MOLECULAR AND CELLULAR BIOLOGY, Nov. 1994, p. 7428–7435
0270-7306/94/$04.00+0

Vol. 14, No. 11

Hierarchies of RNA-Processing Signals in a Trypanosome Surface Antigen mRNA Precursor

MICHAEL HUG, HANS-RUDOLF HOTZ, CLAUDIA HARTMANN, AND CHRISTINE CLAYTON*

Zentrum für Molekulare Biologie, D-69120 Heidelberg, Federal Republic of Germany

Received 17 June 1994/Returned for modification 29 July 1994/Accepted 12 August 1994

Nearly all trypanosome mRNAs are synthesized as polycistronic precursors, from which mature mRNAs are excised by *trans* splicing and polyadenylation. Polyadenylation of a procyclic acidic repetitive protein (PARP, or procyclin) transcript was studied by transient transfection of constructs bearing a chloramphenicol acetyltransferase gene linked to the PARP intergenic region. Polyadenylation usually occurred at A residues, about 100 bases upstream of a *trans*-splicing acceptor signal. The wild-type polyadenylation site has a cryptic *trans*-splicing signal about 100 bp downstream: deletion or inversion of this signal results in polyadenylation at multiple sites, upstream of other cryptic *trans*-splicing signals. The PARP mRNA precursor appears to contain a hierarchy of possible processing signals, the function of cryptic ones being revealed only when the dominant ones are deleted or moved. Correct polyadenylation can be restored by addition of *trans*-splicing signals from other loci. The results indicate that polyadenylation is coupled to downstream *trans* splicing but that the products of the *trans*-splicing reaction are not necessarily functional mRNAs.

Members of the order Kinetoplastida are parasitic protozoa that diverged very early in eucaryotic evolution (33). Members such as *Leishmania* and *Trypanosoma* species have a mode of gene organization that is distinct from that of most other eucaryotes. Nearly all trypanosome mRNAs are synthesized as polycistronic precursors (reviewed in reference 7). Processing into single-gene units is effected by a combination of 5′ *trans* splicing and 3′ polyadenylation. The *trans*-splicing reaction involves addition of a capped 39-nucleotide (nt) sequence in a transesterification reaction (1). Many *trans*-splicing acceptor sites are preceded by polypyrimidine tracts which probably influence the specificity of the reaction (14), but such sequences cannot always be identified easily, and the precise sequence requirements for the reaction have not been defined.

No consensus signals for polyadenylation have been found within kinetoplastid 3′ untranslated regions. Recent results, however, suggest that polyadenylation may be coupled to *trans* splicing. Experiments analyzing the transcription and processing of trypanosome tubulin transcripts in a permeabilized cell system have demonstrated that *trans* splicing occurs cotranscriptionally and obligatorily precedes polyadenylation (35); functional mutations in the α-tubulin *trans*-splicing signal also affected polyadenylation in upstream β-tubulin sequences (22). (Results suggesting that polyadenylation precedes *trans* splicing, for a heat shock protein locus, were more difficult to interpret because the subcellular preparation used in the experiment was not capable of *trans* splicing [13].) Moreover, experiments with *Leishmania major* suggest that polyadenylation tends to occur at multiple sites, at a median distance of 390 nt upstream of a *trans*-splice acceptor site (18).

In *Trypanosoma brucei*, distances between one gene and the next vary considerably, some being as short as 150 bp and others much longer. The major surface protein of *T. brucei* multiplying within the tsetse fly is a procyclic acidic repetitive protein (PARP), also known as procyclin (reviewed in reference 8). In *T. brucei* MiTat 1, the PARP genes are arranged in pairs that are cotranscribed from an upstream promoter (reviewed in reference 7). The distance between the upstream (α) gene polyadenylation site and the downstream (β) gene splice acceptor site is about 600 nt. We find that processing at the PARP α gene polyadenylation site is more accurate than reported for *Leishmania* species, occurring at the same position for all cDNA clones characterized. Both the 3′ untranslated region and the intergenic region contain sequences that can function as splice acceptor sites and may influence the site of polyadenylation; the pattern of processing site usage is clearly influenced by context. Similar results have been reported independently by Schürch et al. (31).

MATERIALS AND METHODS

Construction of plasmids. All plasmids were derived from pJP 44 (32) or pHD 227 (15). To construct plasmid pHD 230, a PCR product synthesized by using primers CZ 286 (5′-ATT CCCGGGCTGCAGTTTAATTTGTTGGAT-3′) and CZ 225 (5′-GAGCATGCCTCGAGTGTAATTGAAGTCTTCAAG-3′), with a PARP A locus genomic clone as a template, was cut with *Pst*I and *Sph*I and cloned into correspondingly digested plasmid pJP 44. To construct plasmids pHD 117 and pHD 119, additional restriction sites were inserted at the *Bam*HI site of plasmid pJP 44. Unidirectional deletions were generated by exonuclease III digestion, followed by treatment with mung bean nuclease and Klenow fragment (12). Plasmids pHD 117 and pHD 119 contain, respectively, 198- and 284-bp deletions starting at the *Bam*HI site and extending to 99 or 13 bp upstream of the endogenous poly(A) site. Plasmid pHD 217 contains the deleted 3′ end of pHD 117 upstream of the pHD 230 intergenic region. Plasmid pHD 220 was constructed by deleting the 165 bp downstream of the *Pst*I site in pHD 230 by using exonuclease III as described above. Plasmid pHD 224 contains the 1,131-bp *Kpn*I-*Pst*I fragment of pHD 119 cloned into similarly digested plasmid pHD 220. Plasmids made by restriction digestion of pHD 230, end filling, and religation were pHD 231 (*Pst*I-*Hin*cII), pHD 232 (*Pst*I-*Apa*I), pHD 233 (*Pst*I-*Bsu*36I), pHD 241 (*Bst*BI-*Xho*I), and pHD 257 (*Pst*I-*Bst*BI). Plasmid pHD 240 was constructed by deleting 52 bp of pHD 230, using exonuclease III, starting from the *Pst*I site

* Corresponding author. Mailing address: Zentrum für Molekulare Biologie, Im Neuenheimer Feld 282, D-69120 Heidelberg, Federal Republic of Germany.

upstream up to 14 bp downstream of the endogenous poly(A) site. To construct pHD 243, a PCR product was made by using pHD 230 and primers CZ 286 (5'-ATTCCCGGGCTGCAGT TTAATTTGTTGGAT-3') and CZ 285 (5'-ATGAAGCTTAA GGAACTATTATTCGAACCC-3'). This product was cut with *Sma*I and *Hin*dIII and cloned into *Sma*I-*Hin*dIII-digested plasmid pHD 227. Plasmid pHD 244 was constructed by self-ligation of the 4,195-bp *Sma*I-*Hin*dIII fragment of plasmid pHD 227. Plasmid pHD 253 was made by cloning the linkers CZ 131 (5'-GCTCGAGTCGACTATTCCCAGGGCCCTGG CTGCA-3') and CZ 132 (5'-GCCAGGGCCCTGGGAATA GTCGACTCGAGCTGCA-3') into *Pst*I-digested pHD 230. Plasmids pHD 254 and pHD 255 were made by cloning, respectively, *Eco*RV-*Sma*I (98 bp) and *Sma*I-*Stu*I (330 bp) fragments of pUT 56 (Cayla, Toulouse, France) into the blunted *Pst*I site of pHD 230. In pHD 256, the 60-bp *Pst*I-*Bst*BI fragment of pHD 230 has been reversed in orientation. To construct plasmid pHD 258, a PCR product made by using pJP 44 and primers CZ 294 (5'-ATAAGCTTCTAGCTTGCAT GCCTGCAG-3') and CZ 295 (5'-ATCCCGGGCGTATTC CATAATTTTAAGCG-3') was cut with *Sma*I and *Hin*dIII and cloned into similarly digested plasmid pHD 227. Plasmid pHD 259 was made by cloning the 75-bp *Apa*I-*Hin*cII fragment of pHD 230 into *Sma*I-*Hin*dIII-digested pHD 227. Plasmid pHD 260 was constructed by cloning the 60-bp *Pst*I-*Bst*BI fragment of pJP 44 into *Pst*I-*Sph*I-digested pHD 227. Plasmids pHD 285, 286, and 287 have, respectively, the 86-bp *Sma*I-*Hin*dIII fragment of pJP 44, the 192-bp *Sma*I-*Hin*dIII fragment of pHD 226 (15), and the 111-bp *Sma*I-*Hin*dIII fragment of pHD 227 cloned into the *Pst*I site of pJP 44. Plasmid pHD 302 was constructed by performing a PCR using plasmid pHD 1 (15) as a template for primers CZ 017 (5'-CCGGAATGCTAATGT CACTGCGAG-3') and CZ 304 (5'-AATAGAAGTGATCT AGAGTAAATTATGGAA-3'). The product was reamplified with plasmid pHD 230 and primers CZ 064 (5'-GTGGGTCA CTGTACTATT-3') and CZ 285 (5'-ATGAAGCTTAAGG AACTATTATTCGAACCC-3'). The product of this second PCR was restriction digested and cloned into pHD 230. Plasmid pHD 303 was made in a similar way, starting with primers CZ 017 and CZ 307 (5'-GCACCAATAATAAGCG TAAGATCCAAATTC-3').

Transfections of trypanosomes. Procyclic *T. brucei* (strain Antat 1.1) was transfected as previously described (32). For RNA analysis, 3 to 10 cuvettes (3×10^7 trypanosomes per cuvette) were transfected with 15 to 40 μg of plasmid DNA per cuvette, and RNA was isolated 8 h later. For transient chloramphenicol acetyltransferase (CAT) assays, 3×10^7 *T. brucei* cells were transfected with 10 μg of plasmid DNA, and CAT assays were performed 18 to 24 h later (4). Each construct was assayed at least four times.

RNA analysis. Total RNA was isolated as previously described (32). Blot hybridization was done using poly(A)$^+$ RNA (QuickPrep mRNA purification kit; Pharmacia, Freiburg, Germany) and Hybond N+ membranes (Amersham Life Science, Braunschweig, Germany); the CAT mRNAs were detected by enhanced chemiluminescence (ECL Kit; Amersham Life Science). The sites of miniexon addition and poly(A) sites were mapped, using total RNA as the substrate, for reverse transcription and PCR (RT-PCR) (15), with the modification that for mapping of poly(A) sites, CZ 205 (5'-CCATGTCGGCA GAATGCTTA-3') instead of CZ 204 was used in the second PCR.

For the mapping of the 5' termini, primer CZ 296 (5'-AGC ACCTTGTCGCCTTGCGTATAA-3') was used to prime the reverse transcriptase reaction. The first PCR was performed with the miniexon primer and CZ 296, and the resulting product was cut with *Pvu*II. The second PCR was done with the miniexon primer and CZ 216 (5'-TATCACCAGCTCACC GTCTT-3'). The resulting product was cut with *Hin*dIII and *Eco*RI and cloned into *Eco*RI-*Hin*dIII-digested pBS SKII+ (Stratagene, Heidelberg, Germany).

To detect *trans*-spliced RNA molecules derived from pJP 44 vector sequences, a reverse transcriptase reaction using RNA from plasmid pJP 44-transfected procyclic *T. brucei* and oligonucleotide CZ 311 (5'-AGTCCCGGGTCATGCCATCCGT AAGATG-3') was performed. With the resulting first-strand cDNA, a PCR was done with oligonucleotide CZ311 and the miniexon-specific primer Z891039 (5'-CGGAATTCAACG CTATTATTAGAACAGTTTCTGTACT-3'). The resulting DNA was cut with *Pvu*II, and a second PCR was performed with oligonucleotides CZ 310 (5'-AAGCGGTTAGCTCCT TCGGT-3') and Z891039. The product was cut with *Sma*I and *Eco*RI, cloned into *Eco*RI-*Sma*I-digested pBluescript SKII+ (Stratagene), and sequenced by using Sequenase (U.S. Biochemical, Div. Amersham Life Science).

To detect *trans*-spliced RNA molecules derived from sequences of the 3' untranslated region of pJP 44, a reverse transcriptase reaction was performed with RNA from plasmid pJP 44-transiently-transfected procyclic *T. brucei* and oligonucleotide CZ 313 (5'-CTGGATCCTAGAGGATCTGGCTAG CGAT-3') followed by a PCR using oligonucleotides CZ 313 and the miniexon-specific primer Z891039. The PCR product was cut with *Nae*I, and a second PCR was performed with oligonucleotides Z891039 and CZ 314 (5'-CTGGATCCTA GAGGATCTGGCTAGCGAT-3'). The product from this second PCR was cut with *Eco*RI and *Bam*HI, cloned into *Eco*RI-*Bam*HI-digested pBluescript SKII+ (Stratagene), and sequenced. PCR conditions were similar to those described for 5'-terminus mapping (15).

RESULTS

RT-PCR cloning of RNA species. To find out what sequences were sufficient to guarantee accurate polyadenylation, we transfected trypanosomes with various derivatives of a plasmid expressing the CAT reporter gene. Transcription was driven by the PARP promoter, and 5' processing was mediated by the PARP Aα *trans*-splicing acceptor site. Supercoiled DNA was transfected into trypanosomes, and RNA was purified 8 h later. The approximate sizes of RNAs produced from selected clones were measured by formaldehyde gel electrophoresis and blot hybridization. To determine polyadenylation sites more accurately, the RNAs were reverse transcribed with an oligo(dT) primer and then subjected to a nested PCR designed to amplify only mRNAs containing the CAT coding sequence. (Molecules lacking the 3' half of the CAT cassette would not be detected.) The PCR products were cloned [by using restriction enzyme sites included within the transfected DNA and on the end of the oligo(dT) primer] so that polyadenylation positions could be pinpointed by sequencing for a small sample of the RNA population.

Accurate polyadenylation is achieved by inclusion of a downstream polypyrimidine tract. To obtain CAT mRNA bearing a PARP 3' untranslated region, we cloned the entire region, from the PARP termination codon to 58 bp downstream of the polyadenylation site (known from cDNA clones [25, 26]), between the CAT gene and the pGEM vector sequence (pJP 44 [32]). The fragment was bounded by synthetic *Bam*HI and *Pst*I sites (Fig. 1 to 3). To our surprise, we found that RT-PCR products made from pJP 44 template RNA were about 100 bp shorter than anticipated (reference 15 and Fig. 2b). Sequencing of nine independently cloned cDNAs

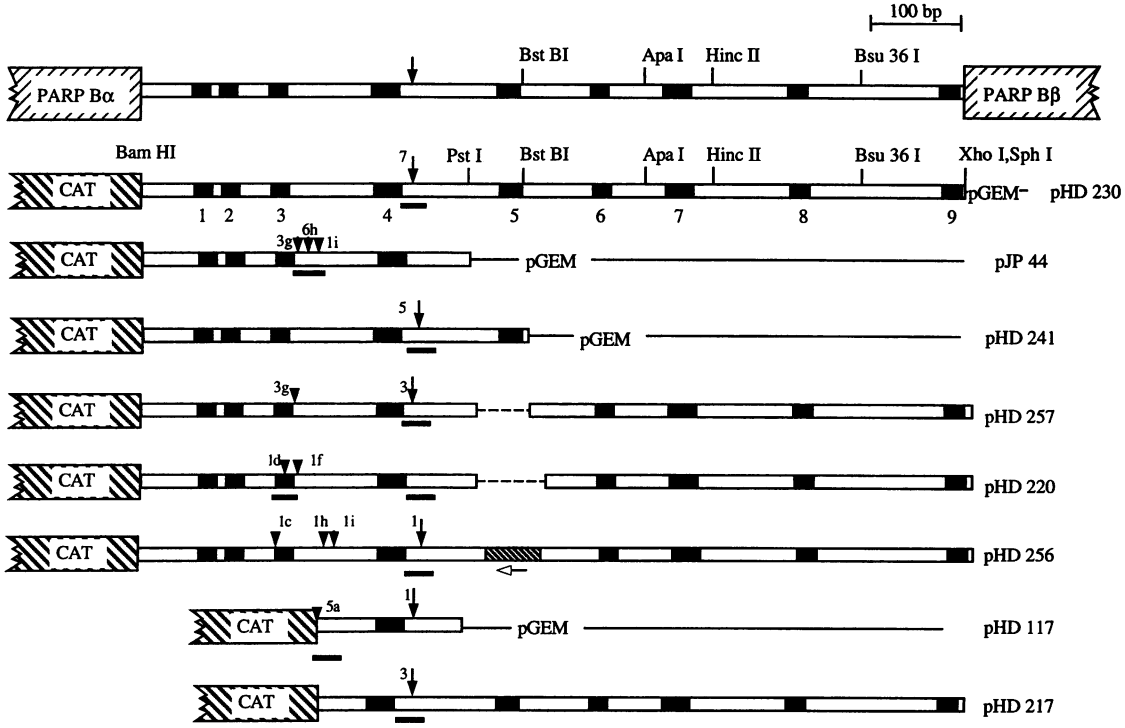

FIG. 1. Structures of the genomic PARP A intergenic region (top) and of the corresponding CAT expression construct (pHD 230). Dashed lines are deletions. Polypyrimidine tracts (defined as a sequence of at least 20 nt with no consecutive purines and at least 80% pyrimidines) are black boxes, and the wild-type polyadenylation site has an arrow. Other observed polyadenylation sites are indicated by filled arrowheads. The number of times each site was observed and the precise position of each site are indicated by numbers and by lowercase letters that are also found in Fig. 3. Short wide bars indicate the predominant PCR products which are displayed in Fig. 2b. The cross-hatched segment (pHD 256) is in the reverse orientation.

revealed polyadenylation at three different sites, 101, 113, and 124 nt upstream of the normal position (Fig. 1, pJP 44; Fig. 3). No wild-type polyadenylation was detected. Either something necessary for wild-type polyadenylation was missing from this construct or else the pGEM vector sequence was having a dominant disruptive effect.

Knowing that *trans* splicing was important for polyadenylation in *Leishmania* species (18), and that there was no trypanosome-derived *trans*-splicing site present downstream of the 3′ end in pJP 44, we decided to restore the intergenic region to its original state by inserting the remaining intergenic sequence down to the start codon of the next gene (PARP Aβ) (pHD 230; Fig. 1 and 3). The RNA synthesized from this plasmid was about 100 nt longer (about 1.4 kb) than that arising from pJP 44 (about 1.3 kb; Fig. 2a), as was the RT-PCR product (Fig. 2b). Every pHD 230 RT-PCR product sequenced was polyadenylated in the wild-type position. A similar result was obtained with constructs lacking two-thirds of the 3′ untranslated region: RT-PCR products obtained after transfection with pHD 117 were mostly polyadenylated on the synthetic *Bam*HI site, wild-type polyadenylation being restored by addition of the intergenic region (pHD 217).

To determine whether *trans* splicing was influencing polyadenylation, we first searched for potential *trans*-splicing signals between the PARP α and β genes. We found nine polypyrimidine tracts, defined arbitrarily as any sequence of at least 20 nt that is more than 80% pyrimidine rich and has no dipurine interruptions. These are labeled 1 to 9 in Fig. 1 and 3. In the plasmid constructs, the next such sequence occurs 450 bp into the pGEM vector. To find out which sequences were required for accurate polyadenylation, we made progressive deletions from the downstream end of the intergenic region. The inclusion of 125 bp downstream of the original polyadenylation site was sufficient to ensure its use (pHD 241; Fig. 1 and 3), yielding mRNA and RT-PCR products of the same sizes as those produced by pHD 230. This 125-bp sequence includes a potential splice acceptor sequence, with a polypyrimidine tract (labeled 5 in Fig. 1 and 3) and an AG dinucleotide (@5 in Fig. 3).

Inversion of polypyrimidine tract 5, yielding a polypurine tract containing oligo(A), resulted in a predominance of shorter RNAs (pHD 256; Fig. 1 and 2b). Deleting it (pHD 257 and 220; Fig. 1 and 3) also clearly disturbed the accuracy of polyadenylation, resulting in shorter products. This short segment clearly plays a role in determining polyadenylation at the wild-type position.

We next made progressive internal deletions in the intergenic region starting from the 5′ end. Clones having longer deletions than pHD 257 showed some accurate polyadenylation once more (Fig. 4, pHD 232, 231, and 233), although some shorter products may also have been present (Fig. 1 and 2a). Finally, a construct with a double deletion that removed DNA to within 13 bp of the wild-type polyadenylation site, and the polypyrimidine tract (tract 5) just downstream (Fig. 2 to 4, pHD 224), yielded predominantly mRNAs polyadenylated about 100 nt upstream of the PARP β gene splice acceptor signal (signal 9) and some polyadenylated a little downstream of the wild-type site.

Does the PARP intergenic region contain cryptic splice sites? In most cases so far, polyadenylation seemed to occur at

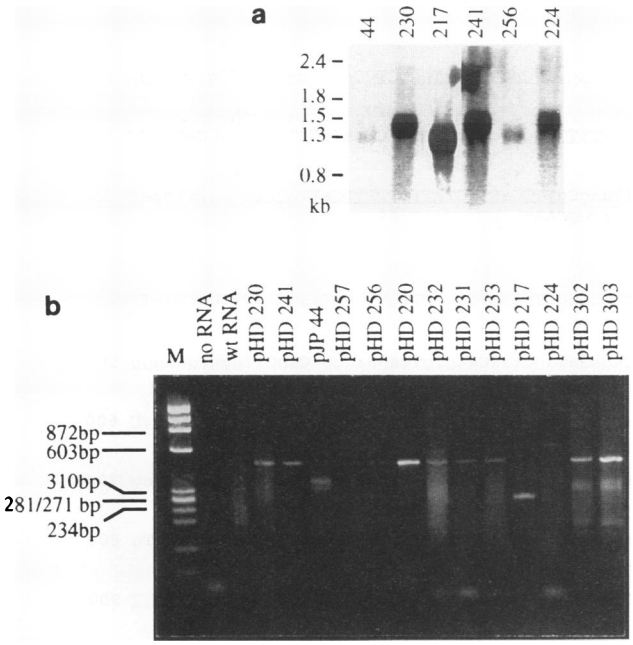

FIG. 2. (a) Northern blots of CAT mRNAs from selected constructs. For each construct, three cuvettes were each transfected with 50 μg of DNA. Poly(A)$^+$ RNA was isolated, and one half of the yield was subjected to formaldehyde gel electrophoresis and blot hybridization using a CAT probe. Yields of RNA were variable, and the intensities of the bands do not necessarily reflect the efficiencies of the constructs; exposures are adjusted for clarity. The name of the transfecting plasmid is noted above each lane. The average sizes (in kilobases) of the detected transcripts and (in parentheses) the sizes predicted from the RT-PCR results for the transcripts [without poly(A) tails] are as follows: pJP 44, 1.3 (1.0); pHD 230, 1.4 (1.2); pHD 217, 1.2 (0.9); pHD 241, 1.4 (1.2); pHD 256, 1.3 (1.0); and pHD 224, 1.6 (1.3). (b) Agarose gel electrophoresis of PCR products. Each lane contained 10 μl of 100-μl reactions and was stained with ethidium bromide. Markers are on the left. The RNA came from trypanosomes transfected with the plasmids indicated.

A residues, 100 bp or so upstream of a polypyrimidine tract. Thus, wild-type polyadenylation appeared to be signalled by sequence 5, that for pJP 44 and pHD 220 appeared to be signalled by polypyrimidine 4, and that for pHD 233 and 224 appeared to be signalled by sequence 9. Polypyrimidine 4 predominated in signalling polyadenylation only in the absence of polypyrimidine 5. Polyadenylation for plasmids pHD 231 and 232 might have been influenced by polypyrimidines 7, 8, or 9 or by some other sequence not recognized by our criteria. We therefore examined whether the ability of a sequence to influence polyadenylation correlated with its ability to signal *trans* splicing.

The PARP Aβ splice acceptor is highly homologous to that for PARP Aα, and the A-locus signals are again homologous to those in the B locus. As the PARP Bα polypyrimidine tract is involved in splicing (14), polypyrimidine 9 probably is as well. The polyadenylation seen for pHD 224 (Fig. 4), about 100 nt upstream of polypyrimidine 9, was therefore not unexpected. As the precise sequence requirements for *trans* splicing are not known, we wished to determine whether any of the other polypyrimidine tracts within the PARP intergenic region could function as a splicing signal. To do this, we cloned three of them, 4, 5, and 7, upstream of the CAT gene, replacing the original actin splice acceptor sequence in plasmid pHD 227 (which has 5′ and 3′ sequences from the actin locus) (15) (Fig. 5). Polypyrimidine tract 7 (pHD 259; Fig. 5) yielded only 2% of the CAT activity of pHD 227. (A negative control lacking a splice acceptor site, pHD 244, yielded no detectable CAT activity.) In contrast, polypyrimidine tract 5, which is sufficient to restore polyadenylation, functions well in *trans* splicing (Fig. 5, pHD 243), yielding products spliced at the nearest downstream AG dinucleotide (@5 in Fig. 3). More surprising was the observation that polypyrimidine tract 4 (Fig. 5, pHD 258) can also specify *trans* splicing with equivalent efficiency. This is interesting for two reasons: first, the sequence is normally within a mature 3′ end, and second, the nearest AG dinucleotide in the construct is more than 80 nt downstream.

RT-PCR experiments were then designed to detect *trans* splicing within the intergenic region in normal procyclic trypanosome RNA. A single clear band was obtained, indicative of splicing at site 5 and consistent with published nuclease protection results (see Discussion). We obtained no evidence that any sequence between polypyrimidine tracts 5 and 9 was capable of directing splicing.

If, in plasmids such as pJP 44 which lack the intergenic region, polypyrimidine 4 is directing polyadenylation, and if the polyadenylation is coupled to *trans* splicing, we would expect to be able to detect the appropriate *trans*-spliced product (spliced at position @4 in Fig. 3). This product was indeed detected by RT-PCR using a primer homologous to the pGEM sequence in combination with a miniexon primer. Thus, polyadenylation of the pJP 44 CAT mRNA in the region around nucleotide 200 (positions g, h, and i in Fig. 3) was correlated with *trans* splicing using signal 4.

Is polyadenylation specified by a scanning mechanism? LeBowitz et al. (18) showed that in *L. major*, decreasing the distance between an upstream gene and a downstream splice signal moved polyadenylation upstream. They suggested that polyadenylation is determined by assembly of the splicing apparatus on the splice site followed by scanning upstream to the nearest acceptable polyadenylation site, which was normally an average of 380 to 390 bp upstream. Our results so far were consistent with this idea, except that the intergenic distances in *T. brucei* were usually shorter. Also, polyadenylation at the wild-type PARP site appeared much more precise than in the *L. major* gene studied: neither we nor Schürch et al. (31) have ever seen polyadenylation anywhere else when the downstream splice site is present (12 PCR products from this study, plasmids pHD 230 and 241, plus other data not shown; 20 PCR products from reference 31).

If spacing between splice sites and polyadenylation sites plays a role in processing, one would expect not only that decreasing the distance between splice signal 5 and the wild-type polyadenylation site would affect polyadenylation but also that increasing the distance would have a similar effect. The results that we obtained showed that this was at least partially true. Decreasing or increasing the distance between the wild-type polyadenylation site and splice signal 5 decreased the accuracy of polyadenylation (Fig. 6, pHD 240, 253, 254, and 255). The inserted sequences were GC rich and lacked polypyrimidine tracts; no polyadenylation within them was detected.

Is a polypyrimidine tract necessary and sufficient for accurate polyadenylation? To determine whether any splice acceptor was sufficient to determine polyadenylation, we inserted splice acceptor sites from other trypanosome genes downstream of the PARP 3′ end (Fig. 7). Polyadenylation was

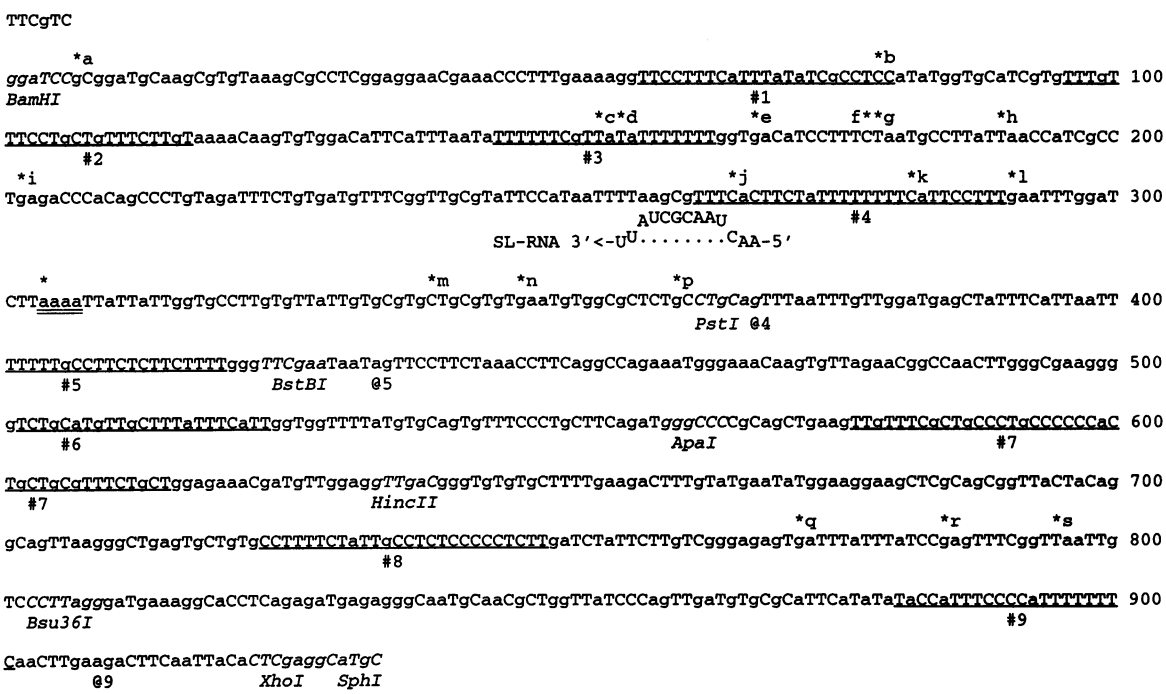

FIG. 3. Sequence of the reconstructed PARP intergenic sequence in pHD 230. The sequence is shown with pyrimidines in uppercase and purines in lowercase. Polypyrimidine tracts (1 to 9) are underlined. Restriction sites are in italics. The *Bam*HI, *Pst*I, *Xho*I and *Sph*I sites are synthetic sites introduced during cloning. @, splice acceptor sites identified experimentally; *, last nucleotide before a poly(A) tail. Each site (apart from the wild-type site) is identified by a lowercase letter correlating with the labeling in Fig. 1, 4, and 6. The wild-type site is double underlined.

obtained in the expected position when splice sites from either PARP Aα (pHD 285), aldolase (pHD 286), or actin (pHD 287) were used, although accuracy was diminished in the latter two cases. If the CAT gene is followed at the 3′ end by a fragment from the actin intergenic region, including the polyadenylation site but lacking the downstream splice site, polyadenylation occurs upstream of the normal position for actin transcripts (15). In this case, insertion of a fragment containing PARP intergenic splice signal 5 was sufficient to restore normal polyadenylation at the actin polyadenylation sites (pHD 260; Fig. 1).

Trypanosome polyadenylation sites, both published and from this work, share no consensus apart from a preference for A residues, but polyadenylation at other nucleotides is clearly possible. To review the role of the four A residues at the wild-type polyadenylation site, we mutated the sequence TTA AAATT to TTACGCTT (pHD 303). Of the six PCR products sequenced, one was now polyadenylated 12 nt upstream of the normal position (position 1 in Fig. 3). A sequence complementary to the trypanosome spliced leader (SL) RNA is present in the PARP Bα 3′ untranslated region (Fig. 3). Neither mutation (pHD 302; Fig. 2b) nor deletion (31) of this complementarity prevented correct polyadenylation, and it is not conserved in the PARP β genes, suggesting that it plays either no role or a marginal one in determining processing.

DISCUSSION

The results in this report suggest that polyadenylation in trypanosome transcripts is coupled to downstream *trans* splic-

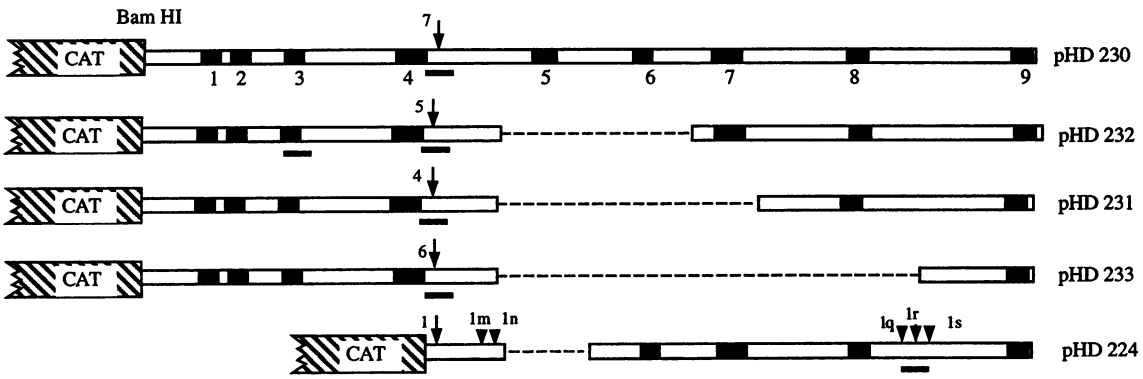

FIG. 4. Effects of further deletions in the intergenic region. Details are as for Fig. 1.

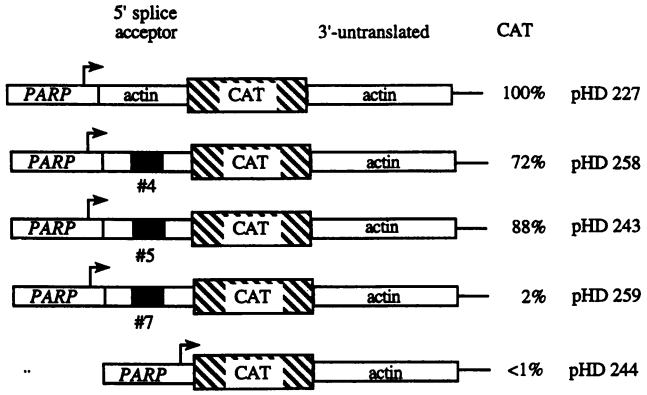

FIG. 5. Function of polypyrimidine tracts in splicing. Plasmids are represented schematically, not to scale. *PARP*, PARP promoter. Other symbols are as for Fig. 1.

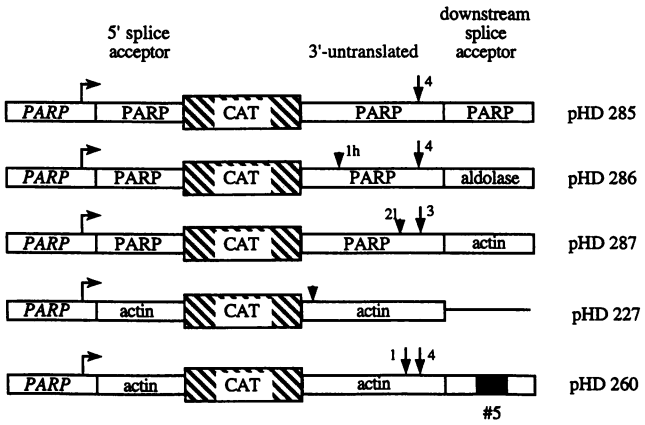

FIG. 7. Function of splicing signals in polyadenylation. Plasmids are represented schematically, not to scale. Other symbols as for Fig. 1 and 5.

ing. This is consistent with the results of kinetic analyses of RNA processing in permeabilized trypanosomes, which indicated that *trans* splicing is cotranscriptional and precedes downstream polyadenylation (35). In the absence of *trans* splicing, polyadenylation was largely inhibited. The results are also in agreement with those obtained for an *L. major* transcript (18) and with others obtained independently for the PARP locus (31, 36) and from mutational analyses of a *T. brucei* tubulin intergenic region (22). In general, the intergenic regions in trypanosomes appear to be shorter than for *Leishmania* species; we also find that the *trans*-splicing signal used to determine trypanosome polyadenylation need not necessarily precede a translated open reading frame. Manipulation of transfected plasmid sequences revealed that 3′ untranslated sequences may contain cryptic *trans*-splicing signals whose products are not normally detectable in mature RNA.

Two principal types of mechanism can be envisaged for a dependence of polyadenylation upon downstream *trans* splicing. One possibility is that both reactions take place in a coordinated fashion, through assembly of a bifunctional RNA processing complex, or through a requirement for a subset of spliceosome components for polyadenylation. This model predicts a polyadenylation site-specific cleavage reaction. Alternatively, the product of the *trans*-splicing cleavage reaction may become a substrate for a nonspecific 3′ → 5′ exonuclease activity and poly(A) polymerase. The latter model predicts that the final intergenic distance (100 to 200 bp for most trypanosome genes; 390 bp for *Leishmania* species) would be determined mainly by the relative efficiencies of exonuclease and poly(A) polymerase. Either process could be affected by secondary structure as well as by as yet unrecognized sequence elements within the precursor. Neither we nor Schürch et al. (31) obtained any evidence that the sequence within the PARP 3′ untranslated region had any role in determining polyadenylation.

A predilection for A residues (especially several consecutive A's) at polyadenylation sites is clear both from this study and from an examination of the literature. That other elements may also play a role is suggested by evidence from two of our mutants (pHD 254 and 255): when segments lacking polypyrimidine tracts (and therefore containing plenty of A's) were inserted between the wild-type polyadenylation site and the splice site immediately downstream, we failed to find polyadenylation within the inserted sequence.

If our interpretation is correct, the structure and abundance of mature RNA products in trypanosomes are determined by the relative strength of *trans*-splicing signals and the stability of the resulting products. Trypanosome noncoding regions in general are very rich in polypyrimidine tracts. It seems likely that although only a minority appear important for generation

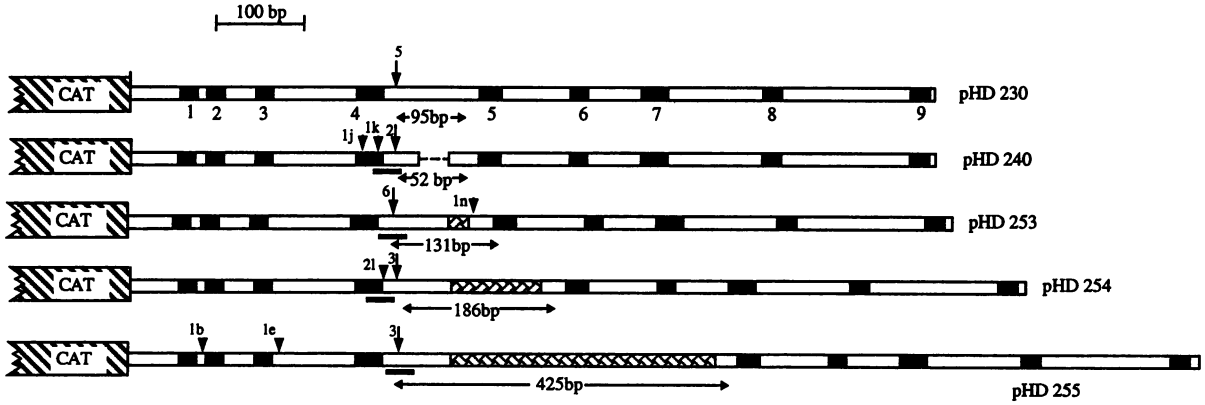

FIG. 6. Insertions and deletions reduce polyadenylation accuracy. Details are as for Fig. 1. The patterned segments are fragments of the gene encoding phleomycin-binding protein, lacking polypyrimidine tracts. Distances shown are from the wild-type polyadenylation site to the start of the downstream polypyrimidine tract.

of stable *trans*-spliced mRNAs, many more are potentially functional. It is not yet clear what constitutes a strong *trans*-splicing signal in trypanosomes. Although polypyrimidine tracts definitely contribute, we do not know why one should constitute a stronger signal than another, whether sequences at branch points play a role, and what distances are allowed between the polypyrimidine tract and the first potential splice acceptor AG dinucleotide. In the PARP 3′ untranslated and intergenic region, we identified nine polypyrimidine tracts. In normal trypanosomes, only two of these appear to be used for splicing. Splice site 9 yields functional PARP β mRNAs. The product of splice site 5 is a minor PARP β RNA with a 300-nt 5′ extension which can be detected on very overexposed Northern (RNA) blots (6a), by primer extension analysis of RNA from UV-irradiated trypanosomes (29), and by RT-PCR (reference 31 and this report). Splice site 5 also directs normal PARP α polyadenylation. Although splice site 9 can direct upstream polyadenylation in artificial constructs (e.g., pHD 224; Fig. 4), neither PARP α mRNAs with elongated 3′ ends nor short intergenic RNAs are detectable. Either these latter species are unstable (the intergenic species would lack a long open reading frame [3, 16]) or they are not made. The activity of polypyrimidine tract 4 in both splicing and polyadenylation is revealed only after deletions or rearrangements (Fig. 1, pJP 44, pHD 220, and pHD 256; Fig. 5, pHD 258). The results of transient assays argue against the idea that the pJP 44 truncated RNAs are highly unstable; therefore, we conclude that although splice signal 4 is equivalent to splice signal 5 in artificial constructs, it is less efficient in the natural context, perhaps because of an 89-nt separation between the polypyrimidine tract and the splice acceptor site. (Normally, the separation is considerably smaller; it is 15 nt for splice site 5.) No clear function of the other polypyrimidine tracts in the intergenic region was detected. Overall, these results suggest a natural hierarchy of processing signals in trypanosome precursor RNAs. Normally, one *trans*-splicing signal predominates in determining polyadenylation. Others may be present but are subordinate, in the sense that their function is revealed only when the dominant signal is deleted or moved.

Both splice site choice and polyadenylation have been implicated in regulation of gene expression in higher eucaryotes (10, 19, 21, 23). In kinetoplastids, *trans* splicing has been postulated as a target for trypanosome developmental regulation (17, 38), and complex RNA processing patterns are not infrequent. Use of alternative polyadenylation sites has been observed in artificial constructs involving the 3′ ends of the actin, β-tubulin, and dihydrofolate reductase genes (6, 15, 18). The region just upstream of the IsTaR 1 variant surface glycoprotein gene contains a variety of alternative splicing and polyadenylation sites and is represented as several overlapping mature mRNA species (30). Use of alternative splice sites is seen for the ESAG 6 and ESAG 7 genes, and an ESAG 6 cDNA with an extended 5′ untranslated region is *trans* spliced just downstream of the ESAG 7 polyadenylation site (29). The variant surface glycoprotein genes lie upstream of telomeres, so there is no downstream intergenic region. Nevertheless, the polyadenylation sites are situated about 80 bp upstream of a conserved polypyrimidine that could well serve as a splice acceptor sequence (2, 9, 24).

Interactions between splicing and polyadenylation have now been documented for a variety of higher and lower eucaryotes. The internal genes of polycistronic *Caenorhabditis elegans* transcripts are preferentially *trans* spliced by using a specific SL variant, SL-2 (34), implying that the *trans*-splicing apparatus has a mechanism for recognizing upstream polyadenylation sites or 3′ untranslated regions. The SL RNA is thought to be the *trans*-splicing equivalent of U1 RNA (5). For the *cis*-spliced, monocistronic transcripts of higher eucaryotes, interactions are seen between the polyadenylation site and the proximal upstream intron. Insertion of a 5′ splice donor site too near a polyadenylation signal can inhibit polyadenylation (28); conversely, mutation of the AAUAAA polyadenylation signal can depress splicing of proximal introns (27). In contrast with the situation in trypanosomes, a base-pairing interaction between the U1 small nuclear ribonucleoprotein and sequences in 3′ untranslated regions (20, 37) may be involved. Alternatively or in addition, interactions between splicing components and the polyadenylation apparatus (11) may explain the apparent coupling.

ACKNOWLEDGMENTS

We thank Isabel Roditi (Universität Bern), Steve Beverley and John Lebowitz (Harvard University), Elisabetta Ullu and Keith Matthews (Yale University), and Hui-Min Ching and Lex van der Ploeg (Merck Research Laboratories) for communicating unpublished results and Isabel Roditi, Steve Beverley, Elisabetta Ullu, and Joan Steitz (Yale University) for interesting discussions.

This work was supported by the Bundesministerium für Forschung und Technologie and the Deutsche Forschungsgemeinschaft.

REFERENCES

1. **Agabain, N.** 1990. *Trans* splicing of nuclear pre-mRNAs. Cell **61:**1157–1160.
2. **Aline, R. F., and K. Stuart.** 1989. Trypanosoma brucei: conserved sequence organisation 3′ to telomeric variant surface glycoprotein genes. Exp. Parasitol. **68:**57–66.
3. **Atwater, J. A., R. Wisdom, and I. M. Verma.** 1990. Regulated mRNA stability. Annu. Rev. Genet. **24:**519–541.
4. **Beverley, S. M., and C. E. Clayton.** 1993. Transfection of *Leishmania* and *Trypanosoma brucei* by electroporation, p. 333–348. *In* J. E. Hyde (ed.), Protocols in molecular parasitology. Humana Press, Totowa, N.J.
5. **Bruzik, J. P., and J. A. Steitz.** 1990. Spliced leader RNA sequences can substitute for the essential 5′ end of U1 RNA during splicing in a mammalian *in vitro* system. Cell **62:**889–899.
6. **Chung, H.-M. M., M. G.-S. Lee, P. Dietrich, J. Huang, and L. H. T. van der Ploeg.** 1993. Disruption of the largest subunit RNA polymerase II genes in *Trypanosoma brucei*. Mol. Cell. Biol. **13:**3734–3743.

6a. **Clayton, C.** Unpublished data.

7. **Clayton, C. E.** 1992. Control of nuclear gene expression in *Trypanosoma brucei*. Prog. Nucleic Acid Res. Mol. Biol. **43:**37–66.
8. **Clayton, C. E., and M. R. Mowatt.** 1989. The procyclic acidic repetitive proteins of *Trypanosoma brucei*: purification and post-translational modification. J. Biol. Chem. **264:**15088–15093.
9. **DeLange, T., J. M. Kooter, P. A. M. Michels, and P. Borst.** 1983. Telomere conversion in trypanosomes. Nucleic Acids Res. **11:** 8149–8165.
10. **Goguel, V., and M. Rosbash.** 1993. Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. Cell **72:**893–903.
11. **Gunderson, S. I., K. Beyer, G. Martin, W. Keller, W. C. Boelens, and I. W. Mattaj.** 1994. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. Cell **76:**531–542.
12. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:**351–359.
13. **Huang, J., and L. H. T. van der Ploeg.** 1991. Maturation of polycistronic pre-mRNA in *Trypanosoma brucei*: analysis of *trans* splicing and poly(A) addition at nascent RNA transcripts from the *hsp70* locus. Mol. Cell. Biol. **11:**3180–3190.
14. **Huang, J., and L. H. T. van der Ploeg.** 1991. Requirement of a poly-pyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3′ splice acceptor site. EMBO J. **10:**3877–3885.

15. **Hug, M., V. Carruthers, D. Sherman, C. Hartmann, G. A. M. Cross, and C. E. Clayton.** 1993. A possible role for the 3′-untranslated region in developmental regulation in *Trypanosoma brucei*. Mol. Biochem. Parasitol. **61**:87–96.
16. **Jackson, R. J.** 1993. Cytoplasmic regulation of mRNA function: the importance of the 3′-untranslated region. Cell **74**:9–14.
17. **Kapotas, N., and V. Bellofatto.** 1993. Differential response to RNA *trans*-splicing signals within the phosphoglycerate kinase gene cluster in *Trypanosoma brucei*. Nucleic Acids Res. **21**:4067–4072.
18. **LeBowitz, J. H., H. Q. Smith, L. Rusche, and S. M. Beverley.** 1993. Coupling of poly(A) site selection and *trans* splicing in *Leishmania*. Genes Dev. **7**:996–1007.
19. **Lou, H., A. J. McCullough, and M. A. Schuler.** 1993. 3′ splice site selection in dicot plant nuclei is position dependent. Mol. Cell. Biol. **13**:4485–4493.
20. **Lutz, C. S., and J. C. Alwine.** 1994. Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. Genes Dev. **8**:576–586.
21. **Maniatis, T.** 1991. Mechanisms of alternative pre-mRNA splicing. Science **251**:33–34.
22. **Matthews, K. R., C. Tschudi, and E. Ullu.** 1994. A common pyrimidine-rich motif governs *trans*-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes Dev. **8**:491–501.
23. **McKeown, M.** 1992. Alternative mRNA splicing. Annu. Rev. Cell Biol. **8**:133–155.
24. **Michels, P. A. M., A. Y. C. Liu, A. Bernards, P. Sloof, M. M. W. Van der Bijl, A. H. Schinkel, H. H. Menke, P. Borst, G. H. Veeneman, M. C. Tromp, and J. H. van Boom.** 1983. Activation of the genes for variant surface glycoproteins 117 and 118 in *Trypanosoma brucei*. J. Mol. Biol. **166**:537–556.
25. **Mowatt, M. R., and C. E. Clayton.** 1988. Polymorphism in the procyclic acid repetitive protein (PARP) gene family of *Trypanosoma brucei*. Mol. Cell. Biol. **8**:4055–4062.
26. **Mowatt, M. R., G. S. Wisdom, and C. E. Clayton.** 1989. Variation of tandem repeats in the developmentally regulated procyclic acidic repetitive proteins of *Trypanosoma brucei*. Mol. Cell. Biol. **9**: 1332–1335.
27. **Niwa, M., and S. M. Berget.** 1991. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. Genes Dev. **5**:2086–2095.
28. **Niwa, M., C. C. MacDonald, and S. M. Berget.** 1992. Are vertebrate exons scanned during splice site selection? Nature (London) **360**:277–280.
29. **Pays, E., H. Coquelet, P. Tebabi, A. Pays, D. Jefferies, M. Steinert, E. Koenig, R. O. Williams, and I. Roditi.** 1990. *Trypanosoma brucei*: constitutive activity of the VSG and procyclin gene promoters. EMBO J. **9**:3145–3151.
30. **Scholler, J. K., R. F. Aline, Jr., and K. D. Stuart.** 1988. Variant specific transcripts from the co-transposed segments of variant surface glycoprotein genes in *Trypanosoma brucei*. Mol. Biochem. Parasitol. **29**:89–104.
31. **Schürch, N., A. Hehl, E. Vassella, R. Braun, and I. Roditi.** 1994. Accurate polyadenylation of procyclin mRNAs in *Trypanosoma brucei* is determined by pyrimidine-rich elements in the intergenic regions. Mol. Cell. Biol. **14**:3668–3675.
32. **Sherman, D., L. Janz, M. Hug, and C. E. Clayton.** 1991. Anatomy of the procyclic acidic repetitive protein promoter of *Trypanosoma brucei*. EMBO J. **10**:3379–3386.
33. **Sogin, M. L., J. H. Gunderson, H. J. Elwood, R. A. Alonso, and D. A. Peattie.** 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. Science **243**:75–77.
34. **Spieth, J., G. Brooke, S. Kuersten, K. Lea, and T. Blumenthal.** 1993. Operons in C. elegans: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. Cell **73**:521–532.
35. **Ullu, E., K. R. Matthews, and C. Tschudi.** 1993. Temporal order of RNA-processing reactions in trypanosomes: rapid *trans* splicing precedes polyadenylation of newly synthesized tubulin transcripts. Mol. Cell. Biol. **13**:720–725.
36. **Vassella, E., R. Braun, and I. Roditi.** 1994. Control of polyadenylation and alternative splicing of transcripts from adjacent genes in a procyclin expression site: a dual role for polypyrimidine tracts in trypanosomes? Nucleic Acids Res. **22**:1359–1364.
37. **Wassarman, K. M., and J. A. Steitz.** 1993. Association with terminal exons in pre-mRNAs: a new role for the U1 snRNP? Genes Dev. **7**:647–659.
38. **Zomerdijk, J. C. B. M., M. Ouellette, A. L. M. A. ten Asbroek, R. Kieft, A. M. M. Bommer, C. E. Clayton, and P. Borst.** 1990. Active and inactive versions of a promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. EMBO J. **9**: 2791–2801.