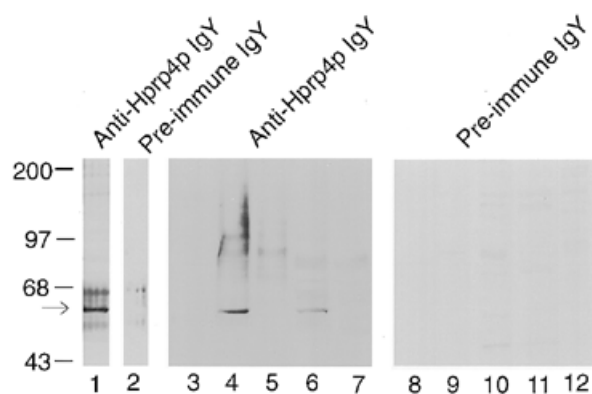


Figure 6. Interaction of Hprp3p with Hprp4p present in HeLa cell nuclear extract. To assess the antibody specificity under the immunoblotting conditions, HeLa cell nuclear extracts were separated on SDS-PAGE and probed directly with anti-Hprp4p IgY (lane 1) or preimmune IgY (lane 2). Ni-NTA resin was mixed with extracts from *E.coli* cells expressing the full size Hprp3p (lanes 4, 5, 9 and 10) or truncated Hprp3p (6, 7, 11 and 12) and washed extensively with buffer (see Materials and Methods). To assess the non-specific background from *E.coli* encoded proteins, a portion of Ni-NTA resin was mixed with extracts of *E.coli* cells containing only the vector plasmid, pET28a (lanes 3 and 8) and washed in the same way. A portion of each batch of the washed beads (lanes 4, 6, 9 and 11) was then mixed with HeLa nuclear extracts and washed again with a buffer containing 500 mM NaCl. The rest of the protein affinity beads were saved for comparison (lanes 5, 7, 10 and 12). Proteins from each sample were separated on 8% SDS-PAGE and probed with anti-Hprp4p IgY or preimmune IgY.

anti-Hprp4p chicken antibodies reacted with Hprp4p in HeLa nuclear extracts (lane 1) while pre-immune chicken IgY from the same hen did not show any non-specific interaction (lane 2). With anti-Hprp4p antibodies, the Hprp3p affinity beads (lane 4), but not the control beads (lane 3) were shown to interact with Hprp4p. The anti-Hprp4p IgY did not cross-react with Hprp3p or any residual *E.coli* proteins bound to Ni beads (lane 5) and pre-immune IgY did not show any non-specific interaction (lanes 8, 9 and 10). These observations suggest that Hprp3p produced in *E.coli* can interact (directly or indirectly) with Hprp4p in HeLa cell nuclear extracts, and are consistent with the previous findings that the yeast U4/U6 associated splicing factors Prp3p and Prp4p genetically interact with each other (21,26).

The sequence alignment (Fig. 1) showed that the C-terminal part, but not the N-terminal part, of Hprp3p is highly conserved. To examine whether the first 100 amino acids of Hprp3p are required for the Hprp3p-Hprp4p interaction, we used a deleted version of Hprp3p in the analysis. As shown in Figure 6 (lane 6), the truncated Hprp3p with the first 100 amino acids deleted was also found to interact with Hprp4p.

The seven WD repeats of Hprp4p

Yeast Prp4p was the first splicing factor found to contain WD repeats and five such repeats were initially detected (28). However, when the amino acid sequence of the human splicing factor, Hprp4, is aligned with that of bovine G- β , seven WD repeats can be identified. More than 50% overall amino acid similarity was observed between the last two domains of human Hprp4p (aa 165-520) and the bovine G- β by using the PILEUP program (within the GCG Wisconsin Package). Interestingly, some of the highly conserved amino acids in β -transducins

G-beta	RRTLGRMTRRRTLGRH...LAKIYAMHWGT..DSRLLLSASQDGKLIW	92
Hprp4p	LHKSLRSLNFCQIGD...DRPISYCHFP...NSKMLATACWSGLCKLWS	257
Prp4p	LQR.MANLELAGSQLV...TKPI.SAVLSLST...DDMVVATGSAWAGDLQV	203
G-beta	SYT...TNKVHAIPLR...SSWVMTCAYAPS.GMYVACGG.LDNICSTY	134
Hprp4p	VPDNLLHLTLRGNHTKKEQLH...SIPKSTVSLDP...KDVNLASCAADGSKLWS	307
Prp4p	SQT.LQPLTKQLDASH...VKGIGADWHPD.SNNQMISCAEGLIKNFQ	247
G-beta	LKTREGNVRVSRRELGH...TGYLSCCRFLD...DNQIVTSS.GDITCALW	179
Hprp4p	LDSDEPVADIEGH...TVRVARVMWHP...SGRFLGTTCYDRSRWLD	349
Prp4p	YSNEEGGLRLLGDLVGH...ERRISDVKYHP...SGKFIGSASHDMTWRLW	293
G-beta	IET...GQQTTFITGH...TGDVMSLSLAPD.TRLFVSGA.CDASAKLW	221
Hprp4p	LEA...QEELHQBEGH...SMGVVDIAPHQD.GSLAGTGG.LDAFGRVW	391
Prp4p	AST...HQELLLQBGH...DRGVFSLSPQCD.GSLVCSGG.MDSLMLW	335
G-beta	VREGMCRQFTTGH...ESDINAICFPFN.GNAFATGS.DDATCRLF	263
Hprp4p	LRTGRCIMFLEGH...LKEIYGINFSPN.GVHIATGS.GDNTCKVW	433
Prp4p	IRSGSKVMTLAGH...SKPIYTVAWSPN.GYQVATGG.GDGIINW	377
G-beta	LR.ADQELMTYSHDNIICIGTSVTSFSSK.GRLLLAGY.DDFNCNVW	307
Hprp4p	LQRRCVYTIPIAH...QNLVTGVKFEPIHGNFLLTGA.YDNTAKIWT	476
Prp4p	IRKRDEGLQILQIAGH...RNIIVTVRFKEDGGKLVSCCYDNLINYS	423
G-beta	ALKADRAGVLAGH...DNRVSCGLVTTD.GMAVATGS.WDSFLKIW	349
Hprp4p	HPGWSPLKTLAGH...EGKVMGLDISSD.GQLIATCS.YDRTFKLMMAE	520
Prp4p	SDTWLRMKSLAGH...TDKIISLDSINN.SHFLVSGG.WDRSIKLWN	465



Figure 7. Alignment of the WD repeats of β -transducin, Hprp4p and Prp4p based on the structural homology detected by THREADER (provided by David Jones) and primary sequence homology revealed by the PILEUP program (within the GCG Wisconsin Package). The amino acid residues in the WD domain of yeast Prp4p corresponding to the single-point mutations that confer lethal or conditional-lethal growth phenotype are also underlined. The residues corresponding to the last two amino acids for each repeat (WD), are in bold. Gaps are represented by periods. The numerical position of the last amino acid in each row is indicated on the right. The arrows at the bottom represent the relative positions of the four β -strands, A, B, C and D, identified in each WD repeat of the β -subunit of G protein (31).

(30,31) can be easily identified in Hprp4p (Fig. 7), indicating that the WD domain of Hprp4p may share a similar structure. Further analysis of yeast Prp4p also revealed two additional putative WD repeats (repeats 1 and 2), although the first repeat is much less conserved (Fig. 7). The presence of seven putative WD repeats in these splicing factors led us to carry out threading analysis and molecular modeling studies of these two proteins.

Threading and molecular modeling

We used the program THREADER written by David Jones (37) to confirm the structural homology of Hprp4p and Prp4p to the topology of β -transducin. Compared with one-dimensional sequence alignment, the threading method utilizes tertiary-structural information contained in the pairwise energy potentials. We added to the default fold library three coordinate sets derived from recent X-ray crystal structures of the β -subunit of the G-protein which has a β -propeller fold with seven propeller blades. As shown in Table 1, these coordinates of the G- β were taken from structures of the transducin $\alpha\beta\gamma$ trimer (31,32), and a structure of the transducin $\beta\gamma$ dimer (30). Sequences from the C-terminal regions of human Hprp4p (aa 212–520) and yeast Prp4p (aa 159–465), both of which have WD sequence motifs (28), were threaded using default gap penalties. The THREADER alignment of the Hprp4p and Prp4p sequences with the β -transducin yielded significant pairwise Z-scores with all three coordinate sets (Table 1), while all other folds in the library gave Z-scores of -3.0 or above. An alignment of the sequences using the PILEUP program (within the GCG Wisconsin Package), with some small manual adjustments to align the conserved residues involved in WD-motif stabilizing interactions, revealed $\sim 25\%$ sequence identity and almost 50%

similarity among the three proteins. In this alignment (Fig. 7), the interactions between the structurally significant His, Ser/Thr, Asp and Trp residues are reasonably well-conserved, with complete conservation in WD repeats 3, 4, 5 and 7. There are substitutions of these conserved residues in all three proteins. For example, substitution of Trp in conserved WD positions of yeast Prp4p for Leu in repeat 1, Phe in repeat 2 and Tyr in repeat 6, are observed; it is not immediately clear what the analogous residue is for the His in repeat 1.

Based on this adjusted sequence alignment, we constructed a homology model of the C-terminal regions of Hprp4p and Prp4p using the program MODELLER (38). Figure 8a and c illustrates the overall topology of the models for Hprp4p and Prp4p WD domains, demonstrating the seven-blade β -propeller fold. The conserved interactions between the His, Ser/Thr, Asp and Trp for WD motifs are illustrated for repeat 5 of Hprp4p in Figure 8b. With the insights gained from this model, we are now able to better interpret the results from our previous mutational analysis of yeast Prp4p (26) and to design experiments to study the relationship between the structure and function of these splicing factors (see Discussion).

Table 1. Threading pairwise energy Z-scores of WD splicing factors with β -transducin

	Structure 1	Structure 2	Structure 3
Hprp4p	-4.29	-3.81	-3.39
Prp4p	-4.79	-5.40	-4.90

Structure 1, coordinates from Wall *et al.* (31); structure 2, coordinates from Lambright *et al.* (32); structure 3, coordinates from Sondek *et al.* (30).

DISCUSSION

Two human genes, *HPRP3* and *HPRP4*, encoding U4/U6 snRNP-associated splicing factors, Hprp3p and Hprp4p, have been isolated and mapped to chromosomes 1q21.2 and 9q31–33, respectively (data not shown). Hprp3p is, in general, a well-conserved splicing factor, especially at its C-terminal part which is highly conserved among *Homo sapiens*, *C.elegans*, *S.pombe* and *S.cerevisiae*. Yeast Prp3p is much smaller than its homologues in human and *C.elegans* (Fig. 1) and it is not clear whether this reflects the difference in complexity of the splicing systems of different organisms.

Two proteins with apparent molecular masses of 60 and 90 kDa have been found in human U4/U6 snRNP (34,35). The 90 kDa protein is likely to be Hprp3p (77 kDa). The discrepancy between the molecular mass reported previously and that predicted from the Hprp3p coding sequence can be attributed, at least partially, to the highly positive charge; its value of predicted isoelectric point (IP) is ~ 9.99 .

Hprp3p co-immunoprecipitates with U4, U6 and U5 snRNAs, suggesting it is the functional homologue of yeast Prp3p which is essential for RNA splicing. Mutations in the yeast *PRP3* or *PRP4* gene have been shown to block RNA splicing (19–21,39) and cause U6 snRNA instability (24–26), suggesting both Prp3p and Prp4p play a critical role in RNA splicing. Genetic evidence indicates that Prp3p and Prp4p interact with each other. For example, a temperature sensitive mutation (*prp3-1*) of the *PRP3* gene can be suppressed by the presence of extra copies of the wild-type *PRP4* gene (21,26), and a double mutant of *prp4-1* and

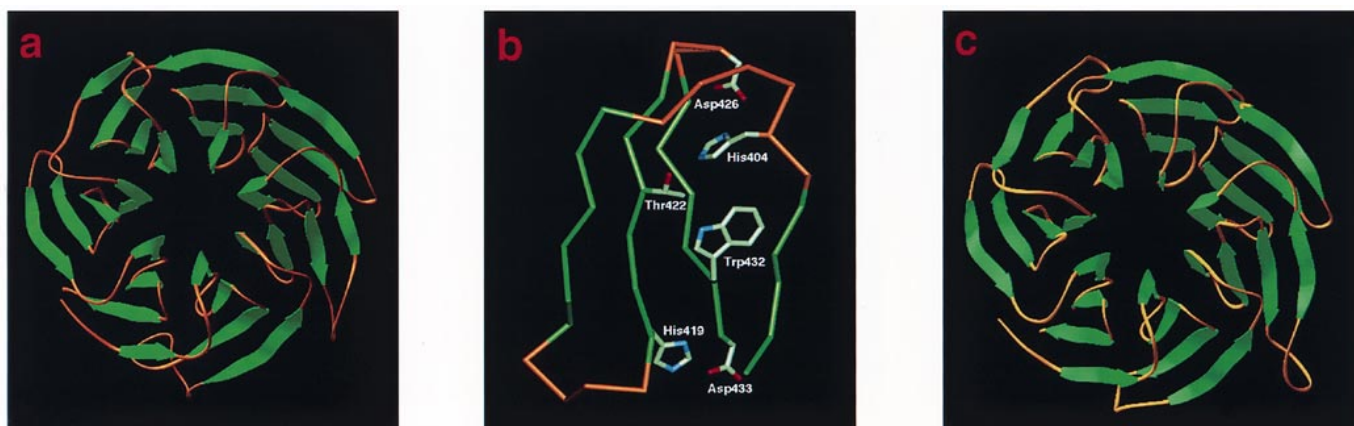


Figure 8. Proposed molecular structures of the WD domains of Hprp4p and Prp4p. (a) Ribbon model of the WD domain of human Hprp4p. (b) The conserved interactions among the His, Ser/Thr, Asp and Trp for WD repeat 5. (c) Ribbon model of the WD domain of yeast Prp4p.

prp3-1 is (synthetic) lethal (40) while both of the single mutants are viable at 25°C. Our observation of Hprp3p and Hprp4p interaction is consistent with these findings reported in yeast.

The human WD protein, Hprp4p, described here is the first identified homologue of the yeast U4/U6 snRNA-associated splicing factor Prp4p. Hprp4p can also be divided into three domains: the N-terminal domain containing the first 164 amino acids, which corresponds to the first 108 amino acids of the yeast Prp4p; the central domain spanning amino acid residues 165–218; and the C-terminal domain containing amino acid residues 219–520. It has been shown by mutational analysis that the central and the C-terminal domains of yeast Prp4p are essential for splicing and cell growth, while the N-terminal part is not essential (26). This is also reflected in the sequence similarities of these two proteins, since the only region of low conservation between these proteins is the N-terminus (Fig. 2). Although the strong amino acid sequence similarity alone is not enough to suggest that Hprp4p is involved in RNA splicing, its association with Hprp3p and snRNAs suggest that it is likely to be the 60 kDa splicing factor in the human U4/U6 snRNP (34,35).

Since the N-terminal domain of yeast Prp4p is not required when yeast cells are grown at 30°C (26), the relatively less conserved N-terminal domain of human Hprp4p may not be required for splicing, although it cannot be ruled out that this domain might be involved in regulation of splicing efficiency. In addition, it may be required for nuclear localization because the amino acid sequences, RRERLR and KKTKK, which reflect typical nuclear localization signal (NLS) sequences (33), are present in this domain between amino acid residues 128 and 150 (Fig. 2). Similarly, the non-typical NLS, RRIR (33), can be found in the same domain of yeast Prp4p. However, this suggestion is difficult to reconcile with the observation that the N-terminal domain of yeast Prp4p is not required for cell growth at 30°C. However, deletion of the N-terminal domain reduces the size of Prp4 significantly and the truncated protein may be able to diffuse freely into the nucleus.

The C-terminal or WD domain of human Hprp4p is highly homologous to that of yeast Prp4p. When the yeast *PRP4* gene was first sequenced, only the C-terminal five WD repeats were identified (28). We have shown here that the yeast protein actually contains seven WD repeats. The WD domain of yeast Prp4p has

been shown to be critical for splicing because many lethal and conditional-lethal mutations are mapped in this domain (26). The proposed structural models enable an explanation of the results of our previous mutational analysis (26). For example, the lethal and conditional-lethal mutations of the yeast *PRP4* gene can be divided into two classes (see the residues underlined in Fig. 7). One class resides in positions known to be involved in maintaining the seven-bladed propeller-structure. As indicated in Figure 7, these mutations affect the residues corresponding to the highly conserved Gly-His, or Asp and Trp-Asp positions (30,31). The second class resides on the turns between β -strands A and B (Fig. 7). Since these residues are surface-accessible, they could be involved in intermolecular interactions. In fact, one of the mutations, S320F, in repeat 4 has been found to affect its interaction with another splicing factor, Prp3p (unpublished results). The corresponding amino acid residue in Hprp4p, as well as those near this position, is also highly conserved compared with yeast Prp4p (Figs 2 and 7). Based on the sequence similarity to yeast Prp4p and structural homology to β -transducin, it is possible that the WD domain of human Hprp4p plays an important role in RNA splicing through its interactions with other splicing factors. These interactions might be important in recycling U4 snRNAs because mutations in the WD domain of yeast Prp4p inhibit the departure of U4 snRNA from the spliceosome and block the spliceosome activation (41). Currently it is not clear whether G α - and γ -like subunits are present in spliceosomes. However, the finding of the brain acetylhydrolase, a member of the phospholipase A2 superfamily with an unusual G-protein-like (α_1/α_2) β trimer (42) suggests that Hprp4p and Prp4p may form G protein-like complexes and the protein-protein interactions in these G-protein-like complexes may play an important role in regulation of conformation rearrangements in U4/U6 snRNP. The co-immunoprecipitation of Hprp4p with Hprp3p and their interaction *in vitro* suggest that these two proteins are present in the same complex, U4/U6-U5 tri-snRNP, perhaps more tightly associated with U4/U6 snRNP (Fig. 4). It is currently unclear whether the interaction of Hprp3p and Hprp4p is direct or indirect and whether the WD domain of Hprp4p is involved in this interaction. However, it is clear that the interaction is very strong because it is stable in the presence of 500 mM NaCl (Fig. 5). Our future experiments will be aimed at

characterization of their interactions and the relevance of their interaction to pre-mRNA splicing.

MATERIALS AND METHODS

Isolation of human Hprp3 cDNA clones

The GenBank non-redundant DNA sequence database was searched for sequences similar to yeast Prp3p at the amino acid level using the computer program, tblastn, which translates the DNA sequence database dynamically in all six reading frames (43). Sequences similar to yeast Prp3p were retrieved from GenBank and multiple sequence alignments were performed with CLUSTAL V (44). To isolate a full size cDNA clone encoding the human homologue of yeast Prp3p, we generated a 700 bp cDNA probe using PCR, based on the sequence homology among yeast Prp3p, its putative homologues in *S.pombe* (accession no. Z66525), and *C.elegans* (accession no. Z49128), and a partially sequenced human cDNA clone (accession no. T65920). We designed two primers (see Fig. 1 for the relative positions of the primers) for PCR amplification: a degenerate oligo, HR3DG (5'-AGGGATCCNRAYATHGARTGGTGG-3' where N = any nucleotide, R = purine, Y = pyrimidine, H = mixture of A, C and T) and a standard DNA oligo, HR3-2 (5'-TTCCTCATCAGACTCCTCATCATC-3'). This degenerate primer was derived from a stretch of conserved amino acids between *S.pombe* and *C.elegans* that contains two tryptophan residues (Fig. 1). The unique primer, HR3-2, was designed based on a short stretch of known sequence of the human cDNA clone. A human cDNA library (a gift from Dr Johanna Rommens at HSC, Toronto) was screened with the DNA probe labeled with digoxigenin-11-dUTP (Boehringer Mannheim). The nucleic acid hybridization was performed as recommended by the manufacturer. Among more than 10⁶ λ phage plaques screened, ~60 showed strong hybridization signals. Six positive isolates were purified and converted into plasmid clones in *E.coli SOLR* (Stratagene); these plasmids were named, pHR3-1, pHR3-2, pHR3-3, pHR3-4, pHR3-5 and pHR3-6. The complete sequence of the insert from pHR3-1 was determined on both strands by the dideoxy chain termination method using a T7 DNA Sequencing Kit (Pharmacia) under conditions recommended by the manufacturer. A portion of DNA sequence was determined by ACGT Corp. (Toronto) with an automated DNA Sequencing System (LI-COR, model 4000L). Other primers used in this study are: HR3-1, 5'-CGAAGG-TAGTTGCCACGTCAGAG-3'; HR3-3, 5'-CCTCTTCATGCG-CTTCTGTG-3'; HR3-4, 5'-TCTGTAACATCAAAGCCATTG-3'; HR3-5, 5'GGCTTTCAGAGTAGGCATGCG-3'; HR3-6, 5'-GATGATTCTACTCTCCGATT-3'; HR3-7, 5'-GGTACAGC-CAAAGACCGGAGC-3'; HR3-NHE, 5'-CTAGCTAGCATGG-CACTGTCAAAGAGGGAG-3'; HU1a, 5'-TGTCTCGGAT-AGAGGACGTATCAG-3'; HU4-2, 5'-CGACTATATTTCAA-GTCGTCATGGC-3'; U5a, 5'-GCTCAAAAAATTGGGTTAAG-ACTCAG-3'; U6a, 5'-CACGAATTTGCGTGTTCATCCTTG-CG-3'

Isolation of Hprp4 cDNA clones

We found that a human cDNA clone (34872) partially sequenced by the Human Genome Center (Lawrence Livermore National Lab, Livermore, CA, USA) shares significant homology at the amino acid level with the C-terminal WD repeat-containing domain of yeast Prp4p. Since WD repeats are present in many other proteins, it was not clear whether this clone, containing only

a small portion of the 3' coding region including the WD repeats, really encodes the human homologue of the yeast Prp4p. A full size cDNA clone was isolated by screening the same cDNA library as mentioned above with a 500 bp *NsiI* fragment from the clone 34872 DNA as a probe. Among more than 10⁶ λ phage plaques screened, ~50 showed strong hybridization signals. Twelve positive isolates were purified and converted into plasmid clones in *E.coli SOLR* (Stratagene). The full sequence of the coding region in one plasmid, pAW-1, was determined on both strands by the dideoxy chain termination method using a T7 DNA Sequencing Kit (Pharmacia) and oligonucleotide primers (ACGT, Corp., Toronto) under conditions recommended by the manufacturer.

Overproduction and purification of His-tagged Hprp3p and Hprp4p

To construct a plasmid that expresses His-tagged Hprp3p, an *NheI* restriction site was introduced to the 5' non-translated region of Hprp3p cDNA. This was done by swapping the 5' portion of Hprp3p cDNA in pHR3-1 with a DNA fragment generated by PCR, containing an *NheI* site introduced by the 5' PCR primer. This plasmid, pHR3NHE, is the same as pHR3-1, except for an *NheI* site present in front of the first ATG codon. The Hprp3 coding region was excised as an *NheI*-*XhoI* fragment and cloned into pET28a (Novagen) between the *NheI* and *XhoI* sites. This plasmid, pETHR3, contains a His-tag in front of the Hprp3 coding region. A deletion construct was generated by taking a *BglIII*-*XhoI* fragment from pHR3-1 and inserting it, in frame, into pET28a between *BamHI* and *XhoI* sites. The resulting plasmid was named pETHR3-d. For protein production, pETHR3 or pETHR3-d was introduced into *E.coli* BL21(DE3). The transformed *E.coli* cells were grown in LB-medium containing 25 μg/ml kanamycin and induced for 6 h with 2 mM isopropylthiogalactoside (IPTG). His-tagged Hprp3p was purified with Ni-NTA resin (Qiagen Inc.) under native conditions as suggested by the manufacturer. The His-tagged Hprp3p is soluble in *E.coli* and 1 mg of purified protein can be obtained from one liter of *E.coli* cells.

The Hprp4p coding region was amplified from pAW-1 for 25 cycles by PCR with *Pfu* polymerase (Stratagene) and inserted into pET28a (Novagen) between *BamHI* and *SalI* sites. The *BamHI* and *SalI* restriction sites in the insert were introduced from the PCR primers (HuR4Bam, 5'-CGGGATCCATGGGCGGCCGCG-CTTCTCGCGAGCCTCTTCC-3'; HUR4SalI, 5'-TTGACGTC-GACTACTTACCTATTCAGCCATCCACAGCTT-3'). This Hprp4p expression plasmid, pETHR4, was introduced in *E.coli* BL21 (DE3) and the transformed cells were induced as described above. His-tagged Hprp4p was purified with Ni-NTA resin (Qiagen Inc.) under denaturing conditions according to the protocol provided by the manufacturer.

Antibody production and purification

To generate polyclonal antibodies against Hprp3p, 500 μg of purified antigen was injected subcutaneously into each of two rabbits. The rabbits were boosted with the same amount of antigen at 1 month intervals. Pre-immune sera and serum samples were collected 2 weeks after each injection and tested by western blot to monitor antibody production. After the second boost injection one rabbit produced high titer antibodies against Hprp3p. Both rabbits were sacrificed and the sera were collected

2 weeks after the third boost injection. Serum processing was carried out as described (45).

To produce chicken antibodies against Hprp4p, purified Hprp4p was resuspended in PBS at 1 mg/ml. An aliquot of the protein suspension (500 µl) was emulsified with an equal volume of complete Freund's adjuvant (Difco, Detroit, MI). The sonified suspension was injected into two egg-laying hens at two sites in the pectoral muscle (46). The hens were boosted twice at 12 day intervals. The eggs were collected daily prior to and after the first injection and stored at 4°C. Extraction and purification of antibodies were carried out with the IgY preparation Kit, Gamma Yolk (Pharmacia) under the conditions recommended by the manufacturer.

Analysis of snRNA and Hprp4p in the immunoprecipitated complex

HeLa nuclear extracts were prepared from cell suspension culture according to a protocol previously described by Dignam *et al.* (47) and modified by Lee and Green (48). Immunoprecipitation was performed as described by Pino-Roma *et al.* (49) except that rabbit polyclonal antibodies against Hprp3p were used, instead of monoclonal antibodies. Under these conditions, the Hprp3p was precipitated together with U4/U6-U5 tri-snRNP. Immunoprecipitation was also carried out in the presence of 500 mM NaCl and 0.5% NP-40 to dissociate U5 snRNP from U4/U6 snRNP. The presence of snRNAs in the complexes was analyzed by primer extension analysis (50) or by northern blot analysis with RNA probes labeled with digoxigenin (DIG). For northern blot analysis, the RNA probes were generated by *in vitro* transcription of linearized plasmids, pSPU1, pSPU2, pSPU4, pSPU5, pT7U6 (gifts from the Steitz's group at Yale University) containing human U1, U2, U4, U5 and U6 snRNA sequences, respectively, using SP6 RNA polymerase in the presence of digoxigenin-11-UTP. Plasmids, pSPU1, pSPU2, pSPU4 and pSPU5 were linearized by *Hind*III digestion and pT7U6 was linearized by digestion with *Eco*RI. The conditions for *in vitro* transcription as well as for subsequent nucleic acid hybridization were described in detail in the DIG System User's Guide from Boehringer Mannheim.

For Hprp4p detection, the immunoprecipitates were resuspended in SDS-PAGE loading buffer and separated on 7% SDS-PAGE. The proteins were then transferred to a nitrocellulose membrane (BioRad) in a semi-dry-blot apparatus (BioRad). The presence of Hprp4p in the immunocomplexes was detected by using anti-Hprp4 IgY (see above). The secondary antibody (Zymed Laboratories, Inc.) and the enhanced chemiluminescent (ECL) reagents (Amersham) were used under the conditions recommended by the manufacturers.

Protein affinity chromatography

Ni-NTA resin (100 µl) was mixed for 1 h at 4°C with extracts of 50 ml IPTG-induced *E.coli* cells containing the expression plasmids pETHR3 or pETHR3-d. As a control, 50 µl Ni-NTA resin was mixed with extracts from *E.coli* cells harboring the expression vector pET28a. The protein-bound resin was washed five times with buffer A (50 mM NaPO₄ pH 8.0, 300 mM NaCl) and five times with buffer B (50 mM NaPO₄ pH 6.0, 500 mM NaCl, 10% glycerol). The beads were then washed once in PBS pH 7.4, and 50 µl of the washed beads was then incubated with 20 µl HeLa nuclear extract at 4°C for 30 min. After washing five times with buffer C (50 mM NaPO₄ pH 7.0, 500 mM NaCl, 10%

glycerol), proteins bound to the beads were dissolved in SDS-gel loading buffer and analyzed by western blot. To detect Hprp4p interacting with Hprp3p, western blot analysis was carried out with preimmune or immune chicken IgY (19).

Threading and molecular modeling

Threading, a promising method for protein fold recognition (37,51–53), was performed with the program THREADER written by David Jones (<http://www.biochem.ucl.ac.uk/~jones/threader.html>). This approach is based on a library of unique protein folds that have been assembled from high-resolution protein structures in the Protein Data Bank (PDB). A set of knowledge-based pairwise potentials has been derived from a statistical analysis of the known protein structures in the PDB, including information such as the correlation between various amino acid types and particular secondary structure or solvent accessibility. Fitting of the query sequence to each fold is performed using double-dynamic programming (54) with the energy computed by summing up the pairwise potential values. Insertions and deletions are taken into account with gap penalties. Folds with the lowest pairwise energies are the most probable matches. Large negative pairwise-energy Z-scores (that are calculated by subtracting average pairwise potential from pairwise potential and dividing by standard deviation of pairwise potential) are considered significant ($Z < -4.0$ for the current library of ~350 folds). Structural models of the Hprp4p and Prp4p C-terminal domains were built using the MODELLER program (38) based on the sequence alignment shown in Figure 7 and the coordinates of the β-transducin kindly provided by Dr Stephen Sprang.

ACKNOWLEDGEMENTS

We would like to thank Paul Sigler, John Sondek, David G. Lambright, Stephen R. Sprang and Mark A. Wall for providing structural coordinates, and Joan A. Steitz and Mei-Di Shu for providing human snRNA clones as probes. We would like to thank Johanna Rommens for providing the human cDNA library, David Koehler for reviewing the manuscript and Delphine Lechardeur, Gergely Lukacs, Anne Freer and C. C. Hui for their technical assistance. This work was supported by start-up funds to JH from the HSC Research Institute and a grant to JH from the Canadian Cystic Fibrosis Foundation. YC holds an MRC/CLA/Glaxo Wellcome postdoctoral fellowship.

REFERENCES

- Guthrie, C. (1991) Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science*, **253**, 157–163.
- Green, M.R. (1991) Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu. Rev. Cell Biol.*, **7**, 559–599.
- Maniatis, T. and Reed, R. (1987) The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature*, **325**, 673–678.
- Kramer, A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.*, **65**, 367–409.
- 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.
- Steitz, J.A. (1988) 'Snurps'. *Sci. Am.*, **258**, 56–60.
- Sharp, P.A. (1987) Splicing of messenger RNA precursors. *Science*, **235**, 766–771.
- Black, D.L., Chabot, B. and Steitz, J.A. (1985) U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell*, **42**, 737–750.

9. 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.
10. Zhuang, Y., Leung, H. and Weiner, A.M. (1987) The natural 5' splice site of simian virus 40 large T antigen can be improved by increasing the base complementarity to U1 RNA. *Mol. Cell. Biol.*, **7**, 3018–3020.
11. Zhuang, Y. and Weiner, A.M. (1989) A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes Dev.*, **3**, 1545–1552.
12. Cheng, S.C. and Abelson, J. (1987) Spliceosome assembly in yeast. *Genes Dev.*, **1**, 1014–1027.
13. 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.
14. Pikielny, C.W., Rymond, B.C. and Rosbash, M. (1986) Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes. *Nature*, **324**, 341–345.
15. 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.
16. 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.
17. Hausner, T.P., Giglio, L.M. and Weiner, A.M. (1990) Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. *Genes Dev.*, **4**, 2146–2156.
18. 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.
19. 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.
20. Björn, 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.
21. Last, R.L., Maddock, J.M. and Woolford, J.L., Jr (1987) Evidence for related functions of the RNA genes of *Saccharomyces cerevisiae*. *Genetics*, **117**, 619–631.
22. Abovich, N., Legrain, P. and Rosbash, M. (1990) The yeast PRP6 gene encodes a U4/U6 small nuclear ribonucleoprotein particle (snRNP) protein, and the PRP9 gene encodes a protein required for U2 snRNP binding. *Mol. Cell. Biol.*, **10**, 6417–6425.
23. Shannon, K.W. and Guthrie, C. (1991) Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs. *Genes Dev.*, **5**, 773–785.
24. Blanton, S., Srinivasan, A. and Rymond, B.C. (1992) PRP38 encodes a yeast protein required for pre-mRNA splicing and maintenance of stable U6 small nuclear RNA levels. *Mol. Cell. Biol.*, **12**, 3939–3947.
25. Galisson, F. and Legrain, P. (1993) The biochemical defects of prp4-1 and prp6-1 yeast splicing mutants reveal that the PRP6 protein is required for the accumulation of the [U4/U6.U5] tri-snRNP. *Nucleic Acids Res.*, **21**, 1555–1562.
26. Hu, J., Xu Y., Schappert, K., Harrington, T., Wang, A., Braga, R., Mogridge, J. and Friesen, J.D. (1994) Mutational analysis of the PRP4 protein of *Saccharomyces cerevisiae* suggests domain structure and snRNP interactions. *Nucleic Acids Res.*, **22**, 1724–1734.
27. Bordonne, 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.
28. Dalrymple, M.A., Björn, S.P., Friesen, J.D. and Beggs, J.D. (1989) The product of the PRP4 gene of *S. cerevisiae* shows homology to β subunits of G proteins. *Cell*, **58**, 811–812.
29. Neer, E.J., Schmidt, C.J., Nambudripad, R. and Smith, T.F. (1994) The ancient regulatory-protein family of WD-repeat proteins [published erratum appears in *Nature* (1994) Oct 27; **371**(6500), 812]. *Nature*, **371**, 297–300.
30. Sondek, J., Bohm, A., Lambright, D.G., Hamm, H.E. and Sigler, P.B. (1996) Crystal structure of a G_A protein β dimer at 2.1 Å resolution. *Nature*, **379**, 369–374.
31. Wall, M.A., Coleman, D.E., Lee, E., Iniguez-Lluhi, J.A., Posner, B.A., Gilman, A.G. and Sprang, S.R. (1995) The structure of the G protein heterotrimer G_{1 α} β γ . *Cell*, **83**, 1047–1058.
32. Lambright, D.G., Sondek, J., Bohm, A., Skiba, N.P., Hamm, H.E. and Sigler, P.B. (1996) The 2.0 Å crystal structure of a heterotrimeric G protein. *Nature*, **379**, 311–319.
33. LaCasse, E.C. and Lefebvre, Y.A. (1995) Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. *Nucleic Acids Res.*, **23**, 1647–1656.
34. Gozani, O., Patton, J.G. and 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.
35. Behrens, S.E. and Lührmann, R. (1991) Immunoaffinity purification of a (U4/U6/U5) tri-snRNP from human cells. *Genes Dev.*, **5**, 1439–1452.
36. Formosa, T., Barry, J., Alberts, B.M. and Greenblatt, J. (1991) Using protein affinity chromatography to probe structure of protein machines. *Methods Enzymol.*, **208**, 24–45.
37. Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) A new approach to protein fold recognition. *Nature*, **358**, 86–89.
38. Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
39. Lustig, A.J., Lin, R.J. and Abelson, J. (1986) The yeast RNA gene products are essential for mRNA splicing in vitro. *Cell*, **47**, 953–963.
40. Ruby, S.W., Chang, T.H. and Abelson, J. (1993) Four yeast spliceosomal proteins (PRP5, PRP9, PRP11 and PRP21) interact to promote U2 snRNP binding to pre-mRNA. *Genes Dev.*, **7**, 1909–1925.
41. Ayadi, L., Miller, M. and Banroques, J. (1997) Mutations within the yeast U4/U6 snRNP protein Prp4 affect a late stage of spliceosome assembly. *RNA*, **3**, 197–209.
42. Ho, Y.S., Swenson, L., Derewenda, U., Serre, L., Wei, Y., Dauter, Z., Hattori, M., Adachi, T., Aoki, J., Arai, H., Inoue, K. and Derewenda, Z.S. (1997) Brain acetylhydrolase that inactivates platelet-activating factor is a G-protein-like trimer. *Nature*, **385**, 89–93.
43. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
44. Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) CLUSTAL V: improved software for multiple sequence alignment. *Comp. Appl. Biosci.*, **8**, 189–191.
45. Harlow, E. and Lane, D. (1988) *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
46. Gassmann, M., Thommes, P., Weiser, T. and Hubscher, U. (1990) Efficient production of chicken egg yolk antibodies against a conserved mammalian protein. *FASEB J.*, **4**, 2528–2532.
47. 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.
48. Lee, K.A. and Green, M.R. (1990) Small-scale preparation of extracts from radiolabeled cells efficient in pre-mRNA splicing. *Methods Enzymol.*, **181**, 20–30.
49. Pinol, R.S., Choi, Y.D. and Dreyfuss, G. (1990) Immunological methods for purification and characterization of heterogeneous nuclear ribonucleoprotein particles. *Methods Enzymol.*, **181**, 317–325.
50. 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.
51. Godzik, A. and Skolnick, J. (1992) Sequence-structure matching in globular proteins: application to supersecondary and tertiary structure determination. *Proc. Natl. Acad. Sci. USA*, **89**, 12098–12102.
52. Sippl, M.J. (1993) Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comp. Aided Mol. Des.*, **7**, 473–501.
53. Bryant, S.H. and Lawrence, C.E. (1993) An empirical energy function for threading protein sequence through the folding motif. *Proteins*, **16**, 92–112.
54. Jones, D. and Thornton, J. (1993) Protein fold recognition. *J. Comp. Aided Mol. Des.*, **7**, 439–456.