The EMBO Journal Vol.16 No.14 pp.4433–4440, 1997

Trans mRNA splicing in trypanosomes: cloning and analysis of a *PRP8*-homologous gene from *Trypanosoma brucei* provides evidence for a U5-analogous RNP

Stephan Lücke, Thomas Klöckner[1], Zsofia Palfi, Michael Boshart[1] and Albrecht Bindereif[2]

Institut für Biochemie, Humboldt-Universität/Charité, Monbijou-Strasse 2a, D-10117 Berlin and [1]Max-Planck-Institut für Biochemie, Genzentrum, Am Klopferspitz 18a, D-82152 Martinsried, Germany

[2]Corresponding author

S.Lücke, T.Klöckner and Z.Palfi contributed equally to this work

In trypanosomes all mRNAs are generated through *trans* mRNA splicing, requiring the functions of the small nuclear RNAs U2, U4 and U6. In the absence of conventional *cis* mRNA splicing, the structure and function of a U5-analogous snRNP in trypanosomes has remained an open question. In *cis* splicing, a U5 snRNP-specific protein component called PRP8 in yeast and p220 in man is a highly conserved, essential splicing factor involved in splice-site recognition and selection. We have cloned and sequenced a genomic region from *Trypanosoma brucei*, that contains a PRP8/p220-homologous gene (p277) coding for a 277 kDa protein. Using an antibody against a C-terminal region of the trypanosomal p277 protein, a small RNA of ~65 nucleotides could be specifically co-immunoprecipitated that appears to be identical with a U5 RNA (SLA2 RNA) recently identified by Dungan *et al.* (1996). Based on sedimentation, immunoprecipitation and Western blot analyses we conclude that this RNA is part of a stable ribonucleoprotein (RNP) complex and associated not only with the p277 protein, but also with the common proteins present in the other *trans*-spliceosomal snRNPs. Together these results demonstrate that a U5-analogous RNP exists in trypanosomes and suggest that basic functions of the U5 snRNP are conserved between *cis* and *trans* splicing.

Keywords: snRNA/snRNP/*trans* splicing/trypanosomes/U5

Introduction

Trans splicing is an essential step in the expression of all protein-coding genes in trypanosomes and involves the addition of a short, non-coding miniexon derived from the spliced leader (SL) RNA onto each protein-coding exon sequence present within long polycistronic precursor transcripts. Besides trypanosomes, this SL-type *trans* splicing occurs also in nematodes, trematodes and *Euglena* (reviewed by Agabian, 1990; Nilsen, 1993, 1995). Although discovered in trypanosomes 10 years ago (Murphy *et al.*, 1986; Sutton and Boothroyd, 1986), *trans* mRNA splicing is still not well understood in comparison to conventional *cis* mRNA splicing. In analogy to *cis* splicing, we assume that *trans* splicing requires the assembly of the two precursor RNAs, the SL RNA and the polycistronic pre-mRNA, into a complex RNA–protein structure, a *trans* spliceosome. Only a few RNA and protein components have been identified in trypanosomes that are relevant for *trans* splicing. In particular, U2, U4 and U6 snRNAs are essential co-factors (Tschudi and Ullu, 1990) and occur in the form of snRNPs in the cell (Michaeli *et al.*, 1990; Cross *et al.*, 1991; Günzl *et al.*, 1992). The SL RNP is an additional, *trans*-splicing specific factor and has been proposed as a U1 snRNP equivalent (Bruzik *et al.*, 1988; discussed in Steitz, 1992); although it contributes its miniexon during the *trans*-splicing reaction, it also resembles an snRNP on the basis of its protein composition (Palfi *et al.*, 1991). Using affinity purification of *Trypanosoma brucei* snRNPs, five common low-molecular-weight polypeptides have been identified that are shared by the U2, U4/U6 and SL RNPs; in addition, this analysis revealed several snRNP-specific protein components (Palfi *et al.*, 1991). In contrast to the yeast and mammalian *cis*-splicing systems, only one snRNP protein gene has been cloned so far from the trypanosome *trans*-splicing system, coding for a U2-specific 40 kDa protein with homology to the *cis*-spliceosomal U2 A′ protein (Cross *et al.*, 1993).

An important open question has been whether there is a *trans*-spliceosomal homologue of the U5 RNA. U5 RNA is the least conserved among the spliceosomal RNAs, and phylogenetic comparisons showed that essentially only the 11-nucleotide 5′ loop sequence is conserved (Guthrie and Patterson, 1988; Frank *et al.*, 1994). The 5′ loop of U5 RNA is functionally important in *cis* splicing, in particular during 5′ and 3′ splice-site recognition and selection; this has been established in yeast using *in vivo* mutational studies (Newman and Norman, 1991, 1992) and also in the mammalian splicing system (Wyatt *et al.*, 1992; Cortes *et al.*, 1993; Sontheimer and Steitz, 1993). A recent *in vitro* snRNP reconstitution study in yeast, however, demonstrated that the U5 loop sequence is not essential for the first step of splicing, but critical for holding the splicing intermediates together (O'Keefe *et al.*, 1996). The latter study pointed to an important role for U5-associated proteins.

The U5 snRNP-specific protein PRP8, which appears to function in close collaboration with the U5 RNA, has first been cloned in yeast (PRP8; Jackson *et al.*, 1988); it has also been identified in man (p220; Anderson *et al.*, 1989) and in other species examined (*Caenorhabditis elegans*, rice, *Arabidopsis*, *Plasmodium*; reviewed in Hodges *et al.*, 1995). This largest known protein in the spliceosome is evolutionarily highly conserved (Hodges *et al.*, 1995) and is required for spliceosome assembly before the first splicing step (Jackson *et al.*, 1988; Brown

and Beggs, 1992), and for uridine-tract recognition and fidelity of 3′ splice-site utilization (Umen and Guthrie, 1995a,b, 1996). Consistent with a central role during splicing, the PRP8/p220 protein makes direct contact with pre-mRNA (Whittacker and Beggs, 1991; Wyatt *et al.*, 1992; Staknis and Reed, 1994; Teigelkamp *et al.*, 1995a,b; Reyes *et al.*, 1996).

Based on the functional correlation between the U5 RNA and the PRP8/p220 protein and the importance of these two components of the *cis* spliceosome, it would in fact be very surprising if there were no equivalents in the *trans* spliceosome. Very recently Dungan *et al.* (1996) reported the RNA sequence of a 62-nucleotide 'U5-like RNA' from *T.brucei* containing the 5′ loop sequence conserved in all known spliceosomal U5 RNAs. Additional evidence for this RNA being the *trans*-spliceosomal U5 counterpart was derived from crosslinking experiments, showing a contact between the U5 loop sequence and the free SL miniexon *trans*-splicing intermediate, an interaction analogous to that between the *cis*-spliceosomal U5 RNA loop and the free first-exon splicing intermediate. In nematodes, where both *cis* and *trans* splicing occur, a recent *in vitro* study demonstrated that U5 RNA is present in *cis* as well as in *trans* spliceosomes (Maroney *et al.*, 1996); in *trans* splicing one role of U5 appears to be to hold onto the free SL exon after the first step of the reaction.

Our approach to address the question of U5 snRNP function was to identify a PRP8/p220 homologous gene in trypanosomes and to search for an associated RNA, which would represent the bona fide U5-analogous RNA in the *trans*-splicing system. We report here the sequence of a *PRP8*-homologous gene of *T.brucei* coding for a 277 kDa protein; using specific antibodies against the trypanosomal protein, we have identified a small RNA of ~65 nucleotides that appears to be identical to the U5-like RNA identified by Dungan *et al.* (1996). We demonstrate that the trypanosome U5 RNA is part of a stable ribonucleoprotein (RNP) complex and associated not only with the p277 protein, but also with the common proteins present in the other *trans*-spliceosomal snRNPs. In sum, a U5-analogous RNP exists in trypanosomes suggesting that basic functions of the U5 snRNP are conserved between *cis* and *trans* splicing.

Results and discussion

Genomic cloning and expression of a PRP8-homologous gene (p277) from T.brucei

In a polycistronic transcription unit encoding two catalytic subunits of the *T.brucei* protein kinase A (T.Klöckner and M.Boshart, manuscript in preparation) an open reading frame homologous to the C-terminal part of the *PRP8* gene from *Saccharomyces cerevisiae* was found by database searches. Using this region as a probe, several phage clones were isolated and mapped from a genomic *T.brucei* λDASH II library (see Materials and methods). Sequencing of ~8 kb of a λ clone revealed a long open reading frame coding for a PRP8-homologous protein of 2403 amino acids with a predicted molecular weight of 276 921 and pI of 8.35 (designated p277 in the following; see accession number Y12638 for genomic nucleotide sequence and data not shown). Southern blot hybridization analysis of

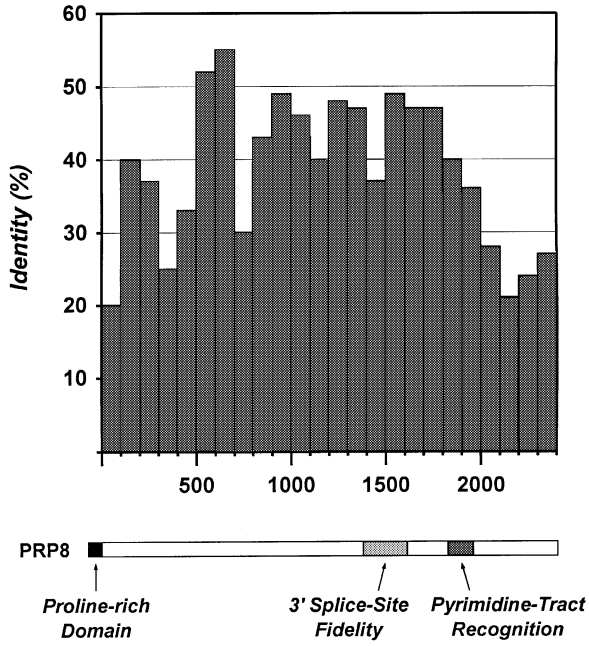

Fig. 1. Graphic representation of the sequence comparison between the *T.brucei* p277 protein and the *S.cerevisiae* PRP8 protein. The percentage of identical amino acids is given for blocks of 100 amino acids each across the entire protein (numbering referring to the *T.brucei* p277 sequence). For comparison, the proposed domain structure of the yeast PRP8 protein (Umen and Guthrie, 1996) is schematically shown below; the N-terminal proline-rich region is represented by a black box, and the domains implied in 3′ splice-site fidelity and pyrimidine-tract recognition by shading. The accession numbers for the *T.brucei* p277 and the *S.cerevisiae* PRP8 sequences are Y12638 and Z24732 respectively.

T.brucei genomic DNA indicated that p277 is encoded by a single copy gene (data not shown). A single transcript of 7.8–8 kb in length was detected at equal abundance in two major stages of the trypanosome life cycle (slender bloodstream forms and procyclic forms; data not shown). The major 3′ splice site (spliced leader acceptor site) was mapped by reverse transcription (RT)–PCR and sequencing of the products to an AG located 112 nucleotides upstream of the putative AUG start codon. This 3′ splice site is preceded by a pyrimidine-rich tract (between positions –7 and –52 relative to the 3′ splice site with 76% pyrimidines). The major polyadenylation sites were mapped by RACE–PCR between 169 and 178 nucleotides downstream of the stop codon. Thus, the calculated length of the transcript without poly(A) tail (7530 bp) is compatible with the observed length on Northern blots, and use of the first ATG downstream of the spliced leader acceptor site is very likely. In addition, Western blot analysis with specific antibodies confirmed that a protein of the expected size is detectable in *T.brucei* extract (see below and Figure 3).

Comparison of the T.brucei p277 protein sequence with known cis-spliceosomal homologues

Since PRP8 conservation is exceptionally high (e.g. 61% identity between the yeast PRP8 protein and the corresponding *C.elegans* sequence; see Hodges *et al.*, 1995), we compared the new sequence of the *T.brucei* p277 protein with the known PRP8 homologues (Figure 1 and data not shown). To maximize the alignment between

p277 and PRP8, several gaps had to be introduced into the trypanosome protein sequence (two of them spanning more than 10 amino acids); the sizes of the three proteins are strikingly similar (*T.brucei*, *S.cerevisiae* and *C.elegans*: 2403, 2413 and 2329 amino acids respectively). Significantly, the trypanosomal p277 protein shares only 40% identical and 61% similar amino acids with the yeast PRP8 sequence (*T.brucei*/*C.elegans*: 41% identical and 64% similar amino acids). The overall sequence conservation of the *T.brucei* sequence is more pronounced in the central region of the protein (approximately amino acids 500–2000) than in the N- and C-terminal regions (see Figure 1 for a schematic representation). There is an area with a particularly high degree of identity (62%) between amino acids 535 and 703. Towards the C-terminus the proportion of identical positions strongly declines, from 56% (between amino acids 1551 and 1600) to values of between 20 and 30% at the C-terminal region (amino acids 2100–2403). Yet all three sequences are almost colinear at their carboxy termini. As previously noted (Hodges *et al.*, 1995), there are two extended highly conserved regions in the yeast and *C.elegans* protein sequences, corresponding to yeast amino acids 1600–1660 (97% identity) and amino acids 1889–1947 (93% identity); significantly, the corresponding regions of the trypanosome p277 protein do not exhibit this high level of homology. In contrast to *C.elegans*, where both *cis* and *trans* splicing occur, trypanosomes had to retain only *trans* splicing. Therefore the function implied by these conserved regions may have adapted during evolution to the particular requirements of the *trans*-splicing machinery.

Similarly as for the yeast and *C.elegans* homologues, no known protein motifs in the trypanosomal p277 sequence could be identified. The only exception is the N-terminal sequence of yeast PRP8, where four runs of proline-rich regions interspersed by acidic residues occur (amino acids 1 to ~80, Hodges *et al.*, 1995). This region appears to be yeast-specific, since it is missing not only in the *C.elegans* (Hodges *et al.*, 1995), but also in the *T.brucei* sequence (this study).

The mutational analysis of the yeast PRP8 protein has revealed two separate domains involved in 3′ splice-site fidelity and in uridine-tract recognition (Umen and Guthrie, 1996; see Figure 1). These two domains are close to each other, and the two highly conserved regions (see above) fall into the same area. First, most of the yeast PRP8 mutations that affected 3′ splice-site fidelity cluster within a small internal region and concern positions that are identical in the *C.elegans* sequence (Umen and Guthrie, 1996). Significantly, three of these five amino acid positions in yeast are identical in the trypanosome sequence (yeast M1399/*T.brucei* M1386; yeast T1565/*T.brucei* T1554; yeast W1609/*T.brucei* W1598), one is similar (yeast W1575/*T.brucei* F1564), and only one differs (yeast E1576/*T.brucei* N1565). A second striking result of the mutational analysis in yeast by Umen and Guthrie (1996) was that a collection of U-tract recognition mutations were all mapped to two positions of yeast PRP8, F1834 and E1960, which were found to be conserved in various species. The former position is not conserved in the trypanosome sequence (yeast F1834→*T.brucei* G1823); the latter position is replaced by a similar amino acid (yeast E1960→*T.brucei* D1951).

Taken together, the high overall conservation of the PRP8/p277 protein sequences implies that basic functions of this protein splicing factor are conserved between *cis* and *trans* splicing. This argues in favour of the same catalytic mechanism operating in both splicing systems. However, not all of the amino acid positions critical for function of the *cis*-spliceosomal PRP8 protein are conserved in p277, suggesting that certain differences exist between PRP8 functions in *cis* and *trans* splicing. For example, one can envision that the same mechanism that ensures 3′ splice-site fidelity operates in *cis* and *trans* splicing, whereas the U-rich region is recognized in a different manner.

The trypanosomal p277 protein is associated with a small RNA

Based on its similarity with PRP8, we expected p277 to be associated with a *trans*-spliceosomal homologue of the U5 RNA. In order to study RNA and protein interactions of p277 in *T.brucei*, in particular to identify a putative *trans*-spliceosomal U5 homologue, we raised polyclonal antibodies against a C-terminal portion of p277 (for a characterization by Western blot analysis, see below). Affinity-purified anti-p277 antibodies and the corresponding non-immune control were used for immunoprecipitation from *T.brucei* extract; in parallel, immunoprecipitation was done with anti-common protein antiserum, which is specific for five protein components shared by the U2 and U4/U6 snRNPs as well as the SL and SLA RNPs (Palfi and Bindereif, 1992; Palfi *et al.*, 1994). RNA was purified from the immunoprecipitate and labelled with [^{32}P]pCp and RNA ligase. As Figure 2 shows, anti-p277 antiserum specifically precipitated a small RNA ('RNA X') migrating as a triplet with a major species of ~65 nucleotides (see lane anti-p277). The other RNA species in the 80-nucleotide range most likely represent non-specifically precipitated tRNAs also seen with the control non-immune serum. Significantly, an RNA of the same mobility as RNA X was recognized by anti-common protein antibodies (see lane anti-CP), in addition to a collection of RNAs, the major species of which have previously been identified as U2, SL, U4, and U6 (Palfi and Bindereif, 1992). Based on partial RNase T1 digestion patterns of gel-purified material, RNA X co-precipitated by either anti-p277 or anti-CP antibodies appeared to be identical (data not shown). In sum, we conclude that a small RNA of ~65 nucleotides is specifically associated with both p277 and common proteins; therefore this RNA represents a likely candidate for the U5-analogous *trans*-spliceosomal snRNA of *T.brucei*.

Stable association of the T.brucei U5 RNA with p277 protein and common proteins: a trans-spliceosomal U5-analogous RNP

Recently Dungan *et al.* (1996) identified a novel RNA called SLA2/U5-like RNA as a good candidate for the U5-homologue of *T.brucei*. Therefore we have tested whether the p277 protein is associated with this RNA, using primer-extension assays and a primer specific for the 3′ terminal region of the SLA2/U5-like RNA (hereafter referred to as U5 RNA).

To characterize the putative trypanosomal U5 RNP, *T.brucei* S100 extract was fractionated through glycerol

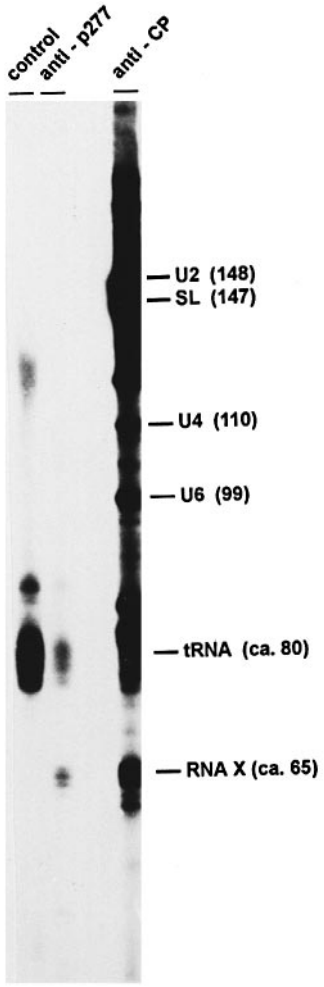

Fig. 2. Immunoprecipitation of small RNAs with anti-p277 and anti-common protein antibodies. RNAs immunoprecipitated from *T.brucei* S100 extract with anti-p277 antibodies (lane anti-p277), the corresponding non-immune serum (lane control) and anti-common protein antibodies (lane anti-CP) were ^{32}P 3′ end-labelled and separated on a denaturing 10% polyacrylamide gel. The positions of immunoprecipitated RNAs and their sizes (in nucleotides) are indicated on the right.

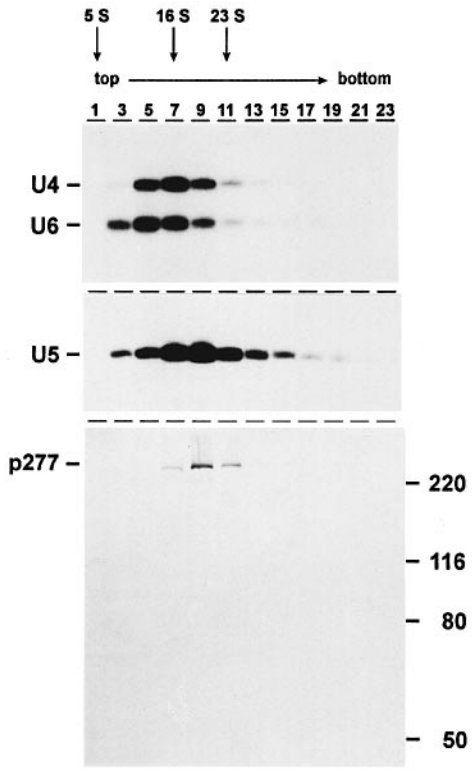

Fig. 3. Glycerol gradient sedimentation of the *T.brucei* U5 and U4/U6 snRNPs. S100 extract was fractionated through a 10–30% glycerol gradient (fractions #1–23, from top to bottom); the distribution of U4, U5 and U6 snRNAs was assayed by primer extension (upper and middle panels), the distribution of p277 by Western blotting (lower panel). The positions of size markers (5S, 16S and 23S) are indicated on the top, protein size markers on the right.

gradient sedimentation. RNA was prepared from gradient fractions and analysed by primer extension for U4, U5 and U6 RNAs. In parallel, protein was purified, and the p277 distribution was tested by immunoblotting with affinity-purified p277 antibodies. Figure 3 shows that most of the U5 RNA is concentrated in fractions #7–11, corresponding to the 20S region. The distribution of U4 and U6 RNAs across fractions #5–9 (corresponding to the 16S region) overlaps with the U5 region, is slightly shifted towards the top of the gradient, and reflects the occurrence of both U4/U6 snRNP and free U6 RNP. As the Western blot analysis shows, p277 was detected above the 220 kDa marker protein as the only protein band, with a peak in fraction #9. This fractionation of p277 reflects the distribution of the U5 RNA at ~20S.

To assess whether U5 RNA and p277 are associated with each other, fractions containing U4, U5, and U6 RNAs were pooled, followed by anti-p277 and anti-common protein immunoprecipitations and U4, U5, and U6 RNA primer extensions. In parallel, non-immune serum was used in control reactions. Figure 4 (gradient fraction lanes) clearly shows that the U5 RNA could be efficiently and specifically immunoprecipitated with anti-p277 antibodies; however, neither U4 nor U6 RNA could be detected in the anti-p277 immunoprecipitate. This confirms that the anti-p277 immunoprecipitated RNA species ('RNA X') we have detected by 3′ end-labelling (see above and Figure 2) and by primer extension with a U5 RNA-specific oligonucleotide (Figure 4) are identical. Using anti-common protein antiserum, U5 RNA was also efficiently immunoprecipitated, as well as U4 and U6 RNAs. Taken together, these data establish that the trypanosomal U5 RNA is stably and specifically associated with both p277 protein and common proteins in an RNP representing the *trans*-spliceosomal U5 snRNP. Somewhat surprisingly U4 and U6 RNAs were not found in this complex; therefore under our conditions the major fraction of U5 RNA (at least 95%) appears not to be present in a U4/U5/U6 triple-snRNP form. Since we might have missed a small fraction of U5 in the triple-snRNP form sedimenting outside of the pooled region of the gradient, the same set of immunoprecipitations was carried out with unfractionated *T.brucei* S100 extract, yielding identical results as with the pooled gradient fractions (Figure 4, S100 extract lanes).

Core particles of snRNPs are characteristically stable during isopycnic centrifugation in CsCl (Lelay-Taha *et al.*, 1986), as shown previously for the RNPs of SL, U2, U4/U6, and SLA RNAs (Michaeli *et al.*, 1990; Cross *et al.*, 1991; Günzl *et al.*, 1992; Palfi *et al.*, 1994). To characterize

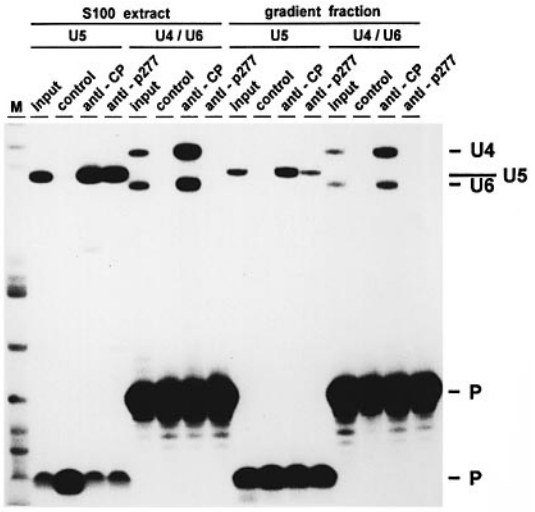

Fig. 4. Immunoprecipitation analysis of the *T.brucei* U5 and U4/U6 snRNPs. snRNPs were immunoprecipitated from *T.brucei* S100 extract (left half) and from pooled glycerol gradient fractions (right half; #5–9 from glycerol gradient shown in Figure 3), using anti-common protein antibodies (lanes anti-CP), the corresponding non-immune serum (lanes control) and anti-p277 antibodies (lanes anti-p277). snRNAs were detected in immunoprecipitates and in 10% of the input (lanes input) by primer extension. The positions of the primer-extension products (U4, U5 and U6) and the primers (P) are indicated on the right. M, pBR322–*Hpa*II marker fragments.

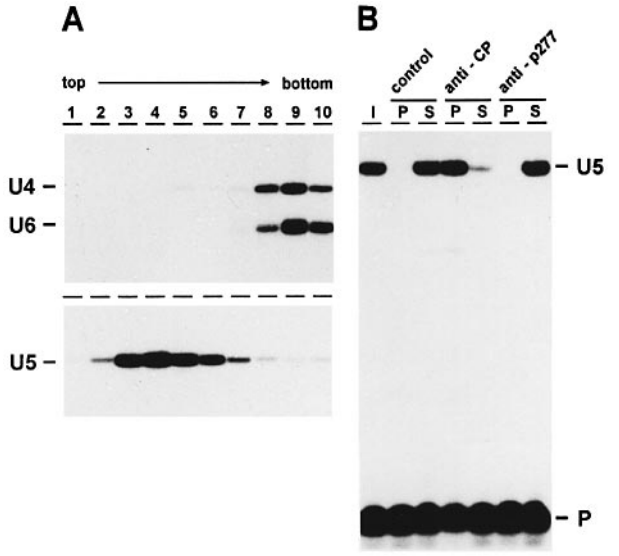

Fig. 5. CsCl density gradient fractionation and immunoprecipitation analysis of the *T.brucei* U5 snRNP. (**A**) *T.brucei* S100 extract was subjected to isopycnic CsCl gradient centrifugation, and gradient fractions #1–10 (from top to bottom) were assayed by primer extension for U4, U5, and U6 RNAs. (**B**) U5 core snRNP-containing peak fractions (#3–5) were pooled and analysed by immunoprecipitation with anti-common protein antibodies (lanes anti-CP), the corresponding non-immune serum (lanes control) and anti-p277 antibodies (lanes anti-p277). U5 RNA (U5) was detected in 10% of the input (lane I), the total pellet (lanes P) and 10% of the supernatant fractions (lanes S) of the immunoprecipitation reactions by primer extension with a U5-specific oligonucleotide (P).

the U5 RNP of *T.brucei* further, S100 extract was fractionated through a CsCl density gradient, RNA from the fractions was prepared, and the distribution of U4, U5, and U6 RNAs was analysed by primer extension. As Figure 5A shows, U5 core snRNPs are concentrated in fractions #3–6 and are clearly separated from U4/U6 core snRNPs (fractions #8–10). The peak fractions of the U5 core RNP (#4) and the U4/U6 core snRNP (#9) correspond to densities of 1.36 g/ml and 1.51 g/ml respectively; based on an empirical formula (Spirin, 1969), the RNA–protein mass ratio of the U5 core snRNP was calculated to be 18:82, corresponding to a total protein mass of ~93 kDa per 62-nucleotide U5 RNA. For the U4/U6 core snRNP, an RNA–protein mass ratio of 43:57 would reflect a total protein mass of ~89 kDa per U4/U6 RNA molecule. The close correspondence between the total mass of the U4/U6 and U5 core protein components (90 kDa) suggests that they share the same set of proteins; however, since the common proteins we have identified previously add up to a total mass of 60 kDa (Palfi *et al.*, 1991), either their stoichiometry is not equimolar, or there are more, unidentified core protein components.

To identify the protein components of the trypanosomal U5 core snRNP immunoprecipitations were carried out (Figure 5B). Pooled CsCl gradient fractions containing U5 core snRNPs were dialysed and reacted with anti-p277 and anti-common protein antibodies, and—as a control for specificity—with non-immune serum. Precipitates and supernatants of the immunoreaction were then analysed by primer extension for their U5 RNA content. The U5 core snRNP was efficiently precipitated by anti-common protein antibodies, but not detectably by anti-p277 antibodies, demonstrating clearly that common proteins, but not the U5-specific p277 protein, are present in the U5 core snRNP.

Comparison between cis and trans-splicing: functional role and interactions of the U5 snRNP

The U5-specific p277 protein of *T.brucei* represents only the second protein factor identified and cloned in the SL-type *trans*-splicing system from trypanosomatids; the first case was the U2-specific 40K protein of *T.brucei*, the N-terminal half of which is homologous to the *cis*-spliceosomal U2 A′ protein (Cross *et al.*, 1993). In contrast, homology between the *trans*-spliceosomal p277 protein and its *cis*-spliceosomal counterparts extends throughout their entire length, although the level of homology changes considerably across the length of the protein (see Figure 1). It is of interest that not all functionally important positions, which were identified in the yeast PRP8 protein and which are conserved in the other *cis*-spliceosomal homologues, occur in the trypanosome p277 sequence. However, there is no evidence at present for an additional, potentially *trans*-splicing specific domain in the p277 protein. Since all known PRP8 homologues lack any recognizable protein motifs (with the exception of the yeast PRP8-specific proline-rich repeats), a mutational analysis of the trypanosomal p277 protein should help to define the domain structure conserved between *cis*- and *trans*-splicing. Finally, because of the large size of this protein factor it might be a key component to identify other interacting protein and snRNA elements in the *trans* spliceosome.

Based on the results of this study and that of Dungan *et al.* (1996) it was surprising to find that conservation of the U5 snRNP is quite different on the protein and RNA level: whereas with the exception of the 5′ loop the trypanosomal U5 RNA sequence greatly differs in its

overall secondary structure and sequence from all the *cis*-spliceosomal U5 RNAs, at least one protein component (PRP8/p220/p277) is highly homologous. This indicates that important functional properties reside in the protein components of the splicing machinery and are conserved between *cis* and *trans* splicing.

One of the most important open questions in the *trans*-splicing field is how both splice sites on the two separate precursor RNAs are recognized and how they interact stably with each other. Although it is now established that a U5-analogous RNP exists in the trypanosome system, we still know only very little about how it acts during *trans*-spliceosome assembly and *trans* splicing. Since a model has been proposed that incorporates both U1 and U5 RNA functions into the trypanosomal SL RNA (Steitz, 1992), the question remains how the U5 RNA—in conjunction with its associated p277 protein—functions in splice-site recognition. In *cis* splicing U5 enters the spliceosome in the form of the U4/U5/U6 triple-snRNP; our analysis in the trypanosome *trans*-splicing system gave no evidence for such a stable snRNP complex in significant abundance, although Dungan *et al.* (1996) showed evidence for a small fraction of U5 existing together with U4 and U6 snRNAs as a triple-snRNP. These results do not rule out that U5 forms *in vivo*—in analogy to the *cis*-spliceosomal complex—a triple-snRNP that functions as a precursor in *trans*-spliceosome assembly, yet is not stable under extract conditions.

In the absence of an *in vitro* system derived from trypanosomes it will be particularly important to establish functional *in vivo* systems to carry out mutational studies of the p277 protein and the U5 RNA components. A detailed comparison between results from such studies and those from *cis*-splicing systems as well as the use of heterologous systems promises to provide novel insights not only into the pathway of *trans*-spliceosome assembly and the basic *trans*-splicing mechanism, but also into the *cis*- and *trans*-specific characteristics of these two important spliceosome components.

Materials and methods

Trypanosome cell culture and extract preparation

Cultures of the procyclic form of *T.brucei* strain 427 (Cross, 1975) were grown at 28°C in SDM-79 medium (Brun and Schönenberger, 1979) supplemented with 5% (v/v) fetal calf serum and 5 mg/l haemin. S100 extract was prepared as described previously (Cross *et al.*, 1991).

Cloning and sequencing of the p277 gene

The C-terminus of p277 was identified by sequencing the genomic clone pC3-1, which harbours a protein kinase A catalytic subunit gene of *T.brucei* (T.Klöckner and M.Boshart, manuscript in preparation). A genomic *T.brucei* (strain AnTat 1.1) library of partially digested *Sau*3A fragments (R.Krämer and M.Boshart, unpublished results) constructed in λDASH II (Stratagene) was screened with a riboprobe covering amino acid positions L2017–I2323 of p277. Four phage clones were isolated and mapped, and λ#6 carrying a 12.4 kb insert with the entire p277 reading frame was selected for subcloning and construction of nested deletions (Pharmacia). Both strands of the region covering the p277 reading frame were sequenced by automated cycle sequencing. Nucleotide sequences were compiled and analysed using the GCG package (version 8); protein sequence alignment and statistical analysis were performed using the GAP program (Devereux *et al.*, 1984).

Northern blot and RACE analysis

Slender bloodstream forms of *T.brucei* strain AnTat 1.1 were purified from infected rat blood. Procyclic forms were differentiated *in vitro* from bloodstream forms as described (Vassella and Boshart, 1996). Total RNA was isolated using the guanidinium isothiocyanate/CsCl cushion method (Ausubel *et al.*, 1987) and 5 µg were separated on a 1% agarose–formaldehyde gel, blotted, and hybridized as described in Boshart *et al.* (1991) with a riboprobe covering amino acid positions L2017–I2323. The polyadenylation site was mapped by RACE techniques using the Marathon cDNA amplification kit (Clontech), poly(A)$^+$-RNA from procyclic forms and the p277-specific primer 5′-GAGGCTGTGT-CTGTGCGTCTG-3′ (corresponding to amino acid positions E2040–L2046). Amplification products were cloned using the internal *Sal*I site at amino acid position S2330 and sequenced. The 5′-end was mapped by RACE techniques, using a p277-specific primer (5′-GGGTACTT-CATCCTTTTG-3′; corresponding to amino acids Q58 to P64) and an SL-specific primer, 5′-AACTAACGCTATTATTAGAA-3′ (nucleotides 1–20 of the *T.brucei* SL RNA), followed by direct sequencing of the PCR product.

Antibodies and bacterial expression of proteins

The C-terminal region of p277 from position P2019 to T2331 was PCR-amplified and cloned in *Bam*HI/*Sal*I cut bacterial expression vectors pGEX4T3 (Pharmacia) and pQE30 (Qiagen) via a primer-derived *Bam*HI and the genomic *Sal*I site at amino acid position S2330. Fusion proteins were expressed using the protocols provided by the suppliers and extracted from the insoluble fraction with 8 M urea followed by affinity purification on Ni-NTA columns (Qiagen) or by SDS–PAGE and blotting onto nitrocellulose. Rabbits were immunized with powdered or DMSO-dissolved nitrocellulose strips (200 µg fusion protein) using either Freund's adjuvant or Titermax™ (Vaxcel, Inc.) and boosted four times at 4–5 week intervals. Sera and preimmune controls were affinity purified on nitrocellulose strips (Pringle *et al.*, 1991) using 200 µg His-tagged fusion protein per ml of serum.

Glycerol gradient centrifugation

One millilitre of *T.brucei* S100 extract was mixed with 0.5 ml of buffer G (20 mM KCl, 20 mM HEPES pH 8.0, 1 mM $MgCl_2$) and fractionated through an 11 ml glycerol gradient in buffer G (10–30% glycerol; SW 40 rotor, 6 h, 40 000 r.p.m., 4°C). Fractions of 0.5 ml were collected from the top of the gradient. RNA was prepared from 100 µl aliquots of every second fraction and assayed for U4, U5 and U6 RNA by primer extension with DNA oligonucleotides complementary to nucleotides 50–62 of U5 RNA (Dungan *et al.*, 1996), 53–72 of U4 RNA and 40–59 of U6 RNA (Cross *et al.*, 1991). For primer extension reactions, RNA and primers were co-precipitated with ethanol and dissolved in 1× Expand™ reverse transcriptase buffer containing 0.5 mM of each dNTP. Following denaturation for 10 min at 65°C and cooling on ice, the reactions were supplemented with RNasin (0.5 U/µl), 10 mM DTT and 25 U of Expand™ reverse transcriptase (Boehringer Mannheim). Incubation was for 10 min at 30°C followed by 45 min at 42°C. Primer-extension products were ethanol precipitated and fractionated by denaturing polyacrylamide gel electrophoresis (10%).

CsCl density gradient centrifugation

CsCl was dissolved in buffer D (Dignam *et al.*, 1983) containing 15 mM $MgCl_2$ to a density of 1.48 g/ml. Two hundred microlitres of *T.brucei* S100 extract were fractionated through a 1 ml gradient (centrifugation in Beckman TL 100 tabletop ultracentrifuge; TLA-100.2 rotor; 90 000 r.p.m. for 20 h at 4°C). Fractions of 100 µl were collected from the top of the gradient. The density distribution across the gradient was measured in a gradient run in parallel to range from 1.31 g/ml (fraction #1) to 1.54 g/ml (fraction #10). RNA prepared from 50 µl of each fraction was assayed by primer extension with U4, U5 and U6 RNA-specific primers. U5 RNA peak fractions were pooled, dialysed against buffer D supplemented with 1 mM DTT and 1 mM PMSF, and used for immunoprecipitation reactions.

Immunoprecipitation and primer-extension analysis

Fifty microlitres of a rabbit serum containing antibodies against the common snRNP proteins of *T.brucei* ('anti-CP'; Palfi and Bindereif, 1992) and the corresponding non-immune serum were coupled to 25 µl of pre-swollen protein A–Sepharose beads (Pharmacia) in 400 µl NET-150 buffer [50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 0.01% Nonidet P-40 (v/v), 0.5 mM DTT] at 4°C overnight. For anti-p277 immunoprecipitations affinity-purified rabbit anti-p277 antibodies were used. The pre-coated beads were washed four times with 0.5 ml of NET-150 each and resuspended in 400 µl of the same buffer. One hundred microlitres of *T.brucei* S100 or gradient fraction respectively, were incubated with the protein A–Sepharose bound antibodies for 2 h at

4°C; then the beads were washed extensively with NET-150 buffer. RNA was released from the beads by phenol/chloroform extractions and precipitated from the aqueous phase with ethanol, followed by 3′ end-labelling with 10 μCi [^{32}P]pCp and T4 RNA ligase, as described by England and Uhlenbeck (1978), and analysed by denaturing polyacrylamide gel electrophoresis. For primer extension analysis RNA was purified from both pellet and supernatant of the immunoprecipitation reactions as well as from 10% of the input and assayed with U4, U5 and U6 RNA-specific primers.

Immunoblot analyis

Protein was prepared from glycerol gradient fractions (300 μl of every second fraction) by phenolization and acetone precipitation, separated on an 8.5% SDS–polyacrylamide gel, and transferred electrophoretically to Hybond ECL nitrocellulose membrane (Amersham). Blots were blocked in PBS, pH 8.0, containing 1× Roti-Block solution (Roth) and 0.2% Tween-20 overnight at 4°C, followed by probing for 4 h with anti-p277 antibodies (diluted 1:250 in blocking solution). Blots were washed in 1× PBS containing 0.2% Tween-20. Immunocomplexes were detected with sheep anti-rabbit IgG-horseradish peroxidase conjugate diluted 1:20 000 in blocking solution using ECL developing reagents from Amersham.

Acknowledgements

We thank Andreas Langer and Shahri Raasi for help with the construction of deletion derivatives and acknowledge Miriam van den Bogaard for technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (A.B.) and German Federal Ministry for Science and Technology (BMFT; grant BEO 21/0316211A; M.B.).

References

Agabian,N. (1990) *Trans* splicing of nuclear pre-mRNAs. *Cell*, **61**, 1157–1160.

Anderson,G.J., Bach,M., Lührmann,R. and Beggs,J.D. (1989) Conservation between yeast and man of a protein asssociated with U5 small nuclear ribonucleoprotein. *Nature*, **342**, 819–822.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds) (1987) *Current Protocols in Molecular Biology*. Greene Publishing Associates and Wiley-Interscience, New York.

Boshart,M., Weih,F., Nichols,M. and Schütz,G. (1991) The tissue-specific extinguisher locus TSE1 encodes a regulatory subunit of cAMP-dependent protein kinase. *Cell*, **66**, 849–859.

Brown,J.D. and Beggs,J.D. (1992) Roles of PRP8 protein in the assembly of splicing complexes. *EMBO J.*, **11**, 3721–3729.

Brun,R. and Schönenberger,M. (1979) Cultivation and *in vitro* cloning of procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. *Acta Tropica*, **36**, 289–292.

Bruzik,J.P., Van Doren,K. Hirsh,D. and Steitz,J.A. (1988) *Trans* splicing involves a novel form of small nuclear ribonucleoprotein particles. *Nature*, **335**, 559–562.

Cortes,J.J., Sontheimer,E.J., Seiwert,S.D. and Steitz,J.A. (1993) Mutations in the conserved loop of human U5 snRNA generate use of novel cryptic 5′ splice sites *in vivo*. *EMBO J.*, **12**, 5181–5189.

Cross,G.A.M. (1975) Identification, purification and properties of clone specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. *Parasitology*, **71**, 393–417.

Cross,M., Günzl,A., Palfi,Z. and Bindereif,A. (1991) Analysis of small nuclear ribonucleoproteins (RNPs) in *Trypanosoma brucei*: structural organization and protein components of the spliced leader RNP. *Mol. Cell. Biol.*, **11**, 5516–5526.

Cross,M., Wieland,B., Palfi,Z., Günzl,A., Röthlisberger,U., Lahm,H.-W. and Bindereif,A. (1993) The *trans*-spliceosomal U2 snRNP protein 40K of *Trypanosoma brucei*: Cloning and analysis of functional domains reveals homology to a mammalian snRNP protein. *EMBO J.*, **12**, 1239–1248.

Devereux,J., Haeberli,P. and Smithies,O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387–395.

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

Dungan,J.M., Watkins,K.P. and Agabian,N. (1996) Evidence for the presence of a small U5-like RNA in active trans-spliceosomes of *Trypanosoma brucei*. *EMBO J.*, **15**, 4016–4029.

England,T.E. and Uhlenbeck,O.C. (1978) 3′-terminal labelling of RNA with T4 RNA ligase. *Nature*, **275**, 560–561.

Frank,D.N., Roiha,H. and Guthrie,C. (1994) Architecture of the U5 small nuclear RNA. *Mol. Cell. Biol.*, **14**, 2180–2190.

Günzl,A., Cross,M. and Bindereif,A. (1992) Domain structure of U2 and U4/U6 small nuclear ribonucleoprotein particles from *Trypanosoma brucei*: identification of *trans*-spliceosomal specific RNA–protein interactions. *Mol. Cell. Biol.*, **12**, 468–479.

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

Hodges,P.E., Jackson,S.P., Brown,J.D. and Beggs,J.D. (1995) Extraordinary sequence conservation of the PRP8 splicing factor. *Yeast*, **11**, 337–342.

Jackson,S.P., Lossky,M. and Beggs,J.D. (1988) Cloning of the *RNA8* gene of *Saccharomyces cerevisiae*, detection of the RNA8 protein, and demonstration that it is essential for nuclear pre-mRNA splicing. *Mol. Cell. Biol.*, **8**, 1067–1075.

Lelay-Taha,M.-N., Reveillaud,I., Sri-Widada,D.M., Brunel,C. and Jeanteur,P. (1986) RNA–protein organization of U1, U5 and U4-U6 small nuclear ribonucleoproteins in HeLa cells. *J. Mol. Biol.*, **189**, 519–532.

Maroney,P.A., Yu,Y.-T., Jankowska,M. and Nilsen,T. (1996) Direct analysis of nematode *cis*- and *trans*-spliceosomes: a functional role for U5 snRNA in spliced leader addition *trans*-splicing and the identification of novel Sm snRNPs. *RNA*, **2**, 735–745.

Michaeli,S., Roberts,T.G., Watkins,K.P. and Agabian,N. (1990) Isolation of distinct small ribonucleoprotein particles containing the spliced leader and U2 RNAs of *Trypanosoma brucei*. *J. Biol. Chem.*, **265**, 10582–10588.

Murphy,W.J., Watkins,K.P. and Agabian,N. (1986) Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. *Cell*, **47**, 517–525.

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

Newman,A.J. and Norman,C. (1992) U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell*, **68**, 743–754.

Nilsen,T.W. (1993) *Trans*-splicing of nematode premessenger RNA. *Annu. Rev. Microbiol.*, **47**, 413–440.

Nilsen,T.W. (1995) *Trans*-splicing: an update. *Mol. Biochem. Parasitol.*, **73**, 1–6.

O'Keefe,R.T., Norman,C. and Newman A.J. (1996) The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing. *Cell*, **86**, 679–689.

Palfi,Z. and Bindereif,A. (1992) Immunological characterization and intracellular localization of *trans*-spliceosomal small nuclear ribonucleoproteins in *Trypanosoma brucei*. *J. Biol. Chem.*, **267**, 20159–20163.

Palfi,Z., Günzl,A., Cross,M. and Bindereif,A. (1991) Affinity purification of *Trypanosoma brucei* small nuclear ribonucleoproteins reveals common and specific protein components. *Proc. Natl Acad. Sci. USA*, **88**, 9097–9101.

Palfi,Z., Xu,G.-L. and Bindereif,A. (1994) Spliced leader-associated RNA of trypanosomes. *J. Biol. Chem.*, **269**, 30620–30625.

Pringle,J.R., Adams,A.E.M., Drubin,D.G. and Haarer,B.K. (1991) Immunofluorescence methods for yeast. *Methods Enzymol.*, **194**, 565–602.

Reyes,J.L., Kois,P., Konforti,B.B. and Konarska,M.M. (1996) The canonical GU dinucleotide at the 5′ splice site is recognized by p220 of the U5 snRNP within the spliceosome. *RNA*, **2**, 213–225.

Sontheimer,E.J. and Steitz,J.A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262**, 1989–1996.

Spirin,A.S. (1969) Informosomes. *Eur. J. Biochem.*, **10**, 20–35.

Staknis,D. and Reed,R. (1994) Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. *Mol. Cell. Biol.*, **14**, 2994–3005.

Steitz,J.A. (1992) Splicing takes a holliday. *Science*, **257**, 888–889.

Sutton,R.E. and Boothroyd,J.C. (1986) Evidence for *trans* splicing in trypanosomes. *Cell*, **47**, 527–535.

Teigelkamp,S., Newman,A.J. and Beggs,J.D. (1995a) Extensive interactions of PRP8 protein with the 5′ and 3′ splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J.*, **14**, 2602–2612.

Teigelkamp,S., Whittaker,E. and Beggs,J.D. (1995b) Interaction of the yeast splicing factor PRP8 with substrate RNA during both steps of splicing. *Nucleic Acids Res.*, **23**, 320–326.
Tschudi,C. and Ullu,E. (1990) Destruction of U2, U4, or U6 small nuclear RNA blocks *trans* splicing in trypanosome cells. *Cell*, **61**, 459–466.
Umen,J.G. and Guthrie,C. (1995a) A novel role for a U5 snRNP protein in 3′ splice site selection. *Genes Dev.*, **9**, 855–868.
Umen,J.G. and Guthrie,C. (1995b) Prp16p, Slu7p, and PRP8p interact with the 3′ splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. *RNA*, **1**, 584–597.
Umen,J.G. and Guthrie,C. (1996) Mutagenesis of the yeast gene *PRP8* reveals domains governing the specificity and fidelity of 3′ splice site selection. *Genetics*, **143**, 723–739.
Vassella,E. and Boshart,M. (1996) High-molecular mass agarose matrix supports growth of bloodstream forms of pleomorphic *Trypanosoma brucei* strains in axenic culture. *Mol. Biochem. Parasitol.*, **82**, 91–105.
Whittaker,E. and Beggs,J.D. (1991) The yeast PRP8 protein interacts directly with pre-mRNA. *Nucleic Acids Res.*, **19**, 5483–5489.
Wyatt,J.R., Sontheimer,E.J. and Steitz,J.A. (1992) Site-specific cross-linking of mammalian U5 snRNP to the 5′ splice site before the first step of pre-mRNA splicing. *Genes Dev.*, **6**, 2542–2553.

Received on February 20, 1997; revised on March 24, 1997