A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes

Keith R. Matthews,1,3 Christian Tschudi,1 and Elisabetta Ullu 1,2,4

1Department of Internal Medicine and 2Cell Biology, Yale University School of Medicine, New Haven, Connecticut 06520-8022 USA

In trypanosomes, the generation of monocistronic mRNAs from polycistronic precursors is achieved via RNA processing, namely trans-splicing of the spliced leader sequence at the 5' end and cleavage/polyadenylation at the 3' end of the mRNA coding region. Recent evidence raised the intriguing possibility that these two reactions are coupled. To begin a dissection of the signals required for mRNA 5'-end and 3'-end formation and to uncover potential interactions between trans-splicing and polyadenylation, we mutagenized the intergenic region between the β- and α-tubulin genes of Trypanosoma brucei. Block substitutions identified the pyrimidine-rich sequences at the α-tubulin 3'-splice-acceptor site as a major determinant for accurate trans-splicing downstream and 3'-end formation upstream. In addition to the utilization of cryptic 3'-splice sites, obliteration of the polypyrimidine tracts led to aberrant poly[A]n site choice, even in the presence of the wild-type poly[A]n site and neighboring sequences. Taken together, these results indicate that the polypyrimidine-rich sequences act as a bifunctional element that affects RNA processing both upstream and downstream from itself. This is consistent with the possibility that the polypyrimidine tract is recognized by both the trans-splicing and polyadenylation machineries, either sequentially or simultaneously.

[Key Words: pyrimidine-rich sequences; trans-splicing; polyadenylation; Trypanosoma brucei]

Received September 27, 1993; revised version accepted December 21, 1993.

In protozoa of the family Trypanosomatidae, which include African and South American trypanosomes and Leishmania, most protein coding genes are transcribed into polycistronic pre-mRNAs [Johnson et al. 1987; Tschudi and Ullu 1988; Muhich and Boothroyd 1989; Rudenko et al. 1990]. The various mRNA coding regions are separated by short spacer regions, ranging in size between one hundred and a few hundred nucleotides. Mature monocistronic mRNA molecules are generated from these precursors post-transcriptionally by two RNA processing reactions: trans-splicing [for review, see Agabian 1990] and 3'-end cleavage/polyadenylation. The former involves a splicing event between two unlinked precursor molecules, namely the spliced leader [SL] RNA, which provides the SL sequence, and the pre-mRNA. Trans-splicing has been shown to be mechanistically similar to cis-splicing of intervening sequences in yeast and higher eukaryotes. For instance, the structure of the 3'-splice acceptor site in trypanosome pre-mRNA closely resembles the consensus sequence established for mammalian introns: it consists of the invariant AG dinucleotide and an upstream polypyrimidine tract of variable length [Huang and Van der Ploeg 1991; Curotto de LaFaille et al. 1992]; however no branch site consensus sequence has been so far identified [Patzelt et al. 1989]. In particular, the efficiency of trans-splicing of the procyclic acidic repetitive protein [PARP] α gene of Trypanosoma brucei has been shown to be extremely sensitive to changes in the polypyrimidine tract at the 3' splice site [Huang and Van der Ploeg 1991].

At present, very little is known about the process of trypanosome mRNA 3'-end formation. It is assumed that the 3' end of mature mRNAs is generated by a mechanism closely related to that operating in higher eukaryotes, namely cleavage of the pre-mRNA at the polyadenylation site followed by addition of the poly[A]n tail [for review, see Proudfoot 1991; Wahle and Keller 1992]. However, no conserved motif like the AAUAAA sequence of higher eukaryotes is present upstream of the trypanosome polyadenylation sites. Furthermore, multiple closely spaced 3' ends are often found for a given species of trypanosome mRNA [Tschudi and Ullu 1988; LeBowitz et al. 1993 and references therein]. These two qualitative aspects of trypano-
some mRNA 3'-end formation underscore a potential similarity with the mechanism of 3'-end formation in yeast (Heidmann et al. 1992, and references therein).

Recent experiments suggested that during maturation of trypanosome polycistronic pre-mRNAs there is a temporal and perhaps mechanistic relationship between trans-splicing and 3'-end formation/polyadenylation. First, when trans-splicing is inhibited by destruction of U2 snRNA in permeabilized cells of T. brucei, 3'-end formation of tubulin mRNA (as well as of the majority of mRNAs) is almost completely inhibited (Ullu et al. 1993). Second, already trans-spliced tubulin transcripts that have not yet acquired a detectable poly(A) + tail are readily identified in nascent RNA (Ullu et al. 1993). Together these observations support the model that at least for the tubulin pre-mRNA trans-splicing precedes and perhaps mechanistically is linked to the formation of polyadenylated transcripts.

To gain further insight into cis-acting signals for RNA processing and the potential mechanistic relationship between trans-splicing and 3'-end formation, we have mutagenized by block substitutions the region between the T. brucei β- and α-tubulin genes. All mutations were introduced in the context of a synthetic dicistronic construct consisting of two reporter genes, namely the bacterial chloramphenicol acetyl transferase (CAT) gene placed upstream and the fire fly luciferase (LUC) gene placed downstream. The results of this analysis demonstrated that the integrity of two blocks of pyrimidine residues upstream of the α-tubulin 3'-splice site was required to specify accurate 3'-end formation of the upstream β- and trans-splicing of the downstream α-tubulin gene.

Results

Deletion of sequences immediately upstream of the α-tubulin 3'-splice site results in shortened transcripts from the upstream CAT gene

To dissect the signals required for efficient and accurate mRNA trans-splicing and 3'-end formation, we assembled synthetic transcription units under the control of the promoter for the PARP α gene of T. brucei (Rudenko et al. 1990; Sherman et al. 1991). We chose the tubulin pre-mRNA as our model because it is one of the best characterized systems in T. brucei. In this organism the tubulin genes are arranged as head to tail α- and β-tubulin repeating units (Thomashow et al. 1983; Kimmel et al. 1985). The first construct [Fig. 1A, plasmid PCATβ/α], which we refer to as monocistronic, consisted of the PARP promoter driving expression of the CAT coding region. At the 3'-end of CAT we placed the sequences that in the trypanosome genome separate the β- and the α-tubulin translated regions [we refer to these sequences as the β/α region], β/α is 643 bp long and contains in a 5'-→3' direction the 3'-untranslated sequences and polyadenylation sites of the upstream β-tubulin gene, a noncoding spacer region, the 3'-splice acceptor site and 5'-untranslated sequence of the downstream α-tubulin gene (Kimmel et al. 1985; Sather and Agabian 1985; Imboden et al. 1986). To test the extent of sequence requirement for β-tubulin 3'-end formation we generated three 3' deletion derivatives of the parent construct [Fig. 1A, plasmids PCATβ/αT1, PCATβ/αT2 and PCATβ/αT3]. The various plasmid DNAs were introduced by electroporation into cultured insect form trypanosomes and after 6 hr total RNA was isolated and analyzed by Northern blot hybridization using antisense CAT RNA as a probe (Fig. 1B). The plasmid PCATβ/α gave rise to abundant CAT RNA [lane 2] of the predicted length (~1100 nucleotides). Rapid amplification of cDNA 3' ends (3'-end RACE) with specific oligonucleotides indicated that polyadenylation of the CAT RNA chimera occurred at the major β-tubulin in vivo sites (data not shown, but see legend to Fig. 1A). As compared to the parent plasmid, all three deletion derivatives of the β/α region [lanes 3–5] gave rise to decreased amounts of CAT RNA, although deletion PCATβ/αT3 had no effect on the size of the CAT RNA produced [lane 5]. In contrast, when we deleted an additional 48 bp [PCATβ/αT2], two new size classes of CAT RNAs were detected [lane 4]: Both were smaller than the wild-type CAT transcript of 1100 nucleotides [PCATβ/α], with sizes of ~1000 and ~800 nucleotides. A more extensive deletion [PCATβ/αT1] gave a result similar to the PCATβ/αT2 construct [lane 3] in terms of RNA size distribution. Analysis of mRNA 3' ends by 3'-end RACE and by oligo(dT) selection showed that the aberrant CAT RNAs were polyadenylated and were shortened at their 3'end due to the utilization of new polyadenylation sites (data not shown, but see Fig. 5). Thus, the deletion derivatives of the β/α region suggested that the 48 bp located between the deletion endpoints of mutants PCATβ/αT2 and PCATβ/αT3 were necessary to specify the positions and perhaps also the efficiencies of utilization of polyadenylation sites.

To confirm the phenotype of the T2 deletion in a different genetic context, we removed the T2 sequence from plasmid PCATβ/aLUCP [Fig. 1A], the dicistronic plasmid that we used for all further mutational analysis. In this construct the β/α sequences were sandwiched between the CAT and the LUC genes. At the 3'-end of the LUC gene we placed sequences spanning the poly-
Pyrimidines in trans-splicing and polyadenylation

Figure 1. Deletion of sequences upstream of the α-tubulin 3' splice site gives shortened CAT transcripts. [A] Schematic representation of plasmid constructs used for transfection. PCATβ/αT1 extended up to nucleotide 431 downstream of the β-tubulin stop codon, that is ~50-60 nucleotides beyond the β-tubulin polyadenylation sites. The deletion endpoints of PCATβ/αT2 and PCATβ/αT3 were at positions 485 and 534, respectively. For the polycistronic construct PCATβ/αLUC, the LUC coding region followed by the PARP 3'-untranslated sequence and polyadenylation site was fused after the α-tubulin 3'-splice site and untranslated sequences. The positions of the wild-type β-tubulin polyadenylation site and α-tubulin 3'-splice site are indicated. From S1 mapping [Imboden et al. 1986] and cDNA sequencing [K.R. Matthews, unpubl.] the polyadenylation sites for the β-tubulin mRNA are known to occur between nucleotides 366- and 373-downstream of the stop codon of p-tubulin mRNA (see Fig. 5 for the nucleotide sequence). The major trans-splicing site for α-tubulin pre-mRNA is at nucleotide 530 [Fig. 5 and Sather and Agabian 1985]. [B] Northern blot analysis of transcripts of deletion constructs. RNAs from the indicated transfectants were hybridized to a CAT antisense RNA probe. Lane 1 contains RNA from mock transfected cells. Lanes 6 and 7 were run on a separate gel. Note that the relative abundance of the two shortened CAT RNA species was different between the T2 and T1 samples, but this varied between different transfection experiments.

denylation site of the PARP gene [Sherman et al. 1991] to provide signals for 3'-end formation of LUC mRNA. Northern analysis of the RNA derived from the plasmid PCATβ/αT2LUCP [Fig. 1A,B, lane 7] showed that also in this different genetic context the T2 deletion still induced the appearance of two size classes of CAT transcripts with faster electrophoretic mobility and in reduced amounts relative to wild type.

To determine whether the decreased amount of 3' shortened CAT RNAs originating from PCATβ/αT2LUCP [Fig. 1B, lane 7] was due to mRNA instability we compared the relative decay rates of T2 and wild-type CAT transcripts after transcriptional inhibition with the drug chloroquine. This drug has been shown previously to diffuse rapidly into trypanosome cells and readily halt RNA synthesis [Laird et al. 1987]. Having determined that transcription of the test constructs is ongoing for at least 4.5 hr after transfection in the absence of chloroquine [data not shown], we performed the experiment shown in Figure 2. Trypanosomes were transfected with constructs PCATβ/αLUCP, PCATβ/αT2LUCP or mock transfected, and after 3.5 hr transcription was inhibited by adding chloroquine to 1 mg/ml. The fate of the transcripts produced prior to drug addition were then assayed by Northern analysis over the following 4.0 hr. Confirming previous observations [Laird et al. 1987], a control hybridization demonstrated that the spliced leader RNA was rapidly turned over, being almost undetectable 30 min after the addition of the drug [Figs. 2A, B]. In contrast, the CAT transcripts from the transfected constructs PCATβ/αLUCP and PCATβ/αT2LUCP [Figs. 2D,E] appeared stable over the following 4 hr. Thus, stability of the shortened T2-CAT RNAs was playing no significant role in governing the final abundance of these transcripts. Instead, instability of the pre-mRNA comp-
bined with a lower efficiency of the RNA processing steps leading to mature mRNA were implicated.

The pyrimidine-rich sequences upstream of the α-tubulin 3'-splice site are required for β-tubulin poly(A)^+ site selection

To search for the specific signals responsible for the 3' shortened CAT transcripts, we generated block substitutions of plasmid PCATβ/αLUCP within the 48 bp defined by the endpoints of the T2 and T3 deletions. This region (nucleotides 486–534) has the interesting feature of encompassing the putative signals for trans-splicing of the downstream α-tubulin RNA (Sather and Agabian 1985; Patzelt et al. 1989), namely the branch sites, a long and a short polypyrimidine tract and the 3'-splice site itself (Fig. 3A). Initially we constructed six mutations, named BM1 to BM6, the structures of which are shown in Figure 3A. The different constructs were then transfected into trypanosomes and RNAs were analyzed by Northern blot hybridization, primer extension analysis and 3'-end RACE. We also measured CAT and LUC enzyme activities, but found that the values obtained for some of the transfectants were not proportional to the amount of mRNA detected by Northern or primer extension analysis [see below]. As an internal control for transfection efficiency and RNA recovery, a marked T. brucei U2 snRNA gene (Fantoni et al. 1994) was included in the transfections and RNA samples were normalized to the amounts of the expressed U2 snRNA. Figure 3D shows the results of a typical primer extension experiment with a U2-specific oligonucleotide as a primer.

To estimate the amount of CAT RNA produced we performed primer extension analysis with a labeled oligonucleotide complementary to nucleotides 26–39 of the CAT coding region (Fig. 3C). This showed that mutations BM1, BM2, BM3, BM4 and BM6 (lanes 3–7) had no significant effect on the accumulation of mature CAT mRNA, when compared with wild-type CAT mRNA (lane 2). Only mutant BM5 (lane 9), in which nucleotides 522–527 were changed, gave rise to ~30% less CAT RNA. This analysis also indicated accurate processing of the 5' end of the CAT transcripts by trans-splicing, since the predicted extension product of 156 nucleotides, in-
cluding the SL sequence, was detected in all samples, but not in the RNA derived from control cells (lane 1). Inspection of the size of the different CAT transcripts by Northern blot hybridization (Fig. 3B, lanes 4–9) further revealed no transcripts of the size range observed with the T2 deletion mutant (lane 3).

The fact that these substitutions, which changed >10 bp or less at the time, failed to reproduce the phenotype uncovered by the endpoints of the T2 and T3 deletions suggested that the β-tubulin 3'-end formation signals were perhaps redundant. Thus, we generated more extensive substitutions to give the two mutants BM7 and BM8 (Fig. 3A). BM7 combines mutations BM2 and BM3, and completely replaces the large polypyrimidine tract, whereas BM8 combines mutations BM7 and BM5 and substitutes both polypyrimidine tracts. Upon transfection both constructs gave rise to substantially lower amounts of CAT mRNA than wild type, as assessed by primer extension (Fig. 3C, lanes 10,11) and by Northern blot analysis (Fig. 3B: BM8, lane 10; BM7, lane11). More interesting, the size distribution of the corresponding CAT transcripts was seriously altered. The BM7 mutation (Fig. 3B, lane 11) gave rise to a heterogeneous collection of CAT RNA species ranging in size between the wild-type RNA of 1100 nucleotides and the shorter T2-derived CAT RNA of ~800 nucleotides (lane 3). In contrast, the size distribution of the BM8-CAT RNA (lane 10) resembled the pattern obtained with the T2 deletion (lane 3): Two size classes of CAT RNA were detected, both smaller than wild type. To exclude the possibility that the effects seen on 3'-end formation were due to the weakening of the 3'-splice site, we generated three additional mutants which combined BM1 and BM3, BM2 and BM6, and BM3 and BM6. Northern blot analysis and 3'-end RACE did not show any major alteration of the 3'-ends of CAT RNA derived from the mutant templates (data not shown). Furthermore, a fourth mutation which combined the polypyrimidine stretch mutation BM7 with the 3'-splice site mutation BM6 produced CAT RNA with a size distribution similar to that of BM7 RNA (data not shown). Thus, we concluded that the pyrimidine-rich sequences immediately upstream of the α-tubulin 3'-splice site were a major determinant for β-tubulin mRNA 3'-end formation.

To localize the polyadenylation sites of the CAT RNAs generated from the various mutant constructs we used 3'-end RACE, using a nested set of CAT oligonucleotides (see Materials and methods). Gel electrophoresis showed that the sizes of the amplified 3'-end cDNA fragments derived from transfectants BM1, BM2, BM3, BM4, BM5 and BM6 were identical to wild type, suggesting that these block substitutions did not affect the site of polyadenylation at a detectable level (data not shown). In contrast, mutants BM7, BM8 and T2 displayed 3'-end fragments of reduced sizes. As summarized schematically in Figure 4 and indicated in Figure 5, by sequence analysis of cloned cDNAs we identified five new regions or sites of polyadenylation [PAII, PAIII, PAIV, PAV and PAVI], in addition to the wild-type site [PAI]. With the exception of site PAIV which was identified by one BM7 cDNA clone at position 135, all other sites displayed poly(A)^+ addition at several closely spaced positions, as it occurs at the wild-type site. In general, there was good correspondence between the sizes of CAT RNA as determined by Northern analysis and the usage of different poly(A)^+ sites. For instance, the major T2- and BM8-CAT RNA poly(A)^+ sites are ~110- and ~320-nucleotides upstream from the wild-type poly(A)^+ site, and their relative position and distance from the wild-type poly(A)^+ site are consistent with the difference in size between the mutant and wild-type CAT mRNAs. On the other hand, the size of BM7 CAT RNA (Fig. 3B) predicted a considerable amount of wild-type size CAT RNA, but we failed to detect BM7 RNA 3'-ends at the wild-type PAI site. Although it is possible that this is an artifact of the amplification procedure, this does not impinge on the conclusion that sequences located between nucleotides 492 and 527, that is 114-bp downstream from the β-tubulin wild-type polyadenylation sites, were major determinants for specifying the position of the β-tubulin mRNA 3'-ends. In particular, obliteration of the polypyrimidine stretches immediately upstream of the 3'-splice site of the α-tubulin pre-mRNA shifted poly(A)^+ addition from the wild type position to new sites located as much as 370-bp upstream.
Figure 5. Sequence of the β/α tubulin region. The positions of the A residues used for polyadenylation at each poly[A] site are indicated by asterisks. Bars above the sequence indicate the extent of each poly[A] site as determined by sequencing cDNA clones. The AG dinucleotides that serve as wild-type and alternate 3'-splice sites are underlined. The wild-type 3'-splice site is denoted by an arrowhead; the branch sites (BS) are indicated by solid dots. PA, poly(A) site. PAI indicates the positions of the major and minor branch sites (BS) are indicated by solid dots. PA, poly(A) site. PAI indicates the positions of the major and minor branch sites (BS) are indicated by solid dots. PA, poly(A) site. PAI indicates the positions of the major and minor branch sites (BS) are indicated by solid dots.

The pyrimidine-rich sequences upstream of the α-tubulin 3'-splice site function in 3'-splice site identification

To investigate whether the polypyrimidine-rich sequences essential for accurate 3'-end formation had any effect on trans-splicing of the downstream LUC gene, we first analyzed the accumulation of LUC mRNA in trypanosome cells transfected with the various mutant constructs. Northern blot analysis with an antisense LUC RNA probe (Fig. 6) revealed that in all cases we could detect LUC RNA of a size indistinguishable from that of wild type, suggesting that the mutations did not abolish trans-splicing. Rather, as compared with the wild-type construct (lane 2), we observed a consistent, but certainly not dramatic reduction, in the accumulation of mature-size LUC mRNA detected by Northern analysis (Fig. 6). One likely explanation is that since T2- and BM7-derived LUC transcripts have altered 5'-ends (see below, and Table 1) made it very difficult to measure the amount of LUC RNA by primer extension analysis. Therefore, to obtain an indirect estimate of the amount of LUC mRNA produced by the various transfectants we determined luciferase protein activity in cell extracts [data not shown]; this analysis indicated that mutations BM4 and BM7 decreased luciferase activity by 25–40%, relative to wild type. The T2 and BM8 mutations had a more severe effect and led to 10- to 20-fold reduction in enzymatic activity. Clearly, these low yields of luciferase activity do not correlate with the abundance of mature size LUC mRNA detected by Northern analysis (Fig. 6). One likely explanation is that since T2- and BM8-derived LUC transcripts have altered 5'-ends (see below, and Table 1), this might decrease the efficiency of translation of these mRNAs. Thus, no firm conclusion about the abundance of LUC mRNA produced by the various transfectants can be drawn. However, it was clear from a number of experiments that the output of LUC mRNA from the BM7 and BM8 constructs was always considerably lower than wild type.

When we analyzed by 5'-end RACE, the effect of the BM series of mutations on the accuracy of trans-splicing at the α-tubulin 3'-splice site, we recovered cDNA clones indicative of aberrant 3'-splice site choice from mutants BM1, BM3, BM5, BM7, and BM8 (Table 1). In contrast, no evidence of alternate use of 3'-splice sites was obtained from the BM2 and BM4 cDNA clones. In contrast, RNA from the BM6 mutant was analyzed by
Table 1. Usage of AG dinucleotides for trans-splicing in wild-type and mutant PCATβ/αLUCP constructs

<table>
<thead>
<tr>
<th>Construct</th>
<th>Position of AG dinucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>530 wild type 5 545 551 556 579 585 588</td>
</tr>
<tr>
<td>T2</td>
<td>5</td>
</tr>
<tr>
<td>BM1</td>
<td>9 1 1 1</td>
</tr>
<tr>
<td>BM2</td>
<td>4</td>
</tr>
<tr>
<td>BM3</td>
<td>3 2</td>
</tr>
<tr>
<td>BM4</td>
<td>4</td>
</tr>
<tr>
<td>BM5</td>
<td>2 4 6</td>
</tr>
<tr>
<td>BM6</td>
<td>PE*</td>
</tr>
<tr>
<td>BM7</td>
<td>2 2 4 1</td>
</tr>
<tr>
<td>BM8</td>
<td></td>
</tr>
</tbody>
</table>

All of the BM mutants were first analyzed by primer extension. However, the interpretation of the results was not straightforward, except in the case of BM6. Therefore, the remaining BM mutants were analyzed by RACE. After amplification of cDNA 5' ends by PCR with specific oligonucleotides, DNA fragments were cloned and sequenced. The numbering of the AG dinucleotides is as indicated in Fig. 5. For each plasmid construct used for transfection the number of cDNA clones demonstrating trans-splicing at the various AGs is indicated.

*PE* Analyzed by primer extension.

primer extension (data not shown). This mutation, that changed the wild-type AG dinucleotide and two nucleotides upstream from it, abolished trans-splicing at this site with the concomitant use of the closest AG dinucleotide located 15 nucleotides downstream (Table 1). Where the long polypyrimidine stretch (mutant BM7) or the short-polypyrimidine stretch (mutant BM5) was substituted, we detected usage of both the wild-type α-tubulin and cryptic 3'-splice sites. However, when both the long- and the short-polypyrimidine tracts were changed (mutant BM8), we only recovered cDNA clones derived from RNAs trans-spliced at cryptic sites. Primer extension analysis (data not shown) was also consistent with the conclusion that in this mutant trans-splicing was predominantly aberrant.

Discussion

The experiments described here show that pyrimidine-rich sequences upstream from the T. brucei α-tubulin 3'-splice site are required to specify selection of the β-tubulin poly[A]+ site upstream and for the identification of the α-tubulin 3'-splice site downstream. Moreover, removal of these pyrimidine-rich motifs resulted in a considerable and simultaneous decrease of the accumulation of both CAT and LUC reporter RNAs, suggesting that not only the accuracy but also the efficiency of these two RNA processing reactions has been affected. The conclusion that the polypyrimidine sequences define an essential element for the identification of the β-tubulin poly[A]+ site is reinforced by considering that all BM mutations never altered the actual poly[A]+ site sequences nor their distance from other potential functional elements. Thus, the usage of the wild-type and cryptic β-tubulin poly[A]+ sites was always assayed in a competitive manner.

Pyrimidine-rich sequences and 3'-end formation

Deletion analysis and block substitution mutagenesis demonstrated that obliteration of the pyrimidine-rich sequences >100 nucleotides downstream from the wild-type β-tubulin poly[A]+ site led to aberrant 3'-end formation of the CAT mRNA through the activation of cryptic sites for polyadenylation. The phenotypes of mutations BM2, BM3, BM5, BM7, and BM8 indicated that ~10 of 20 pyrimidine residues of the long polypyrimidine tract were probably redundant for identification of the poly[A]+ site and that the short polypyrimidine tract was also dispensable, as long as the long polypyrimidine tract was intact. On the other hand, the effect of the short polypyrimidine tract on the choice of poly[A]+ sites was clearly revealed when both polypyrimidine elements were simultaneously changed (mutant BM8). At present the mechanism by which the short and long polypyrimidine tracts interact to aid in the selection of poly[A]+ sites is not clear and might be only clarified when an in vitro system becomes available. Surprisingly, the polypyrimidine stretches that are essential for 3'-end formation are located between 120 and 140 nucleotides downstream from the wild-type poly[A]+ site, whereas, for instance, in mammalian pre-mRNAs the downstream signals for 3'-end formation are found at a much shorter distance from the poly[A]+ site (Proudfoot 1991; Wahle and Keller 1992). It is unlikely that the polypyrimidine-rich sequences serve a role analogous to that of the AAUAAA/GU-rich sequence motifs required for accurate 3'-end formation in higher eukaryotes. This is because deletion of the latter sequences predominantly inactivates 3'-end formation rather than promoting utilization of cryptic poly[A]+ sites (Connelly and Manley 1988).

In comparing the sequences at and surrounding the cryptic poly[A]+ sites with the β-tubulin wild-type site, we found the only clear common feature to be that polyadenylation always starts before or after an adenosine residue. This is also observed in wild-type poly[A]+ sites of a number of trypanosome mRNAs (Tschudi and Ullu 1988) and is in agreement with the results of LeBowitz et al. (1993). In the case of the cryptic poly[A]+ sites, PAI, PAII, PAV, and PAVI we can identify polypyrimidine stretches that could potentially replace the wild-type elements, being at least 10 nucleotides long and being located between 100 and 140 nucleotides downstream from the corresponding poly[A]+ sites. Although further mutational analysis is required to establish whether or not these sequences function in poly[A]+ site identification, these cryptic poly[A]+ sites are probably poor substitutes for the wild-type site, since they are never utilized at a detectable level either for the endogenous β-tubulin mRNA or for the CAT mRNA produced upon
transfection of the PCATβ/αLUCP construct. It is conceivable that the poly(A)+ site itself and other sequence elements upstream and downstream from it might also contribute to poly(A)+ site selection.

In the case of the *Leishmania* DHFR-TS locus it has been proposed that the choice of the polyadenylation site is in part determined by the position of the 3′-splice acceptor site of the downstream gene (LeBowitz et al. 1993). Our results do not favor the hypothesis that the distance of the poly(A)+ site from the 3′-splice site is a primary determinant for poly(A)+ site identification. Although the phenotype of the T2 deletion mutant could support this model, we found that when the α-tubulin 3′-splice sites moved >50 nucleotides downstream from the wild-type position (mutant BM8), the β-tubulin poly(A)+ sites moved in the opposite direction, that is, upstream from the natural poly(A)+ site, as a result the distance between the two RNA processing sites was substantially increased. On the other hand, the results of the deletion analysis of the *Leishmania* DHFR-TS 3′-splice acceptor region are consistent with our finding of a functional role of polypyrimidine sequences in poly(A)+ site identification; unfortunately, a direct comparison between the two experimental systems is difficult at present, due to the different nature (deletion versus block substitution) of the two mutagenesis approaches.

### Pyrimidine-rich sequences and trans-splicing

There is evidence that the polypyrimidine tract in mammalian introns serves a dual function, first for recognition of the branch site and subsequently for 3′-splice site identification (Reed 1989). In yeast, instead, a pyrimidine-rich motif appears to be required for effective identification and utilization of only those 3′-splice sites that are distant from the branch site (Patterson and Guthrie 1991). Our studies with the trypanosome α-tubulin pre-mRNA clearly indicate that here one role of the pyrimidine-rich stretches is in aiding recognition of the 3′-splice site, thus further strengthening the similarity between *cis*-splicing and *trans*-splicing. However, sequence analysis of wild-type and cryptic 3′-splice sites (this paper, Huang and Van der Ploeg 1991; LeBowitz et al. 1993), indicates that the extent of the polypyrimidine sequences upstream of the invariant AG dinucleotide is quite variable among trypanosome 3′-splice acceptor sites. Functional studies showed that in the case of the *T. brucei* α-tubulin 3′-splice acceptor site, the length of the polypyrimidine tract can be substantially decreased without affecting the accuracy of *trans*-splicing or the accumulation of mRNA [this paper]. This is different from what has been reported for the PARP 3′-splice acceptor site of *T. brucei* (Huang and Van der Ploeg 1991) and could reflect intrinsic differences or perhaps different structural constraints between 3′-acceptor sites that are located immediately downstream from the promoter (PARP) relative to those that are within polycistronic pre-mRNAs (α-tubulin).

Substitution of nucleotides −13 to −6 (which include four uridines of the polypyrimidine tract) relative to the AG dinucleotide at the 3′-splice site of the PARP α gene reduced the expression of the CAT reporter gene to <10% of the wild-type level (Huang and Van der Ploeg 1991). In contrast, a similar mutation at the α-tubulin 3′-splice site (BM5), which substituted nucleotides −10 to −5, did not cause such a dramatic diminution of the expression of the LUC reporter gene as measured by enzymatic activity [data not shown] and by the accumulation of LUC mRNA [Fig. 6]. Differences in the turn-over rate and/or the rate of *trans*-splicing of the two pre-mRNAs might be responsible for this variability. It has been shown that only when the rate of splicing falls below the rate of degradation of the pre-mRNA, one observes a decrease of the steady-state accumulation of the corresponding mRNA (Pikielny and Rosbash 1985).

### Relationship between trans-splicing and 3′-end formation

Previously, we have shown that inhibition of *trans*-splicing by destruction of U2 snRNA results in inhibition of 3′-end formation/polyadenylation of the α-tubulin pre-mRNA as well as of the majority of trypanosome pre-mRNAs (Ullu et al. 1993). On the basis of these observations we proposed that there is a hierarchical order of reactions in trypanosome pre-mRNA processing, with *trans*-splicing preceding polyadenylation. Moreover, the recent paper by LeBowitz et al. (1993) supports the model that in *Leishmania* polycistronic pre-mRNA polyadenylation requires active *trans*-splicing or, minimally, recognition of the 3′-splice acceptor site. The evidence presented here further strengthens the notion of a close interaction or even a functional coupling between the *trans*-splicing and polyadenylation machineries.

The bifunctional nature of the polypyrimidine tract in RNA processing suggests at least two plausible mechanisms through which *trans*-splicing and polyadenylation might be coupled. For instance, the dual role of the polypyrimidine-rich sequences could be rationalized if they were sequentially recognized by different components of the *trans*-splicing and polyadenylation machineries. In polycistronic pre-mRNAs this could take place irrespective of whether *trans*-splicing at the downstream gene temporally precedes or follows polyadenylation at the upstream gene. After cleavage at the polyadenylation site the polypyrimidine sequences could still function for *trans*-splicing of the downstream gene and, vice versa, after cleavage at the 3′-splice site the same element could direct polyadenylation upstream. This mechanism would result in the apparent functional coupling of *trans*-splicing and polyadenylation because the two reactions would depend on a common functional element. Another possibility is that the primary role of the pyrimidine-rich sequences in the β/α-tubulin pre-mRNA is to bind a factor that is shared by the 3′-end formation and *trans*-splicing machineries, thus effectively coupling the two processes. This shared factor could be a protein or a protein complex or even a small
nuclear ribonucleoprotein particle (snRNP). A precedent for a protein involvement is the Drosophila tra-2 protein whose binding to dsk pre-mRNA appears to be necessary for sex-specific splicing and polyadenylation (Hedley and Maniatis 1991). Interestingly, recent in vitro cross-linking experiments suggest a role for the U1 snRNP in the coupling of cis-splicing and polyadenylation at terminal exons (Wassarman and Steitz 1993). In trypanosomes no U1 snRNA has been so far identified and it has been proposed that the spliced leader RNA is endowed with U1-like functions (Bruzik et al. 1988). It is tempting to speculate that the spliced leader RNA might have a dual function in trans-splicing and polyadenylation.

Support for a model coupling polyadenylation and trans-splicing also comes from other systems. For example, analysis of cis-splicing and polyadenylation at 3’-terminal exons of mammalian pre-mRNA has provided evidence that these two reactions are functionally interconnected (Niwa et al. 1990; Niwa and Berget 1991). Moreover, recent work from Blumenthal’s group has shown that also in Caenorhabditis elegans the polyadenylation and trans-splicing machineries of polycistronic transcription units may interact (Spieth et al. 1993). However, it can be anticipated that in trypanosomes trans-splicing and polyadenylation might also occur independently of each other. This might be the case for genes that are located at the beginning or at the end of a polycistronic transcription unit. Perhaps these genes possess some features that mark them as “terminal” genes for RNA processing, similar to the definition of a 3’-terminal exon in cis-splicing (Niwa et al. 1990, 1992, Niwa and Berget 1991).

In trypanosomes, the search for RNA polymerase II promoter sequences for protein-housekeeping genes has been unsuccessful so far. This has been taken as an indication that either this type of promoter is absent or very rare in the trypanosome genome. The current view is that in these organisms modulation of gene expression in terms of mRNA output on a per gene basis is primarily achieved post-transcriptionally by regulatory loops that involve pre-mRNA turn over in combination with differential rates of trans-splicing and polyadenylation, and mRNA turn over. In this scenario, it is conceivable that coupling of trans-splicing and polyadenylation within a polycistronic pre-mRNA could serve a number of different but not mutually exclusive roles. For instance, it could insulate that similar amounts of mRNAs coding for different or identical gene products are produced from polycistronic pre-mRNA or could simply increase the rate of production of mature mRNA 5’ and 3’ ends. This latter mechanism could be important to avoid premature degradation of the uncapped pre-mRNA 5’ end that is generated by cleavage at the polyadenylation site. An additional role for coupling the two reactions would be to overcome the absence of transcription termination within polycistronic transcription units, because there seems to be a tight association between 3’-end formation and primary transcript termination in other organisms (Connelly and Manley 1988; Proudfoot 1991; Wahle and Keller 1992).

Materials and methods

Trypanosomes

Procytic forms of T. brucei rhodesiense were used throughout and maintained as described previously (Ullu et al. 1993). Transfection was performed essentially as described by Sherman et al. (1991) except that the cell suspension was pulsed twice with a Bio-Rad gene pulser set at 1.5 kV and 25 μF.

Nucleic acids analyses

Total RNA was isolated by guanidinium thiocyanate/acid phenol extraction. Residual DNA was removed by digestion with RNase-free DNase I. Northern blotting was performed using standard procedures after separating the RNA by electrophoresis through a 1.2% agarose-formaldehyde gel. Blots were hybridized with antisense RNA probes prepared by in vitro transcription of either the CAT or the luciferase gene cloned into the vector pT3T7 (Boehringer Mannheim). Hybridizations were carried out at 55°C in 50% formamide, 5× SET (5× SET is 150 mM NaCl, 10 mM Tris-HCl at pH 7.5 and 1 mM EDTA), 5× Denhardt’s, 50 μg/ml of tRNA, 0.5% SDS, and washed at the same temperature in 0.1× SET, 0.5% SDS.

For primer extension analysis of CAT or U2 RNA, RNA from 1 to 5 × 10⁶ cell equivalents were used. Complementary specific oligonucleotides were labeled at the 5’ end with [γ³²P]ATP and polynucleotide kinase. After purification, total RNA was primer extended with a U2-specific oligonucleotide complementary to nucleotides 87–103 of the U2 snRNA or a CAT-specific oligonucleotide complementary to positions 26–39 of the coding region. The cDNAs were fractionated on a 6% polyacrylamide gel containing 7 M urea. For the U2 primer extension, this produced two cDNA products: one of 103 nucleotides, which is derived from endogenous U2 snRNA, and a second one of 117 nucleotides, which represents the transcripts originating from the marked U2 snRNA gene (Fantoni et al. 1994). Individual bands, corresponding to the transfected U2 RNA, were excised and the associated radioactivity determined by liquid scintillation counting. The RNA input of each sample to be used for further analysis was adjusted according to the U2 primer extension results.

cDNA synthesis for 3’- and 5’-end RACE was carried out with unphosphorylated oligonucleotides. The oligonucleotide used for the first strand synthesis for 3’-end RACE was PWMSANC (5’-CGGTGGCGACGAGCCCACCTTTTTTTTTTTTTTTTTT-3’). The resulting cDNAs were amplified with oligonucleotides PWMSeci (5’-CGAGAATTCGCTGCTGCGCACCAC-3’) and CATDN2 (5’-CGGTCTGTATGGCTTCCATG-3’, 80 nucleotides upstream from the CAT termination codon). A second round of amplification was carried out by nested PCR with oligonucleotide CATDN1 (5’-TGCAATCCCAGGGTCAAAC-3’, 17 nucleotides upstream from the CAT termination codon) and PWMSeci. For 5’-end analysis of LUC mRNAs, cDNA synthesis was primed with oligonucleotide LUC7 (5’-ATGCAAGTTGTCCTTCCAGGCG-3’, 81 nucleotides downstream from the ATG initiation codon of LUC). This was followed by amplification with the SL-specific oligonucleotide Eco-SL (5’-GGGAATTCCTCGATTATTAGAACGTTTCT-3’) and LUC7. Nested PCR was carried out with oligonucleotide LUCPRIME (5’-GAATGGCGGCGGCGCTTTT-3’, 43 nucleotides downstream from the LUC Eco-SL codon) and Eco-SL. PCR products were cleaved with the appropriate restriction enzymes, fractionated by gel electrophoresis, purified by conventional methods and cloned in the plasmid vector pT3T7. Individual clones were sequenced using the Sequenase system.

GENES & DEVELOPMENT 499
**Plasmid constructions**

All constructs were assembled from gene cassettes obtained by PCR amplification with specific oligonucleotides carrying unique restriction sites at their ends. The construction pathways and the sequence of the various specific oligonucleotides are available upon request. The PARP promoter and 3′-splice acceptor site (from nucleotide −399 to −1, relative to the CAT initiation codon), the CAT gene and the PARP gene 3′-untranslated sequence and poly(A) site (from the nucleotide after the CAT stop codon to 353 nucleotides downstream) were derived from plasmid pJP44 (Sherman et al. 1991), that is the nucleotide preceding the start codon of the α-tubulin gene was derived from a genomic lambda clone λGT804 [C. Tschudi, unpubl.] The LUC gene was a generous gift of B. Crise (Yale University, New Haven, CT). Unless otherwise stated, all constructs were in the same orientation in the plasmid vector pBluescript KS(−) (Stratagene, La Jolla, CA). The PARP promoter being proximal to the T7 RNA polymerase promoter and the 3′-PARP untranslated region being proximal to the T3 promoter.

**Mutagenesis** was performed in the dicistronic construct PCATβ/aLUCP. A NeoN/NotI restriction fragment spanning from 110 nucleotides upstream from the CAT termination codon to 50-nucleotides downstream of the LUC initiation codon was excised and subcloned into a pBluescript II KS(−) derivative which had been modified by insertion at the Smal site of a linker containing a NotI and NeoI restriction sites. The BM series of mutants was constructed by PCR-mediated mutagenesis. After cloning and sequencing the mutated fragments were inserted in the dicistronic construct, and resequenced.

**Acknowledgments**

We thank David Sherman and Christine Clayton for providing the pJP44 plasmid, Katherine Rogers for excellent technical assistance, Philippe Male for photography and Tim Nilsen and Nanci Steitz for valuable criticism on the manuscript. K.R.M. was supported by National Institutes of Health grant 5R01 GM 41161-02A1/A2. The plasmid construct was a generous gift of Drs. D. Sherman and C. Clayton. The β/α-tubulin region [from the nucleotide after the β-tubulin stop codon to nucleotide 643] [Imboden et al. 1986], that is the nucleotide preceding the start codon of the α-tubulin gene] was derived from a genomic lambda clone xGT804 [C. Tschudi, unpubl.] The LUC gene was a generous gift of B. Crise (Yale University, New Haven, CT). Unless otherwise stated, all constructs were in the same orientation in the plasmid vector pBluescript II KS(−) [Stratagene], with the PARP promoter being proximal to the T7 RNA polymerase promoter and the 3′-PARP untranslated region being proximal to the T3 promoter.

**References**


Pyrimidines in trans-splicing and polyadenylation


A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes.

K R Matthews, C Tschudi and E Ullu

Genes Dev. 1994, 8:
Access the most recent version at doi:10.1101/gad.8.4.491

References
This article cites 30 articles, 12 of which can be accessed free at:
http://genesdev.cshlp.org/content/8/4/491.full.html#ref-list-1

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.