The EMBO Journal vol.14 no.18 pp.4540–4550, 1995

Molecular characterization of the spliceosomal proteins U1A and U2B″ from higher plants

Gordon G.Simpson[1,2], Gillian P.Clark[1], Helen M.Rothnie[2], Wilbert Boelens[3], Walther van Venrooij[3] and John W.S.Brown[1,4]

[1]Cell and Molecular Genetics Department, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK, [2]Friedrich Miescher Institute, PO Box 2543, CH-4002 Basel, Switzerland and [3]Department of Biochemistry, University of Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands

[4]Corresponding author

In addition to their role in pre-mRNA splicing, the human spliceosomal proteins U1A and U2B″ are important models of how RNP motif-containing proteins execute sequence-specific RNA binding. Genes encoding U1A and U2B″ have been isolated from potato and thereby provide the only evolutionary comparison available for both proteins and represent the only full-length genes encoding plant spliceosomal proteins to have been cloned and characterized. *In vitro* RNA binding experiments revealed the ability of potato U2B″ to interact with human U2A′ to enhance sequence-specific binding and to distinguish cognate RNAs of either plant or animal origin. A comparison of the sequence of U1A and U2B″ proteins indicated that multiple residues which could affect RNP motif conformation probably govern the specific distinction in RNA binding by these proteins. Since human U1A modulates polyadenylation in vertebrates, the possibility that plant U1A might be exploited in the characterization of this process in plants was examined. However, unlike vertebrate U1A, neither U1A from potato nor *Arabidopsis* bound their own mRNA and no evidence for binding to upstream efficiency elements in polyadenylation signals was obtained, suggesting that plant U1A is not involved in polyadenylation.

Keywords: polyadenylation/RNA binding/RNP motif/spliceosomal protein

Introduction

The processing of nuclear pre-mRNA is a fundamental aspect of gene expression and an important level at which it can be regulated. Primary transcripts are capped at the 5′ end, introns are removed and exons ligated in a process known as pre-mRNA splicing, and the 3′ end is cleaved and polyadenylated. The analysis of these processes in higher plants has revealed significant differences in the functional *cis*-acting elements in both pre-mRNA splicing and polyadenylation compared with those signals which operate in the relatively well characterized systems of yeast and mammals (Wahle and Keller, 1992; Moore *et al.*, 1993; Filipowicz *et al.*, 1994; Hunt, 1994). For example, efficient pre-mRNA splicing in higher plants requires AU-rich intron sequences, but neither an absolutely conserved branchpoint sequence (required in the yeast, *Saccharomyces cerevisiae*) nor a 3′ splice site proximal polypyrimidine tract (required in mammals) (Goodall and Filipowicz, 1989; Filipowicz *et al.*, 1994) is required. In addition, mammalian poly(A) signals are not properly recognized in plants (Hunt *et al.*, 1987), indicating that there are differences in the *cis*-acting signals or the mechanism of polyadenylation. In contrast to the situation in mammals where each poly(A) signal generally directs processing at a single cleavage site, the 3′ ends of mRNAs from a single plant gene can be heterogeneous, indicating multiple cleavage sites (Dean *et al.*, 1986). The 'canonical' poly(A) signal, AAUAAA, is absent from a large number of plant genes (Joshi, 1987) and, even if present, is either not required (Sanfaçon, 1994) or is recognized with less stringent sequence specificity than in mammals (Rothnie *et al.*, 1994). Specific downstream sequence elements are not part of the poly(A) signal but, rather, sequences upstream of the cleavage site contain the important *cis*-acting signals (reviewed in Hunt, 1994).

Progress in the characterization of these events in higher plants has been restricted by the lack of plant extracts capable of faithfully processing pre-mRNA *in vitro* and by the lack of characterization of *trans*-acting factors. Although genes encoding all the spliceosomal UsnRNAs have been isolated from plants (Solymosy and Pollák, 1993), the *trans*-acting protein components remain almost wholly uncharacterized, and no *trans*-acting factors functioning in polyadenylation in plants have been described. The characterization of higher plant proteins involved in the processing of nuclear pre-mRNA is essential to understanding basic aspects of plant gene expression and allows evolutionarily distinct comparisons to be made with the well characterized yeast and mammalian systems. To facilitate research into higher plant nuclear pre-mRNA processing, we have previously isolated a cDNA encoding the U2snRNP-specific protein, U2B″, from potato (Simpson *et al.*, 1991). This remains the only full-length plant spliceosomal protein gene to have been characterized to date.

Although their functions in pre-mRNA splicing are currently unknown, U2B″, together with a second related spliceosomal protein, the U1snRNP-specific U1A protein, have been studied intensely as a model of RNA–protein interactions. To date, U2B″ has been isolated from human (Sillekens *et al.*, 1987) and potato (Simpson *et al.*, 1991), while U1A has been obtained from mouse (Bennet *et al.*, 1993), *Xenopus* (Scherly *et al.*, 1991) and *Drosophila* (Harper *et al.*, 1992), where it may function in splice site selection (Flickinger and Salz, 1994). A probable functional homologue of U1A has also been isolated from *Saccharomyces cerevisiae* (Liao *et al.*, 1993). Both U1A

© Oxford University Press

and U2B″ proteins consist of two RNP motifs [alternatively referred to as RNA binding domains (RBD), RNA recognition motifs (RRM) or RNP-80 motifs (Bandziulis *et al.*, 1989; Keene and Query, 1991; Kenan *et al.*, 1991)] separated by a central region. The RNP motif consists of 80–90 amino acids containing two short, highly conserved sequences (RNP1 and RNP2) and is found in numerous proteins involved in a range of RNA processing events (Kenan *et al.*, 1991; Birney *et al.*, 1993; Mattaj, 1993). Although poorly conserved at the primary sequence level, the RNP motif exhibits a characteristic fold: the highly conserved RNP1 and RNP2 sequences are found in the two central strands of a four-stranded anti-parallel β sheet packed against two α helices (Nagai *et al.*, 1990; Görlach *et al.*, 1992). In both human U1A and U2B″, only the N-terminal RNP motif is required for specific RNA binding and the minimal required region is 75% identical between the two proteins. Since the target UsnRNA sequences are also highly related, these proteins have been studied intensely by mutational swap analyses in order to determine how sequence-specific binding is achieved in RNP motif-containing proteins (see below). In view of the ubiquity of the RNP motif, an understanding of how it interacts with RNA is essential to the dissection of the mechanism of diverse RNA processing events.

Despite the relatedness of the proteins and the similarity of their target sequences, human U1A binds loop II of U1snRNA on its own (Scherly *et al.* 1989, 1990a, 1991; Lutz-Freyermuth *et al.*, 1990; Boelens *et al.*, 1991b), whereas U2B″ requires a second protein, U2A′, to bind U2 stem–loop IV specifically (Scherly *et al.*, 1990a,b; Bentley and Keene, 1991; Boelens *et al.*, 1991a; Fresco *et al.*, 1991). Mutation swap analyses of the most variable region between human U1A and U2B″ have identified distinct residues in the U2B″ β2 strand/β2–β3 protein loop and amino acids on the face of the A helix which are involved in the U2B″–U2A′ interaction that promotes U2snRNA-specific binding (Scherly *et al.*, 1990a; Bentley and Keene, 1991). However, swapping of the corresponding region from U1A into U2B″ did not promote U1snRNA-specific binding by the mutated U2B″ derivatives (Scherly *et al.*, 1990a; Bentley and Keene, 1991). In addition, mutation analyses of U1A have identified multiple residues which are critical for RNA binding, but almost all are precisely conserved between U1A and U2B″ (Nagai *et al.*, 1990; Jessen *et al.*, 1991). This was confirmed by the co-crystal structure, where residues which make contacts to the loop sequence are precisely conserved between U1A and U2B″ (Oubridge *et al.*, 1994), leaving the identification of residues in U1A which govern specific binding incompletely resolved. However, the co-crystal structure of U1A with U1 stem–loop II RNA (Oubridge *et al.*, 1994) reveals that the β2–β3 loop protrudes through the RNA loop, allowing the loop nucleotides AUUGCAC to fit the groove between the β2–β3 loop and the C-terminal end of the RNP domain, indicating that they play an important role in RNA target discrimination.

U1A is of further interest because, in addition to its usefulness as a model of RNP motif–RNA interactions and its role in pre-mRNA splicing, U1A can modulate polyadenylation. U1A autoregulates its expression by blocking polyadenylation of its own mRNA, by binding to an RNA element in the 3′ untranslated region (UTR) and inhibiting poly(A) polymerase (PAP) by specific protein–protein interactions (Boelens *et al.*, 1993; van Gelder *et al.*, 1993; Gunderson *et al.*, 1994). The mechanism of U1A autoregulation appears to be conserved in vertebrates, since two copies of a U1 loop II-like sequence can be found at a conserved spacing relative to the poly(A) signal in the 3′ UTR of U1A mRNA from man, mouse and *Xenopus* (Boelens *et al.*, 1993). U1A may also positively regulate polyadenylation efficiency by interacting with the upstream efficiency element of the simian virus 40 (SV40) late poly(A) signal (Lutz and Alwine, 1994). Functional upstream efficiency elements strikingly similar in sequence to those in the poly(A) signal of SV40 late have been identified in the poly(A) signal of cauliflower mosaic virus (CaMV) (Rothnie *et al.*, 1994) and notably the SV40 poly(A) signal is functional in higher plants (H.M.R., unpublished results). The characterization of plant U1A may therefore additionally provide a handle on *trans*-acting factors functioning in plant polyadenylation.

In this paper, we report the first functional analysis of proteins involved in the processing of nuclear pre-mRNA in higher plants. RNA binding studies demonstrate the ability of potato U2B″ to interact with human U2A′ to effect U2snRNA binding, thereby confirming it as the functional counterpart of human U2B″ and the only analogue of U2B″ to have been cloned to date. The relatedness of U1A and U2B″ was exploited by using potato U2B″ as a probe to isolate potato U1A. As the human proteins are intensely studied as models of how specific RNA binding is achieved by RNP motif-containing proteins, the characterization of plant U1A and U2B″ importantly provides the only evolutionary distinct comparison available for both proteins. Since U1A also modulates polyadenylation in vertebrates, the possibility that the isolation of U1A might be exploited to analyse polyadenylation in higher plants was examined.

Results

Molecular characterization of U2B″ from higher plants

Potato U2B″ binds specifically to stem–loop IV of U2snRNA in the presence of human U2A′. Human U2B″ binds specifically to U2snRNA stem–loop IV only in the presence of a second protein, U2A′. Specific binding is conferred on U2B″ by U2A′ in two steps. The first step involves interaction with U2B″ amino acids between positions 37 and 46, and is sufficient to allow binding to U2snRNA. The second step, which results in reduction of non-specific binding, is due to the interaction of U2A′ with U2B″ amino acids between positions 14 and 25, particularly E21 and R25 (Scherly *et al.*, 1990b). The corresponding sequences are extensively conserved in the potato U2B″ cDNA (Simpson *et al.*, 1991). Of the residues corresponding to positions 37–46 (α helix A–β2 loop/β2 strand), eight are identical to human U2B″ and another is a conservative I→V change, while both residues which correspond to E21 and R25 (face of α helix A) are identical. This raises the possibility that sequence-specific binding by potato U2B″ is achieved in a similar manner to human U2B″, requiring U2A′. As a first step in the functional characterization of this protein and in the

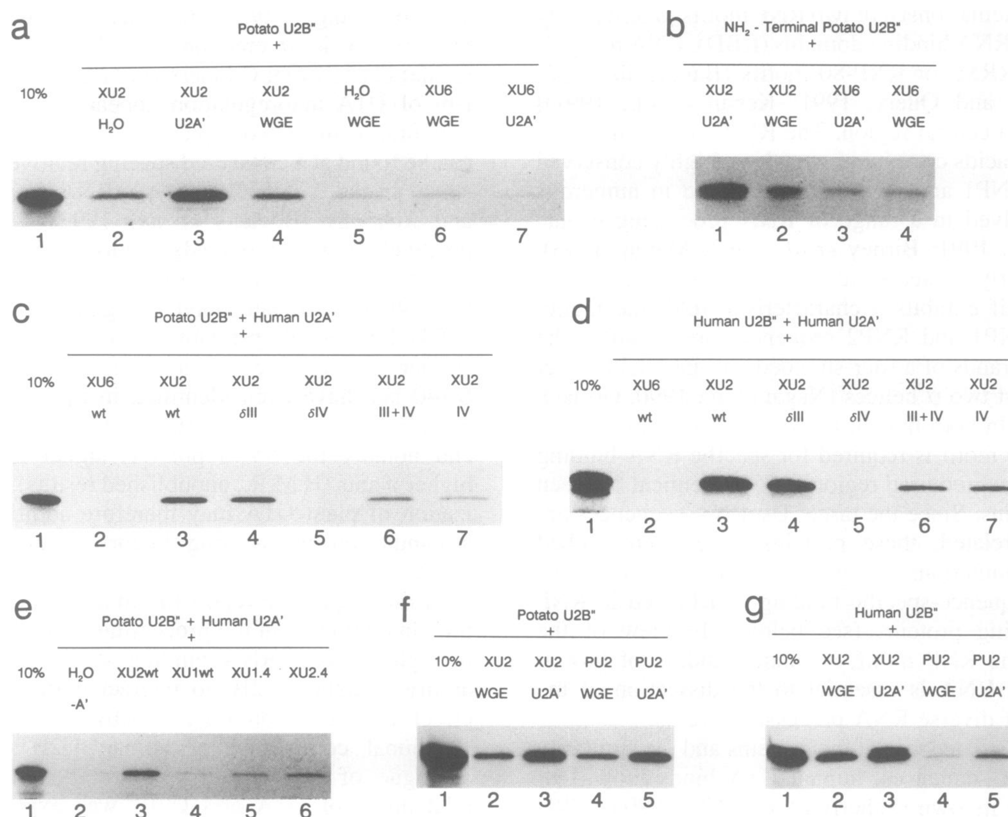

Fig. 1. RNA binding properties of potato U2B″. (**a**) Human U2A′ enhances specific binding of potato U2B″ to U2snRNAs. Binding of *in vitro* translated, ^{35}S-labelled potato U2B″ to *Xenopus* U2snRNAs (lanes 2–4) and U6snRNAs (lanes 6 and 7) in the presence of wheat germ extract primed with human U2A′ (lanes 3 and 7) or with control, unprimed wheat germ extract (WGE) (lanes 4–6). Lane 1 shows 10% of input U2B″ protein per binding assay. Enhancement of specific binding by human U2A′ was 3- to 4-fold. (**b**) Binding of a potato U2B″ derivative consisting of only the N-terminal 103 amino acids to *Xenopus* U2snRNAs (lanes 1 and 2) and U6snRNAs (lanes 3 and 4) in the presence (lanes 1 and 3) and absence (lanes 2 and 4) of human U2A′. (**c**) Binding of potato U2B″ with human U2A′ to *Xenopus* mutant U2snRNAs. Lane 1: 10% of input U2B″ protein per binding assay; lane 2: *Xenopus* U6; lane 3: wild-type U2; lanes 4 and 5: deletions of stem–loop III and IV respectively; lane 6: stem–loops III and IV only; lane 7: stem–loop IV only. (**d**) Binding of human U2B″ and human U2A′ to the same series of RNAs as described in (c). (**e**) Binding of potato U2B″/human U2A′ to U1 and U2 loop swap mutant RNAs. U1.4 is *Xenopus* U1snRNA, mutated in loop II to possess the sequence of *Xenopus* U2 loop IV, and U2.4 is the corresponding exchange mutant. Lane 1: 10% of input U2B″ per binding assay; lane 2: no U2A′, no UsnRNA control; lane 3: *Xenopus* U2 wild-type; lane 4: *Xenopus* U1 wild-type; lane 5: U1.4; lane 6: U2.4. (**f**) Binding of potato U2B″ to *Xenopus* (lanes 2 and 3) and potato U2snRNA (lanes 4 and 5) in the presence (lanes 3 and 5) or absence (lanes 2 and 4) of human U2A′. Lane 1: 10% of input U2B″ per assay. (**g**) Binding of human U2B″ to *Xenopus* (lanes 2 and 3) and potato U2snRNA (lanes 4 and 5) in the presence (lanes 3 and 5) or absence (lanes 2 and 4) of human U2A′. Lane 1: 10% of input U2B″ per assay.

absence of a cloned potato U2A′, we performed *in vitro* RNA binding studies with potato U2B″ and human U2A′.

All the RNA binding experiments described in this paper are based on the method established by Scherly *et al.* (1989) that has been used extensively in the characterization of the RNA binding activity of human U1A and U2B″ proteins (Scherly *et al.*, 1989, 1990a,b, 1991; Boelens *et al.*, 1991a,b, 1993). Briefly, [^{35}S]methionine-labelled proteins made *in vitro* in wheat germ extract were incubated with RNA into which biotinylated-UTP had been incorporated. Protein specifically bound to RNA could then be recovered by precipitation on streptavidin–agarose and analysed by SDS–polyacrylamide gel electrophoresis. Protein binding to previously characterized *Xenopus* and potato UsnRNAs was examined.

In the absence of human U2A′, potato U2B″ exhibited a weak preference for *Xenopus* U2snRNA over *Xenopus* U6snRNA. However, in the presence of human U2A′, specific binding to U2snRNA was significantly enhanced 3- to 4-fold (Figure 1a). Such specific binding and enhancement by U2A′ was also observed with a mutant version of potato U2B″ consisting of only the N-terminal 103 amino acids, indicating that sufficient sequence information for U2-specific binding and U2A′ association resides in this part of the protein (Figure 1b). We confirmed the prediction that stem–loop IV was the binding site of potato U2B″ by analysing the binding of the protein to a series of mutated *Xenopus* U2snRNAs (Hamm *et al.*, 1989). Binding of U2B″ to these mutant RNAs was only detected if stem–loop IV was present, and was undetectable with mutant U2snRNAs lacking stem–loop IV (Figure 1c). Similar results were obtained in parallel experiments with human U2B″ (Figure 1d). Since U2A′ itself exhibits weak affinity for stem IV (Boelens *et al.*, 1991a), it was possible that the specificity of the interaction we observed was due to sequence-specific binding by this protein and its association with potato U2B″. Since at least a significant part of the binding site of U2B″ is the loop IV sequence, we investigated the ability of U2B″ and human U2A′ to bind two loop-swap mutant RNAs. In construct U1.4, *Xenopus* U1snRNA loop II has been replaced by the sequence of *Xenopus* U2snRNA loop IV and in construct

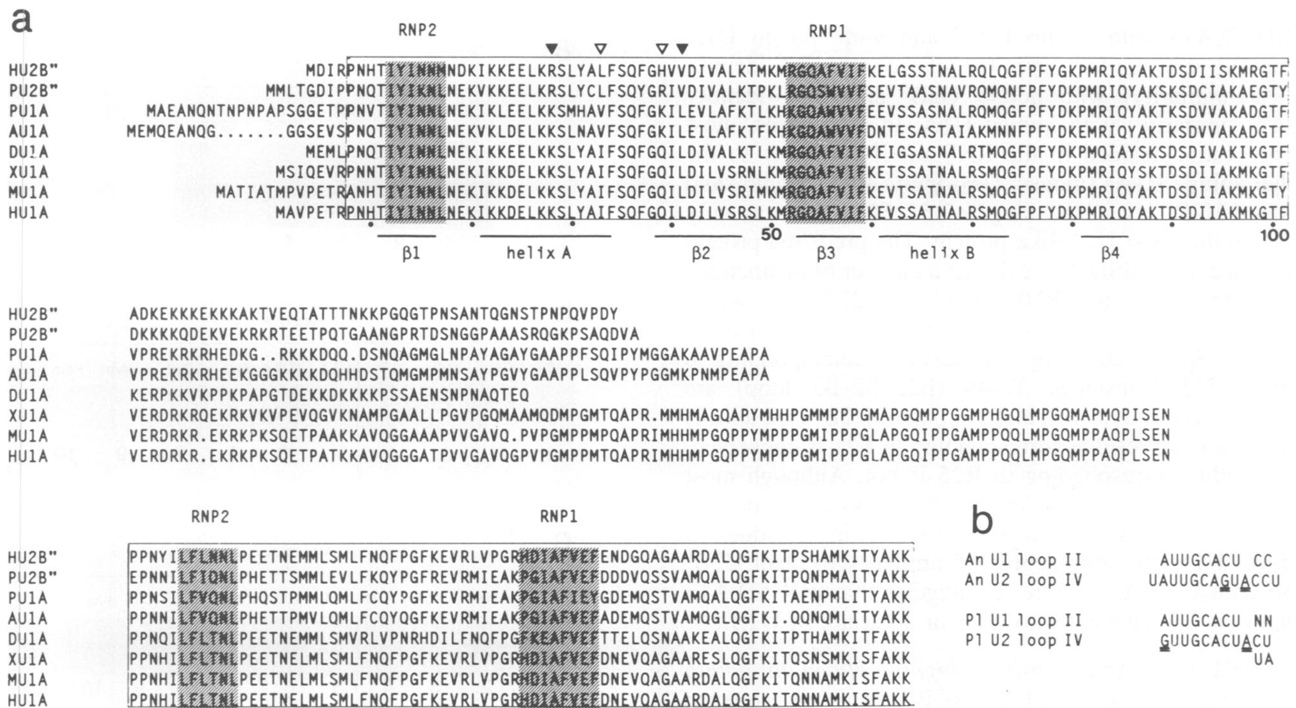

Fig. 2. (**a**) Predicted amino acid sequence of potato and *Arabidopsis* U1A and alignment with published U1A and U2B″ protein sequences: potato U2B″ (Simpson *et al.*, 1991), human U2B″ (Habets *et al.*, 1987), human U1A (Sillekens *et al.*, 1987), *Drosophila* D25 (Harper *et al.*, 1992), *Xenopus* U1A (Scherly *et al.*, 1991) and mouse U1A (Bennet *et al.*, 1993). The sequences have been arranged to highlight the N- and C-terminal RNP motifs (boxed) and the central domains. RNP1 and RNP2 are highlighted and triangles indicate positions of amino acid distinctions between U1A and U2B″ proteins (see text); closed triangles, conserved; open triangles, non-conserved. Numbering is based on the human U1A sequence. (**b**) U1 and U2snRNA target loop sequences from animals and plants. Sequence distinctions between animal U1 loop II and U2 loop IV are underlined. The 3′-most nucleotides in the plant loop sequences are variable: NN in U1snRNA, C/U and U/A in U2snRNA. EMBL accession numbers are Z49990 and Z49991.

U2.4, the converse replacement has been made (Scherly *et al.*, 1990b). We found that potato U2B″ could bind both mutant RNAs more strongly than U1snRNA, but binding was always weaker than with U2snRNA (Figure 1e), indicating that both U2B″ and U2A′ contribute to the specificity of binding observed here. We also confirmed that potato U2B″ bound specifically to potato U2snRNA in the presence of human U2A′ (Figure 1f). These results indicate that the manner of sequence-specific binding by U2B″, in requiring the presence of U2A′ for enhanced specific binding, and the U2snRNA target site, stem–loop IV, is conserved across nature. Notably, the ability of potato U2B″ to bind either potato or *Xenopus* U2snRNA appeared similar, but human U2B″ exhibited a preference for *Xenopus* over potato U2snRNA (Figure 1g). Given the similarity of U2B″ and U1A proteins and their respective target sequences, specificity of binding of U2B″ to U2snRNA was further demonstrated by the lack of binding to *Xenopus* U1snRNA either in the presence (Figure 1e) or absence of U2A′ (results not shown). This confirmed our previous observation that, *in vivo,* potato U2B″ distinguishes between U1 and U2snRNAs, as only U2snRNA was detected in immunoprecipitates from potato extracts with a U2B″-specific monoclonal antibody (Simpson *et al.*, 1991).

Molecular characterization of U1A from higher plants

Isolation of genes putatively encoding potato U1A. Since human U1A and U2B″ proteins exhibit extensive sequence conservation, particularly in the N- and C-terminal RNP motifs (Sillekens, 1987), we attempted to isolate genes encoding potato U1A by employing the potato U2B″ cDNA as a probe in a library screening programme. We obtained a genomic clone and corresponding cDNA clone, pGGSD.A13, related to, but distinct in sequence from, potato U2B″ (see Materials and methods). The predicted amino acid sequence of the cDNA clone is shown in Figure 2.

The cDNA encodes a protein with a predicted mol. wt of 28 kDa and protein translated *in vitro* in wheat germ extract from this cDNA migrates with an apparent mol. wt of 30.5 kDa in SDS–polyacrylamide gels. Like U1A and U2B″, the predicted amino acid sequence of the protein encoded by GGSD.A13 consists of two RNP motifs separated by a central domain and, notably, a tryptophan residue is found at the fifth position of RNP1 in the N-terminal RNP motif of the protein. This, the most conserved position of RNP1, is usually occupied by phenylalanine in RNP motif-containing proteins (Kenan *et al.*, 1991; Birney *et al.*, 1993) and the only other such protein with a tryptophan at this position is potato U2B″ (Simpson *et al.*, 1991).

A search of the NCBI non-redundant sequence database, employing the BLAST program, showed the protein encoded by pGGSD.A13 to be most related to potato U2B″, but U1A from human, mouse, *Xenopus* and *Drosophila* exhibited similarly high scores for relatedness. These sequences are aligned in Figure 2 and pairwise comparison of the full-length predicted amino acid sequence of

pGGSD.A13 with potato U2B″ and with human U1A revealed very similar scores for relatedness: 59% identity (79% similarity) with potato U2B″ and 57% identity (74% similarity) with human U1A. In both cases, most conservation resided in the RNP motifs. Identity to the other animal U1As (mouse, *Drosophila* and *Xenopus*) and human U2B″ lay in the range of 51–55%, but was only 22% to the yeast U1A-like protein. The predicted protein sequence of pGGSD.A13 exhibits a number of distinctions with the N-terminal RNP motif of U2B″ in residues important for U2A′ association and U2snRNA-specific binding. Specifically, only four residues corresponding to human U2B″ residues 37–46 (β2; β2–β3 loop) are identical, compared with eight in potato U2B″ (see above) and, while the residue corresponding to E21 is conserved, the residue corresponding to R25 is not. Although most of the sequence distinctions in these positions do not exhibit conservation with the corresponding residues in human U1A, the high degree of similarity to plant U2B″ and animal U1As (Figure 2) suggests that pGGSD.A13 might encode either potato U1A or a variant of U2B″.

A cDNA from Arabidopsis is highly related to potato pGGSD.A13. A search of the NCBI-expressed sequence tag database (EST) employing the BLAST program with the sequence of the putative potato U1A, revealed a partial sequence, EST21158, highly related to pGGSD.A13, which, notably, also contained a tryptophan residue at the fifth position of RNP1. Complete sequencing of the clone for EST21158 showed the predicted amino acid sequence to have 77% identity and 89% similarity to that of pGGSD.A13 (Figure 2). Significantly, several of the sequence distinctions between potato U2B″ and the protein encoded by pGGSD.A13, which reside in regions important for U2A′ association and U2snRNA-specific binding, were conserved between pGGSD.A13 and *Arabidopsis* EST21158.

Potato pGGSD.A13 and Arabidopsis EST21158 encode U1A proteins. The proteins encoded by potato pGGSD.A13 and *Arabidopsis* EST21158 are clearly related to both the spliceosomal proteins U1A and U2B″. We therefore examined the ability of the protein encoded by pGGSD.A13 to bind to *Xenopus* U1 and U2snRNA in the presence or absence of human U2A′ (Figure 3a). We were unable to detect any binding to *Xenopus* U2snRNA even in the presence of human U2A′. However, this protein bound specifically to *Xenopus* U1snRNA in a manner unaffected by the presence of human U2A′ (Figure 3a). To examine this further, we used the two RNAs, U1.4 and U2.4, described above. Potato GGSD.A13 protein bound specifically to U1snRNA and U2.4snRNA (Figure 3b), indicating that the RNA binding activity was specific for the U1snRNA loop II sequence, the binding site of human U1A. The protein also bound to potato U1snRNA but not potato U2snRNA (Figure 3b). This pattern of sequence-specific binding was mirrored by the *Arabidopsis* EST21158 protein, illustrating its functional relatedness (Figure 3c). The sequence conservation and RNA binding specificity of these proteins suggests that they are plant U1A homologues.

Conservation of two intron positions between human and potato genomic U1A sequences. To analyse the gene

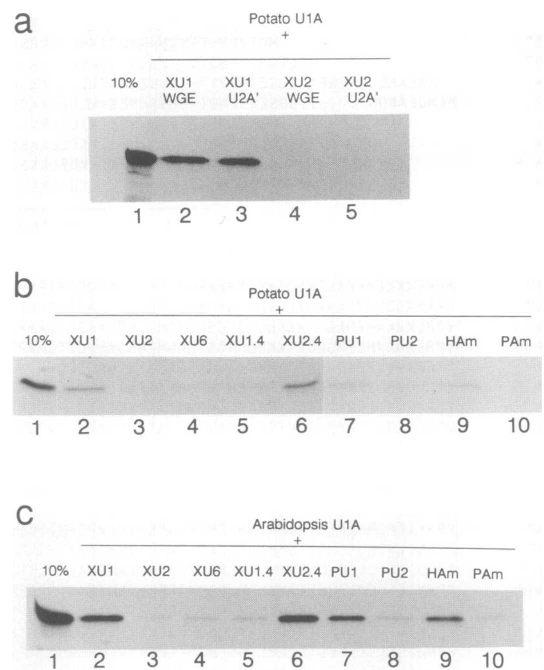

Fig. 3. RNA binding properties of potato U1A and *Arabidopsis* U1A. (**a**) Specific binding of potato U1A to *Xenopus* U1snRNA. Lane 1: 10% of input protein per binding assay. Binding of potato U1A to *Xenopus* U1snRNA (lanes 2 and 3) and *Xenopus* U2snRNA (lanes 4 and 5) in the presence of human U2A′ primed wheat germ extract (lanes 3 and 5) and control unprimed wheat germ extract (lanes 2 and 4). (**b**) Binding of potato U1A to *Xenopus* U1snRNA (lane 2), U2snRNA (lane 3), U6snRNA (lane 4), U1.4 and U2.4 loop swap transcripts (lanes 5 and 6), potato U1snRNA (lane 7), potato U2snRNA (lane 8), human U1A 3′ UTR RNA (lane 9) and potato U1A mRNA (lane 10). (**c**) Binding of *Arabidopsis* U1A to the same series of RNAs as described in (b) with the exception that the target RNA in lane 10 was *Arabidopsis* U1A mRNA

structure of potato U1A, a genomic clone was isolated and subcloned as two 4.5 kb fragments which overlap each other by 50 bp. This 9 kb cloned region contained the potato U1A-coding region organized into five exons and four introns (Figure 4) and ~2.5 kb upstream of the translation start codon and ~1.4 kb downstream of the translation stop codon. Introns I and II interrupt the coding sequence in the N-terminal RNP motif in precisely the same positions as intron I and II of the human U1A gene (Nelissen *et al.*, 1991). The other two plant U1A introns interrupt the central domain, which is unrelated in sequence to that of human U1A. The human U1A gene contains three further introns, two of which interrupt the central domain in different positions to the plant gene, and one which interrupts the C-terminal RNP motif (Figure 4). The conservation of the first two intron positions between human and plant U1A genes further underlines their relatedness. Intron conservation has also been reported for the first introns of three other RNP motif-containing proteins: tobacco 31 kDa, maize AAIP and human hnRNP A1 genes (Li *et al.*, 1991). These introns are located ~50 nt further downstream in the N-terminal RNP than those of potato and human U1A, and both examples point to the presence of introns in ancestral RNP protein genes. Estimations of gene copy number by Southern blot and by single strand conformational polymorphism analysis

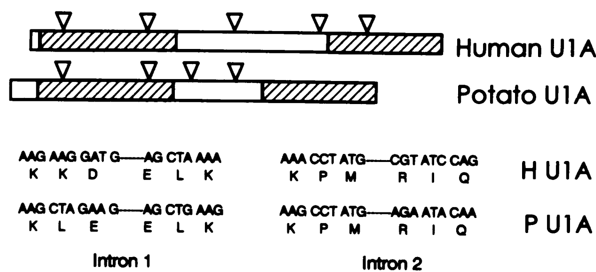

Fig. 4. Comparison of intron positions in potato and human genomic U1A. RNP motifs are hatched and central domains are unshaded. Intron positions are indicated by inverted triangles. The DNA and amino acid sequences at the sites of introns I and II are aligned.

indicate the presence of only one or possibly two potato U1A and U2B genes per haploid genome (not shown).

Taken together, the above results indicate that the proteins encoded by pGGSD.A13 and by *Arabidopsis* EST21158 encode plant homologues of human U1A protein. In summary, both proteins are highly related in sequence to known U1A proteins, both bind specifically to U1snRNA loop II sequences, and a potato genomic clone reveals precise conservation of location of two intron positions with human U1A. In addition, we have found that potato U1A is targeted to nuclei, as judged by immunofluorescent microscopy of protoplasts of *Nicotiana plumbaginifolia* in which epitope-tagged protein had been transiently expressed (A.van Dijken and G.G.Simpson, unpublished results).

U1A and polyadenylation

Neither potato nor Arabidopsis U1A proteins bind their own mRNA. Human U1A autoregulates its expression by inhibiting poladenylation of its own pre-mRNA. It binds two copies of a sequence, highly related to U1snRNA loop II, located upstream of the poly(A) signal and blocks PAP by specific protein–protein interactions (Gunderson *et al.*, 1991; Boelens *et al.*, 1993; van Gelder *et al.*, 1993). We examined whether such an autoregulatory mechanism might exist in plants by investigating the ability of both potato and *Arabidopsis* U1A to bind their own mRNAs. Although each was able to bind to the 3′ UTR of human U1A mRNA, neither protein was able to bind to RNA transcribed from their own cloned cDNAs (Figure 3b and C). Consistent with this, sequence analysis failed to reveal any U1snRNA loop II-like sequences in the 3′ UTR of either the potato (Figure 5b) or *Arabidopsis* U1A cDNAs. Although such sequences are located at a conserved distance from the poly(A) signal in vertebrates (Boelens *et al.*, 1993), the *cis*-acting elements which guide 3′ end formation and poly(A) addition in plants are distinct from vertebrates (reviewed in Hunt, 1994). Since 3′ end formation often occurs at multiple sites in the 3′ UTR of plant pre-mRNAs, we examined whether alternative sites of 3′ end formation and polyadenylation of potato U1A pre-mRNA exist and whether U1snRNA loop II-like sequences could be found further downstream of the 3′ UTR sequence found in the cDNA. A fragment of the genomic clone pGGSD.A1, consisting of 1.4 kb of sequence 3′ of the stop codon, was subcloned downstream of a CAT reporter gene in the plant expression vector R-CAT*. After transient expression in protoplasts of

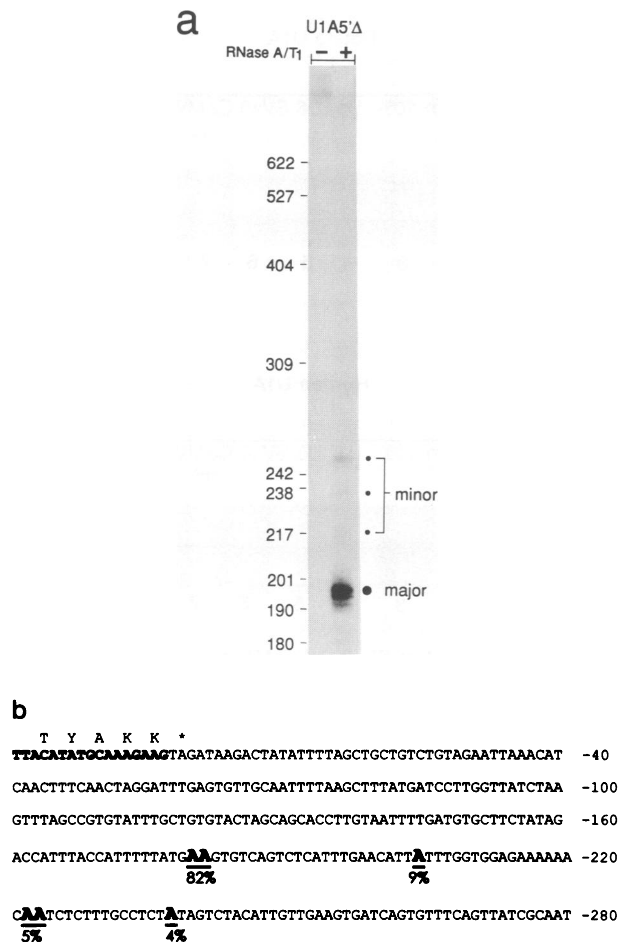

Fig. 5. Mapping of 3′ end processing sites of potato U1A. (**a**) RNase A/T1 mapping of transcripts from transiently expressed U1A*5′Δ. (**b**) Sequence of potato U1A 3′ UTR and positions of major sites of 3′ end formation.

N.plumbaginifolia, the 3′ ends of transcripts produced from this chimeric gene were analysed by RNase A/T1 mapping (Figure 5a). The major site of 3′ end formation (82% of transcripts) corresponded to the 3′ end of the U1A cDNA (Figure 5b). A further three minor sites were mapped within the following 60 nt. No further RNA processing events were detected within the entire 1.4 kb region downstream of the U1A stop codon. Even including the 60 nt of sequence covering the minor 3′ end processing sites, no sequences resembling loop II of U1snRNA were present in the 3′ UTR of the potato U1A mRNA (Figure 5b). Taken together, these results show that U1A does not autoregulate its expression in plants by the same mechanism that is used in vertebrates.

Potato U1A is unable to bind upstream efficiency elements in plant polyadenylation signals. Human U1A has been implicated in playing a positive role in polyadenylation by virtue of a report of it binding the upstream efficiency element of the SV40 late poly(A) signal (Lutz and Alwine, 1994). Extensive mutational analysis of the poly(A) signal of CaMV revealed a functional role in plants for an upstream efficiency element strikingly similar in sequence to that of the SV40 late poly(A) signal (Rothnie *et al.*, 1994). However, we failed to detect any binding of potato U1A to RNAs containing either the CaMV or SV40 late

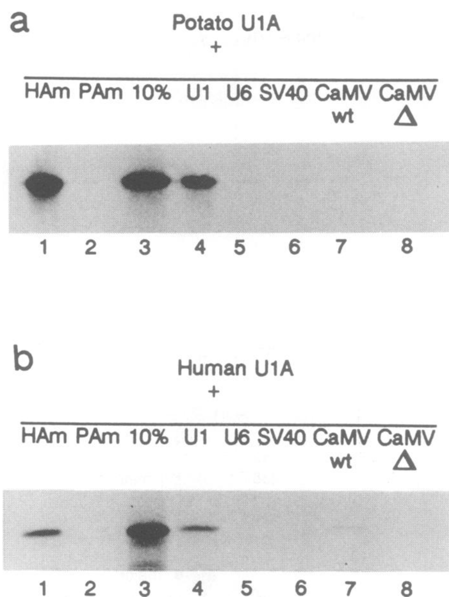

Fig. 6. Binding of U1A to SV40 and CaMV polyadenylation signal transcripts containing upstream efficency elements. (**a**) Binding of potato U1A protein: Lane 1: human U1A mRNA (full-length mRNA produced from U1A cDNA), lane 2: potato U1A mRNA; lane 3: 10% input protein; lanes 4 and 5: *Xenopus* U1snRNA and U6snRNA respectively; lanes 5 and 6: SV40 and CaMV 35S polyadenylation upstream efficiency elements, and lane 7: CaMV 35S transcript with deletion of upstream efficiency elements. (**b**) Binding of human U1A to the same series of RNAs as described in (a).

upstream efficiency elements (Figure 6a). We also failed to detect binding of human U1A to these signals even under conditions where clear binding to human U1A mRNA 3′ UTR and U1snRNA was observed (Figure 6b).

Discussion

Potato U2B″, like human U2B″, binds specifically to stem–loop IV of U2snRNA (of either vertebrate or plant origin) in a manner which is significantly enhanced by a second protein, human U2A′. The conservation of this protein–protein–RNA interaction between either human or potato U2B″, human U2A′ and U2snRNA underlines its importance in the determination of binding specificity. It seems probable that a homologue of U2A′ should also exist in higher plants and a cDNA, characterized as part of the *Arabidopsis* cDNA sequencing program, which encodes a protein with >60% amino acid sequence identity (over its entire length) to human U2A′ has been described (EMBL Accession No. X69137). We have exploited the characterization of U2B″ in two ways. Firstly, in a separate study, we have used it to describe the sub-nuclear organization of spliceosomal components in higher plants (Beven *et al.*, 1995), and here we have used it as a probe to isolate cDNA and genomic clones encoding the related spliceosomal protein, U1A from potato. We have characterized the RNA binding properties of both the potato and *Arabidopsis* U1A proteins. U1A from higher plants is highly related to vertebrate U1A and binds specifically to loop II of U1snRNA of either vertebrate or plant origin.

A striking feature of spliceosomal UsnRNAs in higher plants is the abundance of multiple, expressed sequence variants (reviewed in Solymosy and Pollák, 1993). Since some UsnRNA variants are expressed in a development-specific manner (Hanley and Schuler, 1991), there has been much speculation as to the functional significance of these variants. Variation among plant U1 and U2snRNAs in the binding sites of U1A and U2B″ (U1 loop II and U2 loop IV respectively) is, with one exception, restricted to bases in the 3′ part of each loop (Simpson *et al.*, 1991; Musci *et al.*, 1992). Our gene-copy analyses indicate that multiple copies of U1A or U2B″ genes do not exist in potato and therefore the presence of variant UsnRNPs distinct in these components seems unlikely. The plant U1A and U2B″ proteins were able to bind specifically to U1 and U2snRNA of either plant or vertebrate origin, even though the sequences at the 3′ positions of the loops were distinct. This indicates that these residues are not critical for binding activity. This is consistent with three pieces of work with the human U1A protein. First, *in vitro* selection experiments yielded the sequence AUUGCAC, which corresponds to the 5′ U1 loop II sequence, but no specific selection of the bases corresponding to the last three residues in the loop was found, suggesting that they are not specifically recognized (Tsai *et al.*, 1992). Secondly, resolution of the co-crystal structure of human U1A with U1 loop II RNA revealed that the terminal three bases of the loop form no apparent contacts with the β sheet and extend into solution (Oubridge *et al.*, 1994). Thirdly, mutation of nucleotides in the U1 loop sequence suggested that only the 5′ seven nucleotides (AUUGCAC) were important for protein contact (Hall, 1994; Stump and Hall, 1995). Although our results indicate that the variation in the sequence of the 3′ positions of U1 loop II and U2 loop IV is not critical for binding by plant U1A and U2B″ proteins, in the absence of functional *in vivo* analyses we cannot definitively rule out a role for these variants.

Human U1A and U2B″ have been studied intensely in order to characterize the mechanism of sequence-specific binding by RNP motif-containing proteins. This is because the target loop sequences of their cognate RNAs are highly related, as are the minimal regions required for sequence-specific binding by these proteins. Extensive mutational analyses have identified U2B″ residues important for U2snRNA-specific binding, but the converse mutations in U1A have not delineated as clearly the amino acids which govern its specificity of binding. For example, a mutant protein of U1A consisting of the N-terminal 102 amino acids with residues 40–49 (β2; β2–β3 loop) replaced by the corresponding residues of human U2B″ (plus other changes resulting from introduction of restriction sites) no longer bound U1snRNA, but acquired the ability to bind U2snRNA in the presence of U2A′, illustrating the importance of these residues to U2snRNA-specific binding (Scherly *et al.*, 1990a,b). The converse mutations made in U2B″, however, did not result in specific binding of the mutant protein to U1snRNA and, in fact, the mutant protein failed to bind significantly to either U1 or U2snRNA (Scherly *et al.*, 1990a). Similar results were obtained by Bentley and Keene (1991) with related mutants where residues 44–48 [β2; β2–β3 loop ≡ variable region (Scherly *et al.*, 1990; Bentley and Keene, 1991)] were exchanged. While U1A, mutated to incorporate the U2B″ residues, still bound U1snRNA and gained the ability to bind U2snRNA, the converse mutant of U2B″ containing

the U1A sequence bound more weakly than wild-type U2B″ to both U1 and U2snRNA. In other studies, a number of individual amino acid residues were found to be critical for the binding of human U1A to U1snRNA loop II, but almost all are precisely conserved in both U1A and U2B″ (Nagai *et al.*, 1990; Jessen *et al.*, 1991). The isolation and characterization of plant U1A and U2B″ allows evolutionary comparisons of these well-studied RNP motif-containing proteins to be made and it indicates the tolerance of multiple changes in a protein sequence while still maintaining functionality.

Higher plant U1A and U2B″ proteins bind U1snRNA loop II and U2snRNA stem–loop IV respectively. The sequence of the target RNA loops in higher plants closely resembles the corresponding vertebrate loop sequences and both human and plant proteins were able to distinguish U1 and U2snRNA of either plant or vertebrate origin. Such properties are important if any evolutionary comparison aimed at identifying residues which might govern specific binding is to be significant. The sequences of the respective loops in plants and vertebrates are not identical, but the ability of the proteins to distinguish U1 and U2snRNA suggests that at least some of the distinctions are not critical determinants of specificity. For example, position 7 in U1 loop II is a C, but a G is found in the corresponding position in U2 loop IV in vertebrates. This distinction is not conserved in plants: a C is found at this position in both U1 loop II and U2 loop IV. While both human and potato U2B″ can distinguish higher plant U1 from U2snRNA, the strength of binding of human U2B″ to the plant U2snRNA is much weaker. This might indicate that, although the G–C distinction plays a relatively minor role in sequence-specific binding (our results; Scherly *et al.*, 1989; Jessen *et al.*, 1990), it might contribute to the strength of the interaction of human U2B″ with vertebrate U2snRNA. The weaker binding of human U2B″ to potato U2snRNA might also be explained by the A–G distinction which is found at the beginning of loop IV between vertebrate and plant U2snRNAs. Notably, mutation of the A at the first position of human U1 loop II RNA to a G (K.Nagai, personal communication) or a C (Hall, 1994; Stump and Hall, 1995) resulted in considerably weaker binding of human U1A protein, suggesting that this residue is critical to complex formation. In addition, the vertebrate U2snRNA loop IV is flanked by two U residues, making the loop size 13 nucleotides. The smaller loop size of only 11 nucleotides in plant U2snRNAs may also affect the ability of human U2B″ to bind stably to plant U2snRNA. The only sequence distinction in the loops which is conserved between vertebrates and plants is the additional A at position 9 in loop IV, indicating that this residue is a critical determinant of sequence-specific binding.

A multiple sequence alignment of the N-terminal RNP motif of all eight available U1A and U2B″ proteins reveals extensive conservation with, for example, 58 amino acids being identical in seven or eight of the protein sequences (Figure 2). In only two positions do amino acids, which are conserved in all U1A proteins, differ from an amino acid conserved in U2B″. In position 28 (human U1A nomenclature) (α helix A), U1A contains lysine while U2B″ proteins contain arginine, and in position 41 (internal on β2) U1A proteins contain leucine and U2B″ valine. In addition, there are two other positions which are distinct between the U1A and U2B″ proteins, but at which the identity of the residue is not conserved. At position 33 (α helix A) U2B″s contain leucine while U1As contain isoleucine or valine, and at position 39 (β2) U2B″s contain histidine or arginine and U1As glutamine and lysine (Figure 2). K28 and Q39 have been mutated previously in human U1A without affecting U1snRNA binding (Nagai *et al.*, 1990; Jessen *et al.*, 1991 respectively). The remaining residues, I33 and L41 are unlikely to make critical contacts. The importance of these distinctions, therefore, may be that, in concert, they alter the conformation of the RNP motif and thus affect RNA binding specificity: L41 and I33 are in close proximity and their coupled mutation could alter the projection of β2. Since multiple and diverse contacts between the RNP motif and the cognate RNA exist (Nagai *et al.*, 1990; Jessen *et al.*, 1991; Görlach *et al.*, 1992; Howe *et al.*, 1994), sequence differences between the U1A proteins in one part of the RNP motif may be compensated for by differences either elsewhere in the motif or in the remainder of the protein (Scherly *et al.*, 1991), making interpretations from multiple sequence alignments difficult. The comparison may instead illustrate the range of solutions which the plasticity of the RNP motif exhibits in executing specificity of recognition. It will be interesting to study the effect of mutation of these distinct residues on the binding of U2B″ to U1snRNA.

Although the N-terminal RNP motifs of the proteins described here are highly related, one distinction is notable: potato U2B″, and *Arabidopsis* and potato U1A each possess a tryptophan residue at the fifth position of RNP1 in the N-terminal RNP motif. These are the only known RNP motif-containing proteins that contain a tryptophan residue at this position. This, the most conserved residue of the RNP motif, is normally occupied by phenylalanine (Birney *et al.*, 1993). The interchelation of RNP1 and RNP2 aromatic residues with bases in the cognate RNA is probably a general feature of RNP motif–RNA interactions (Kenan *et al.*, 1991; Oubridge *et al.*, 1994; Stump and Hall, 1995). Indeed, the co-crystal structure of human U1A with U1 loop II RNA revealed such an interaction between Phe56 of RNP1 and target RNA bases (Oubridge *et al.*, 1994). In view of this property, the extreme rarity of tryptophan residues in RNP1 is striking, particularly since tryptophan might provide a more extensive interaction with purines in such ring stacking interactions. The presence of tryptophan in position 5 of RNP1 in these plant proteins is clearly indicative of the ability of this amino acid to function in this context, and human U1A with Phe56 mutated to tryptophan retains the ability to bind U1snRNA (G.G.Simpson, C.Oubridge, K.Nagai and J.W.S.Brown, unpublished results). Although the tryptophan in this position of RNP1 is unique to these plant proteins, tryptophan is found in approximately one-third of all known RNP motifs (M.Cusick, personal communication) and is known to occur in strands of the RNA-interacting surface of some RNP motif-containing proteins (Kenan *et al.*, 1991; Birney *et al.*, 1993), notably in the most conserved position of RNP2 in yeast PRP24 (Shannon and Guthrie, 1991) and in position 3 of RNP1 of the third RNP motif of yeast PES4 (EMBL Accession No. D26442) (M.Cusick, personal communication). A second notable

feature of potato U2B″ is the presence of a proline residue in the β2–β3 loop of its N-terminal RNP motif. Prolines are present in the analogous position in the C-terminal RNP motif of human U1A (and U2B″) which does not appear to bind RNA (Scherly *et al.*, 1989; Lu and Hall, 1995). In U1A–U1snRNA recognition, the β2–β3 loop protrudes through the RNA loop; the proline in the potato U2B″ β2–β3 loop probably restricts its flexibility and therefore may be important in U1snRNA rejection by U2B″.

In addition to the relatedness of the N-terminal RNP motifs of these proteins, even greater conservation of the C-terminal RNP motifs was observed. For example, the C-terminal motif of potato U1A is 93% similar to that of potato U2B″. The C-terminal RNP motif is not required for sequence-specific binding by either U1A or U2B″ and its function is unknown. The extensive conservation of this motif, however, indicates a related function in both proteins. The least similar part of the proteins is the central domain which separates the two RNP motifs. These domains function as a nuclear localization signal (NLS) in both human U1A and U2B″ proteins (Kambach and Mattaj, 1992, 1994) and, as such, they represent quite novel NLSs, in terms of their large size and complexity. Notably, both human U1A and U2B″ shuttle repeatedly between the nucleus and cytoplasm (Kambach and Mattaj, 1992, 1994). This region in the plant proteins is unrelated to the human proteins and quite unrelated between the plant proteins. However, each plant protein contains a conserved stretch of basic amino acids [consensus: $(R/K)_4$ $(X)_{4-5}$ $(R/K)_4$] which might function in nuclear targeting, and experiments to address this are in progress.

Human U1A autoregulates its expression by regulating polyadenylation of its own pre-mRNA. It binds two copies of a loop II-like sequence in the 3′ UTR of its own mRNA which are found at a conserved position relative to the poly(A) signal in vertebrates (Boelens *et al.*, 1993). No U1snRNA loop II sequences are present in the 3′ UTR of either the potato or *Arabidopsis* U1A genes, and neither protein was able to bind to its own mRNA. U1snRNA loop II sequences were also absent in the 3′ UTR of the *Drosophila* U1A mRNA (Harper *et al.*, 1992; Flickinger and Salz, 1994), indicating that this mechanism of autoregulation of U1A may not be a general phenomenon, although conserved in vertebrates. Differences in the mechanism of polyadenylation between vertebrates, *Drosophila* and higher plants may explain why the same form of autoregulation is not found in the latter organisms. Either the same kind of interactions may not be possible between the plant proteins, or, if established, may not result in effective inhibition of polyadenylation, or U1A may have a normal function in polyadenylation in vertebrates, but not other organisms, and this has been exploited in the evolution of the autoregulatory mechanism (Gunderson *et al.*, 1994). The significance of vertebrate U1A autoregulation is not yet known and it may be that either plants do not need to tightly regulate free U1A protein or that regulation is achieved by a different mechanism. The isolation of cDNA and genomic clones, together with the ease of genetic transformation of plants, make a study of the regulation of U1A in higher plants readily feasible.

A possible role for vertebrate U1A in polyadenylation has been suggested from the reported binding of U1A to upstream sequences in the SV40 late poly(A) signal (Lutz and Alwine, 1994). Since similar sequences are functionally required in the CaMV poly(A) signal in plants (Rothnie *et al.*, 1994), and the SV40 late poly(A) signal is functional in plants (H.M.Rothnie, unpublished results), we tested whether plant U1A could bind such sequences. Under conditions where binding to U1snRNA and human U1A mRNA was detected, no binding of plant U1A to RNAs containing all known SV40 or CaMV poly(A) signals was observed and thus plant U1A does not appear to be involved in the activation of polyadenylation. In contrast to the results of Lutz and Alwine (1994), binding of human U1A to SV40 late poly(A) signal-containing RNA was not observed, even though binding to U1snRNA and U1A mRNA was readily detected. However, consistent with the lack of U1A binding to SV40 poly(A) sequences here, Lu and Hall (1995) recently reported the inability of the C-terminal RNP motif of U1A to bind an RNA containing the proposed SV40 target sequence (Lutz and Alwine, 1994) even under conditions where the N-terminal RNP motif bound RNA non-specifically.

In this paper, we have reported the characterization of plant U1A and U2B″ spliceosomal proteins. As the only full-length, characterized plant spliceosomal proteins to be described, they provide a useful handle on the analysis of pre-mRNA processing in higher plants. The isolation of clones encoding these proteins from such an evolutionarily distinct source should yield insight into the mechanism of RNA-specific binding by RNP motif-containing proteins and the mechanism of autoregulation by vertebrate U1A.

Materials and methods

Oligonucleotides

All oligonucleotides were synthesized at the Scottish Crop Research Institute. Oligo#1, 5′ aggctcctgctcctccaaata 3′; oligo#2, 5′ ctacttctttgcatatgtaatcaa 3′; oligo#3, 5′ attaggatcccatatgatgcttacaggagac 3′; oligo#4, 5′ ggcaagcttcactagaattaagaatgc 3′; oligo#5, 5′ ttaaagctttcaataggtcccctctgcctt 3′; oligo#6, 5′ gcaagcttgtggggtgcgccaggcccgtgc 3′; oligo#7, 5′ gggaattcatacctttctcggccttttgg 3′.

Library screening and sequence analysis

Isolation of a genomic clone encoding potato U1A. A *Solanum tuberosum* cv. Saturna genomic library in λEMBL 4 was screened with a 780 bp *Bfr*I–*Eco*RI restriction fragment of pPotB2-1 (Simpson *et al.*, 1991) at moderate stringency by standard procedures (Sambrook *et al.*, 1989). A positive phage was plaque purified and contained an insert of ~18 kb. An ~4.5 kb *Eco*RI fragment and an overlapping ~4.5 kb *Pst*I fragment (50 bp overlap) were subcloned separately into pUC13 and pGEM3Zf(+) (Promega) respectively, generating pGGSD.A1 and pGGSD.A3 respectively, and sequenced.

Isolation of a cDNA encoding potato U1A. A *S.tuberosum* cv. Cara leaf cDNA library in λgt11 (Hedley *et al.*, 1994) was screened according to standard procedures with a 248 bp probe generated by PCR amplification of pGGSD.A1 with oligos #1 and #2. A positive phage was plaque purified and the insert (999 bp) was subcloned into pUC13, generating pGGSD.A13 and sequenced.

Plasmids

For *in vitro* transcription of potato U2B″ RNA, pPU2B″FL, a plasmid containing the full-length potato U2B″-coding sequence, was prepared by PCR amplification of the original cDNA clone, pPotB2-1 (Simpson *et al.*, 1991) with oligos #3 and #4. Discrete amplified DNA products were purified on agarose gels, treated with Klenow and cloned into the *Sma*I site of pGEM3Zf(+). The plasmid pPU2B″NT, containing a sequence coding for the first 103 amino acids of potato U2B″, was prepared similarly, except that PCR amplification was performed with

oligos #3 and #5. In both cases, PCR amplification was performed with *Pfu* DNA polymerase (Stratagene) and the cloned amplified products were verified by complete sequencing. For *in vitro* transcription of potato U1A RNA, pPU1A, a plasmid containing the full-length potato U1A-coding sequence, was prepared by subcloning the full-length cDNA as an *Eco*R1 fragment from pGGSDA.13 into the corresponding site in pGEM3Zf(+). pAU1A, containing the *Arabidopsis* U1A cDNA in pZipLoc2, was obtained as EST21158 from the MSU-DOE Plant Research Laboratory, Michigan State University, USA. The cDNA insert was subcloned into pGEM3Zf(+) and sequenced completely. Constructs encoding human U2B″ and U2A′ have been described previously (Boelens *et al.*, 1991a).

pPU2, a plasmid containing a potato U2snRNA gene sequence for *in vitro* transcription, was prepared by PCR amplification of gU2-22 (Waugh *et al.*, 1991) with oligos #6 and #7. Discrete amplified DNA fragments were purified on 8% non-denaturing polyacrylamide gels, digested with *Eco*RI and *Hin*dIII and subcloned into the corresponding sites of pGEM3Zf(+). pPU1, a construct used for the production of potato U1snRNA, has been described (Beven *et al.*, 1995). Plasmids containing the *Xenopus* U1, U2 and U6snRNA sequences and the mutant constructs, U1.4 and U2.4 used for the production of RNA target transcripts have been described previously (Hamm *et al.*, 1989; Scherly *et al.*, 1989, 1990a,b).

Full-length human U1A mRNA was produced from a plasmid containing the human U1A cDNA (a gift from I.Mattaj). Human U1A 3′ UTR RNA was produced from pAmWT which contains human U1A sequences from positions VI 842 to VI 897 (Nelissen *et al.*, 1991) cloned into the *Eco*RI and *Hin*dIII sites of pGEM3Zf(+). A construct for transcription of SV40 late poly(A) signal-containing RNA was generated from a plasmid containing SV40 expression signals (pSVmT; Muser *et al.*, 1989). A 196 nt *Bam*HI–*Dra*I fragment containing all the known elements of the SV40 late poly(A) signal was subcloned into *Bam*HI–*Sma*I-digested pGEM1, resulting in plasmid pGSV40pA from which sense transcripts covering the SV40 late poly(A) signal were generated. To construct templates for CaMV poly(A) signal transcripts, *Eco*RI–*Pst*I fragments containing the CaMV and nopaline synthase (*nos*) gene poly(A) signals from plasmids R-CAT* and Δ53–32 (Rothnie *et al.*, 1994) were subcloned into the corresponding sites in pGEM2. *In vitro* transcription with T7 RNA polymerase generated transcripts corresponding to the wild-type CaMV poly(A) signal and a mutant derivative in which the tandemly repeated UUUGUA motif at positions –53 to –32 is deleted.

In vitro transcription/translation and protein–RNA binding assay

In vitro transcription of biotinylated RNAs was performed as described in Boelens *et al.* (1991a), except that biotin-21-UTP (Clontech) was used. *In vitro* transcription of mRNA transcripts and *in vitro* translation in wheat germ extract were carried out as described previously (Boelens *et al.*, 1991a). Protein–RNA binding assays were performed according to the method established by Scherly *et al.* (1989) with modifications (Boelens *et al.*, 1991a). Specifically, RNA binding was carried out in KHN buffer (150 mM KCl/20 mM HEPES-KOH, pH 7.9/0.01% NP-40) with 20 ng biotinylated U1 or U2snRNA derivatives and 2 μg of *Escherichia coli* tRNA. Following incubation, recovery on streptavidin-agarose beads and SDS-polyacrylamide gel electrophoresis, the gel was fixed in 5% v/v methanol, 7.5% v/v acetic acid and treated with Amplify (Amersham) prior to drying and exposure to X-ray film. Quantitative estimations of binding enhancement were carried out on autoradiographs using a BioImager.

Mapping of the potato U1A poly(A) site

An *Nde*I–*Eco*RI fragment covering the last four codons of the ORF and ~1.35 kb of downstream sequence from the potato U1A gene was introduced in place of the CaMV ‘R’ region in plasmid R-CAT* (Rothnie *et al.*, 1994). In the resulting plasmid (U1A*5′Δ), the chloramphenicol acetyltransferase (CAT) gene under the control of the CaMV 35S promoter is followed by the U1A 3′ end fragment and then a second poly(A) site from the *nos* gene. This latter poly(A) site acts as a ‘trap’ for any transcripts not processed within the U1A sequence.

Nicotiana plumbaginifolia leaf protoplasts were transfected with 5 μg plasmid DNA as described by Goodall *et al.* (1990). Total RNA was isolated from the protoplasts 6 h after transfection and subjected to RNase A/T1 protection analysis according to published protocols (Goodall *et al.*, 1990). Radioactively labelled antisense RNA probes were synthesized by *in vitro* transcription in the presence of *Pst*I-linearized U1A*5′Δ as a template. Protected fragments were resolved on 6% polyacrylamide denaturing gels and visualized by autoradiography. The positions of the processing sites in the U1A sequence were deduced by comparing the sizes of protected fragments with radioactively labelled single-stranded DNA size markers (pBR322/*Hpa*II). Fragments corresponding to transcripts processed within the *nos* region were identified on the basis of their size. Radioactivity in the protected fragments was quantified using a PhosphoImager (Molecular Dynamics). The percentage of transcripts processed at each site was calculated, taking into account the number of labelled nucleotides in each fragment. Processing efficiencies expressed as percentages represent the mean values from at least three separate transfections, rounded to the nearest whole percent.

Acknowledgements

We thank the Agriculture and Food Research Council (now the Biotechnology and Biological Sciences Research Council) and the Scottish Office Agriculture and Fisheries Department for financial support. We are also grateful to Kiyoshi Nagai, Chris Oubridge, Craig Simpson and Robbie Waugh. We thank Mike Cusick for information from his RNP motif database. G.G.S. was supported also by a short-term EMBO fellowship and thanks Professor Jean Beggs and Susan Anderson for their help at Edinburgh University.

References

Bandziulis,R.J., Swanson,M.S. and Dreyfuss,G. (1989) RNA binding proteins as developmental regulators. *Genes Dev.*, **3**, 431–437.

Bennet,M.M., Baron,M.A. and Craft,J. (1993) Nucleotide sequence analysis of the A-protein of the small nuclear ribonucleoprotein particle—the murine protein contains a 5′ amino terminal TAG. *Nucleic Acids Res.*, **21**, 4404.

Bentley,R.C. and Keene,J.D. (1991) Recognition of U1 and U2 small nuclear RNAs can be altered by a 5-amino-acid segment in the U2 small nuclear ribonucleoprotein particle (snRNP) B″ protein and through interactions with U2snRNP-A′ protein. *Mol. Cell. Biol.*, **11**, 1829–1839.

Beven,A.F., Simpson,G.G., Brown,J.W.S. and Shaw,P.J. (1995) The organization of spliceosomal components in the nuclei of higher plants. *J. Cell Sci.*, **108**, 509–518.

Birney,E., Kumar,S. and Krainer,A. (1993) Analysis of the RNA-recognition motif and RS and RGG domains—conservation in metazoan pre-messenger RNA splicing factors. *Nucleic Acids Res.*, **21**, 5803–5816.

Boelens,W., Scherly,D., Beijer,R.P., Jansen,E.J.R., Dathan,N.A., Mattaj,I.W. and van Venrooij,W. (1991a) A weak interaction between the U2A′ protein and U2snRNA helps to stabilise their complex with the U2B″ protein. *Nucleic Acids Res.*, **19**, 455–460.

Boelens,W., Scherly,D., Jansen,E.J.R., Kolen,K., Mattaj,I.W. and van Venrooij,W. (1991b) Analysis of *in vitro* binding of U1-A protein mutants to U1snRNA. *Nucleic Acids Res.*, **19**, 4611–4618.

Boelens,W.C., Jansen,E.J.R., van Venrooij,W.J., Stripecke,I.R., Mattaj,I.W. and Gunderson,S.I. (1993) The human U1snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell*, **72**, 881–892.

Dean,C., Tamaki,S., Dunsmuir,P., Favreau,M., Katayama,C., Dooner,H. and Bedrook,J. (1986) mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. *Nucleic Acids Res.*, **14**, 2229–2240.

Filipowicz,W., Gniadkowski,M., Klahre,U. and Liu,H.-X. (1994) Pre-mRNA splicing in plants. In Lamond,A.I. (ed.), *Pre-mRNA Processing*. R.G.Landes Company Publishers, pp. 65–77.

Flickinger,T.W. and Salz,H.K. (1994) The *Drosophila* sex determination gene, *SNF*, encodes a nuclear protein with sequence and functional similarity to the mammalian U1A snRNP protein. *Genes Dev.*, **8**, 914–924.

Fresco,L.D., Harper,D.S. and Keene,J.D. (1991) Leucine periodicity of U2 small nuclear ribonucleoprotein particle (snRNP) A′ protein is implicated in snRNP assembly via protein–protein interactions. *Mol. Cell. Biol.*, **11**, 1578–1589.

Goodall,G.J. and Filipowicz,W. (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell*, **58**, 473–483.

Goodall,G.J., Wiebauer,K. and Filipowicz,W. (1990) Analysis of premessenger RNA processing in transfected plant protoplasts. *Methods Enzymol.*, **181**, 148–161.

Görlach,M., Wittekind,M., Beckman,R.A., Mueller,L. and Dreyfuss,G.

(1992) Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. *EMBO J.*, **11**, 3289–3295.

Gunderson,S.I., Beyer,K., Martin,G., Keller,W., Boelens,W.C. and Mattaj,I.W. (1994) The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell*, **76**, 531–541.

Habets,W.J., Sillekens,P.T.G., Hoet,M.H., Schalken,J.A., Roebroek, A.J.M., Leunissen,J.A.M., van de Ven,W.J.M. and van Venrooij,W.J.. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 2421–2425.

Hall,K.B. (1994) Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry*, **33**, 10076–10088.

Hamm,J., Dathan,N.A. and Mattaj,I.W. (1989) Functional analysis of mutant *Xenopus* U2snRNAs. *Cell*, **59**, 159–169.

Hanley,B.A. and Schuler,M.A. (1991) Developmental expression of plant snRNAs. *Nucleic Acids Res.*, **19**, 6319–6325.

Harper,D.S., Fresco,L.D. and Keene,J.D. (1992) RNA binding specificity of a *Drosophila* snRNP protein that shares sequence homology with mammalian U1-A and U2-B″ proteins. *Nucleic Acids Res.*, **20**, 3645–3650.

Hedley,P.E., Machray,G.C., Davies,H.V., Burch,L. and Waugh,R. (1994) Potato (*Solanum tuberosum*) invertase-encoding cDNAs and their differential expression. *Gene*, **145**, 211–214.

Howe,P.W.A., Nagai,K., Neuhaus,D. and Varani,G. (1994) NMR studies of U1snRNA recognition by the N-terminal RNP domain of the human U1A protein. *EMBO J.*, **13**, 3873–3881.

Hunt,A.G. (1994) Messenger RNA 3′ end formation in plants. *Plant Mol. Biol.*, **45**, 47–60.

Hunt,A.G., Chu,N.M., Odell,J.T., Nagy,F. and Chua,N.-H. (1987) Plant cells do not properly recognise animal gene polyadenylation signals. *Plant Mol. Biol.*, **8**, 23–35.

Jessen,T.-H., Oubridge,C., Teo,C.H., Prichard,C. and Nagai,K. (1991) Identification of molecular contacts between the U1A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.*, **10**, 3447–3456.

Joshi,C.P. (1987) Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. *Nucleic Acids Res.*, **15**, 9627–9640.

Kambach,C. and Mattaj,I.W. (1992) Intracellular distribution of the U1A protein depends on active transport and nuclear binding to U1snRNA. *J. Cell Biol.*, **118**, 11–21.

Kambach,C. and Mattaj,I.W. (1994) Nuclear transport of the U2 snRNP-specific U2B″ protein is mediated by both direct and indirect signalling mechanisms. *J. Cell Sci.*, **107**, 1807–1816.

Keene,J.D. and Query,C.C. (1991) Nuclear RNA-binding proteins. *Nucleic Acids Res. Mol. Biol.*, **41**, 179–202.

Kenan,D.J., Query,C.C. and Keene,J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.*, **16**, 214–220.

Li,Y., Ye,L., Sugita,M. and Suguira,M. (1991) Tobacco nuclear gene for the 31 kDa chloroplast ribonucleoprotein: genomic organisation, sequence analysis and expression. *Nucleic Acids Res.*, **19**, 2987–2991.

Liao,X.C., Tang,J. and Rosbash,M. (1993) An enhancer screen identifies a gene that encodes the yeast U1snRNP A protein: implications for snRNP protein function in pre-mRNA splicing. *Genes Dev.*, **7**, 419–428.

Lu,J. and Hall,K.B. (1995) An RBD that does not bind RNA: NMR secondary structure determination and biochemical properties of the C-terminal RNA binding domain from the human U1A protein. *J. Mol. Biol.*, **247**, 739–752.

Lutz,C.S. and Alwine,J.E. (1994) Direct interaction of the U1snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. *Genes Dev.*, **8**, 576–586.

Lutz-Freyermuth,C., Query,C.C. and Keene,J.D. (1990) Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem–loop II of U1 RNA. *Proc. Natl Acad. Sci.*, **87**, 6393–6397.

Mattaj,I.W. (1993) RNA recognition: a family matter? *Cell*, **73**, 837–840.

Moore,M.J., Query,C.C. and Sharp,P.A. (1993) In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.

Musci,M.A., Egeland,D.B. and Schuler,M.A. (1992) Molecular comparison of monocot. and dicot. U1 and U2 snRNAs. *Plant J.*, **2**, 589–599.

Muser,J., Kaech,S., Moroni,C. and Ballmer-Hofer,K. (1989) Stimulation of pp60^{c-src} kinase activity in FDC-P1 cells by polyoma middle-T antigen and haematopoietic growth factors. *Oncogene*, **4**, 1433–1439.

Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.

Nelissen,R.L.H., Sillekens,P.T.G., Beijer,R.P., Guerts van Kessel,A.H.M. and van Venrooij,W.J. (1991) Structure, chromosomal localisation and evolutionary conservation of the gene encoding human U1snRNP-specific A protein. *Gene*, **102**, 189–196.

Oubridge,C., Ho,N., Evans,P.R., Teo,C.-H. and Nagai,K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.

Rothnie,H.M., Reid,J. and Hohn,T. (1994) The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of mRNA 3′-end formation in plants. *EMBO J.*, **13**, 2200–2210.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanfaçon,H. (1994) Analysis of figwort mosaic virus (plant pararetrovirus) polyadenylation signal. *Virology*, **198**, 39–49.

Scherly,D., Boelens,W., van Venrooij,W.J., Dathan,N.A., Hamm,J. and Mattaj,I.W. (1989) Identification of the RNA binding segment of human U1A protein and definition of its binding site on U1snRNA. *EMBO J.*, **8**, 4163–4170.

Scherly,D., Boelens,W., Dathan,N.A., van Venrooij,W.J. and Mattaj,I.W. (1990a) Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B″ and their cognate RNAs. *Nature*, **345**, 502–506.

Scherly,D., Boelens,W., Dathan,N.A., van Venrooij,W.J. and Mattaj,I.W. (1990b) The U2B″ RNP motif as a site of protein–protein interaction. *EMBO J.*, **9**, 3675–3681.

Scherly,D., Kambach,C., Boelens,W., van Venrooij,W.J. and Mattaj,I.W. (1991) Conserved amino acid residues within and outside of the amino-terminal ribonucleoprotein motif of U1A small nuclear ribonucleoprotein involved in U1 RNA-binding. *J. Mol. Biol.*, **219**, 577–584.

Shannon,K.W. and Guthrie,C. (1991) Suppressors of a U4snRNA mutation define a novel U6snRNP protein with RNA-binding motifs. *Genes Dev.*, **5**, 773–785.

Sillekens,P.T.G., Habets,W.J., Beijer,R.P. and van Venrooij,W.J. (1987) cDNA cloning of the human U1snRNA-associated A protein: extensive homology between U1 and U2 snRNP-specific proteins. *EMBO J.*, **6**, 3841–3848.

Simpson,G.G., Vaux,P., Clark,G., Waugh,R., Beggs,J.D. and Brown,J.W.S. (1991) Evolutionary conservation of the spliceosomal protein, U2B″. *Nucleic Acids Res.*, **19**, 5213–5217.

Solymosy,F. and Pollák,T. (1993) Uridylate-rich small nuclear RNAs (UsnRNAs), their genes and pseudogenes, and UsnRNPs in plants: structure and function. A comparative approach. *Crit. Rev. Plant Sci.*, **12**, 275–369.

Stump,W.T. and Hall,K.B. (1995) Crosslinking of an iodo-uridine-RNA hairpin to a single site on the human U1A N-terminal RNA binding domain. *RNA*, **1**, 55–63.

Tsai,D.E., Harper,D.S. and Keene,J.D. (1992) U1snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. *Nucleic Acids Res.*, **19**, 4931–4936.

van Gelder,C.W.G., Gunderson,S.I., Jansen,E.J.R., Boelens,W.C., Polycarpou-Schwartz,M., Mattaj,I.W. and van Venrooij,W.J. (1993) A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation. *EMBO J.*, **12**, 5191–5200.

Wahle,E. and Keller,W. (1992) The biochemistry of 3′ end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.*, **61**, 419–440.

Waugh,R., Clark,G., Vaux,P. and Brown,J.W.S. (1991) Sequence and expression of potato U2snRNA genes. *Nucleic Acids Res.*, **19**, 249–256.

Received on April 18, 1995; revised on June 26, 1995