

Accurate Polyadenylation of Procyclin mRNAs in *Trypanosoma brucei* Is Determined by Pyrimidine-Rich Elements in the Intergenic Regions

NADIA SCHÜRCH, ADRIAN HEHL, ERIK VASSELLA, RICHARD BRAUN, AND ISABEL RODITI*

Institut für Allgemeine Mikrobiologie, Universität Bern, Bern, Switzerland

Received 1 November 1993/Returned for modification 6 December 1993/Accepted 22 February 1994

Polycistronic precursor RNAs from trypanosomes are processed into monocistronic mRNAs by the excision of intergenic sequences and the addition of a 39-nucleotide spliced leader by *trans* splicing. These mRNAs are also polyadenylated, yet they do not contain the hexamer AAUAAA within their 3' untranslated regions (UTRs). To identify the signals required for the accurate polyadenylation of mRNAs, we tested the effects of deletions in either the procyclin 3' UTR or the downstream intergenic region on the polyadenylation of transcripts from a reporter gene. Deletion of the entire 3' UTR does not affect polyadenylation, but a crucial element is located in the intergenic region and includes a pyrimidine-rich sequence from positions 79 to 112 followed by an AG dinucleotide. Related motifs are also found a similar distance downstream of other genes in both the procyclin and the variant surface glycoprotein expression sites. These sequences bear a strong resemblance to splice acceptor sites, but they are generally several hundred base pairs upstream of the major splice acceptor site of the next gene in the transcription unit. There is evidence, however, that some of them can give rise to alternatively spliced transcripts with unusually long 5' UTRs.

The genome of *Trypanosoma brucei* is organized into clusters of genes which are coordinately transcribed as polycistronic precursor RNAs (reviewed in reference 3). These RNA precursors are subsequently processed into monocistronic mRNAs, each with a 5' minixon or spliced leader sequence which is acquired by *trans* splicing (21, 33) and a poly(A) tail. The sequences determining the 3' splice acceptor site for *trans* splicing in trypanosomes, a polypyrimidine tract followed by the dinucleotide AG, are essentially the same as those used by other organisms for *cis* splicing (10). In contrast, mRNAs from trypanosomes do not contain the canonical polyadenylation signal AAUAAA, which is found 10 to 30 bases upstream of the poly(A) addition site in most mRNAs from higher eukaryotes (36), and no alternative signals have been defined.

The first indication that *trans* splicing and polyadenylation might be linked in *T. brucei* stemmed from the observation that 3' processing of α -tubulin mRNAs was blocked by inhibitors of *trans* splicing (34). These results were summarized in a model in which addition of the minixon to the 5' end of a nascent transcript preceded cleavage at the 3' end. More recently, an analysis of the dihydrofolate reductase-thymidylate synthase locus in a related parasite, *Leishmania major*, has shown that the polyadenylation of transcripts is affected by the 3' splice acceptor site of the next gene in the transcription unit (15). This is still consistent with a coupling of the two processes, but in this case 3' end formation of one mRNA would be linked to *trans* splicing of the downstream transcript.

Some of the most intensively studied transcription units in *T. brucei* are those encoding the major surface glycoproteins—the procyclins (26–29) or procyclic acidic repetitive proteins (18–20)—which are expressed in procyclic forms of the parasite and the variant surface glycoproteins (VSGs) which are expressed in bloodstream forms (reviewed in reference 5). In contrast to

normal housekeeping genes, the procyclin and VSG expression sites are transcribed by an RNA polymerase that is resistant to high levels of α -amanitin (12, 14, 30). There are four procyclin expression sites in procyclic-form trypanosomes (30), each containing two or three tandemly linked procyclin genes (12, 18) followed by a procyclin-associated gene (PAG) (13). While all of the procyclin expression sites appear to be transcribed in procyclic forms of the parasite (12, 13, 19), only one VSG expression site at a time is active in bloodstream-form trypanosomes (5). The active VSG transcription unit is extremely large and encompasses at least eight expression site-associated genes (ESAGs) between the promoter and the VSG gene (23, 25).

We have recently begun to define regulatory elements associated with procyclin genes and have shown that correct polyadenylation of transcripts can occur in the absence of the splice acceptor site from the downstream gene (9). In this study we have now localized a pyrimidine-rich element in the intergenic region between tandemly linked procyclin genes and have shown that it is required for accurate polyadenylation of mRNAs. Related motifs are also found at similar distances downstream of other genes in the procyclin and VSG expression sites. All of these elements contain sequences which resemble splice acceptor sites, and we have demonstrated that at least two of them can give rise to alternatively spliced forms of mRNA with unusually long 5' untranslated regions (UTRs) (35).

MATERIALS AND METHODS

Trypanosome strains. Procyclic forms of *T. brucei* (strain 427) (6) were cultured in SDM-79 (2).

Plasmid constructs. A series of constructs with nested deletions of the procyclin intergenic region (PIGs) were generated by exonuclease III digestion of the wild-type chloramphenicol acetyltransferase (CAT) construct described by Hehl et al. (9). This plasmid is derived from a 4.7-kb *Pvu*II fragment from the Pro A locus (pAP2 [12, 22]) and contains the procyclin promoter, the entire 3' UTR, and 261 bp of the intergenic region. The plasmid was digested with *Xba*I to

* Corresponding author. Mailing address: Institut für Allgemeine Mikrobiologie, Baltzerstrasse 4, CH-3012 Bern, Switzerland. Phone: 31-631 46 47. Fax: 31-631 46 84. Electronic mail address: roditi@imb.unibe.ch.

produce an exonuclease III-sensitive end and either *SacI* (PIG 23 and PIG 73) or *BstXI* (PIG 126 and PIG 187) to produce an exonuclease III-resistant end. Exonuclease III and S1 nuclease digestion, religation, and cloning were carried out according to manufacturer's (Pharmacia, Lucerne, Switzerland) instructions. Individual clones were sequenced to determine the extent of deletion and then designated by the initials PIG followed by the remaining number of base pairs from the intergenic region.

The plasmid Δ PIG, which contains a deletion between positions 76 and 140 of the intergenic region, was constructed as follows: the upstream fragment between the CAT gene and position 76 was amplified from the wild-type construct by PCR with a CAT-specific oligonucleotide (9) and an oligonucleotide corresponding to the intergenic region (5'-ATCCAACAATTAAAAGGC-3') and then digested with *Bam*HI; the downstream fragment between position 140 and the vector was amplified with a second oligonucleotide corresponding to the intergenic region (5'-ATCAGGCCAGAAATG-3') and the Bluescript forward primer (9) and then digested with *Xba*I. The underlined sequences constitute the two halves of an *EcoRV* site; the nucleotide marked in boldface type is mutated from the T in the original sequence. The two fragments were ligated simultaneously into Bluescript which had been digested with *Bam*HI and *Xba*I and subsequently used to replace the wild-type 3' UTR and intergenic region. The correct assembly of Δ PIG was confirmed by digestion with several restriction enzymes. The construct Δ PYR contains a deletion of the polypyrimidine tract from positions 79 to 112 of the intergenic region. This construct was derived by insertion of the oligonucleotide 5'-GAG↓GGGTTCGAATAATAGTTCCTTCTAAACC-3' (and its complement) into the unique *EcoRV* site of Δ PIG. The arrow marks the position of the deletion.

The plasmids Δ 27 and Δ 28 contain internal deletions extending from within the 3' UTR to beyond the wild-type polyadenylation site. These were generated from Δ 16, a plasmid in which a conserved 16-mer in the wild-type 3' UTR has been replaced by a unique *EcoRV* site (9). The plasmid was linearized with *EcoRV*, digested briefly with *Bal* 31 and then recircularized with T4 ligase. The precise borders of the deletions were determined by nucleotide sequencing. Δ UTR contains a deletion extending from a synthetic *Bam*HI site immediately downstream of the CAT coding region (9) to a *Sau*3A site 7 bp upstream of the procyclin polyadenylation site; the intergenic sequences are the same as in the wild-type construct.

Electroporation and CAT assays. Transient transfection of procyclic-form trypanosomes and CAT assays were performed as previously described (9, 38).

RNA isolation and Northern blot analysis. A total of 10^8 procyclic-form trypanosomes were transfected with 20 μ g CsCl-purified plasmid DNA. Cells were harvested 7 h after transfection, and total RNA was isolated essentially as previously described (29) except that the precipitation with LiCl was preceded by digestion with 10 U of RNase-free DNase (Boehringer-Mannheim) to remove any vestiges of plasmid DNA. The procedures used for Northern (RNA) blot analysis were optimized for the detection of rare transcripts. Total RNA (15 μ g) was denatured with glyoxal and dimethyl sulfoxide and separated on a 1.2% agarose gel (31). The RNA was transferred to a nylon membrane (Zeta Probe; Bio-Rad) and fixed by baking at 80°C for 1 h. Prehybridization was carried out for 2 h at 42°C in 6 \times SSC (1 \times SSC is 0.15 M NaCl plus 0.015 M sodium citrate)-0.5% sodium dodecyl sulfate (SDS)-400 μ g of herring sperm DNA ml⁻¹-5 \times Denhardt's solution-50% formamide. For hybridization, the solution was replaced with one in which the concentration of herring sperm DNA was reduced to

100 μ g ml⁻¹. Filters were hybridized at 42°C for 4 h with 5 \times 10⁶ cpm of DNA ml⁻¹ which had been labelled by random priming. Final posthybridization washes were performed in 1 \times SSC-0.5% SDS at 68°C.

Analysis of polyadenylation sites. The amplification of sequences corresponding to the 3' ends of CAT mRNAs was performed as previously described (9). Poly(A) addition sites were determined by nucleotide sequence analysis.

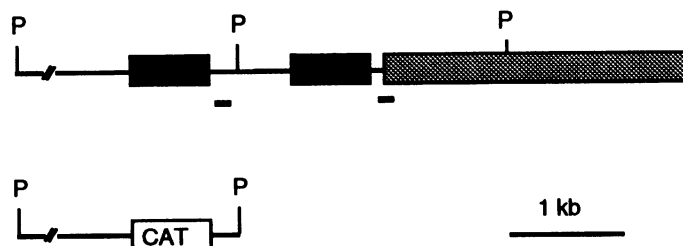
RESULTS

Conserved sequences downstream of the procyclin genes are required for accurate polyadenylation. The starting point of this analysis was the observation that conserved sequences are found at a similar distance downstream of all procyclin genes. In the Pro A locus, which contains tandemly linked procyclin genes followed by the gene PAG 1 (13), one copy of this sequence is found in the intergenic region between procyclin genes, 66 to 168 bp downstream of the polyadenylation site of the procyclin α gene. A related sequence, with 73% identity over 103 bp, starts 74 bp downstream of the procyclin β gene and extends from the intergenic region into PAG 1 (Fig. 1). This sequence can be regarded as two domains which are separated by the splice acceptor site for PAG 1: upstream, a pyrimidine-rich region with 66% identity over 64 bp, and downstream, a more conserved region of 85% over 39 bp.

We have recently analyzed the expression of CAT from a plasmid containing the entire procyclin α gene 3' UTR followed by the first 261 bp of the 603-bp intergenic region (9). This construct does not include the splice acceptor site for the downstream procyclin gene (Fig. 1), but the CAT transcripts derived from it are polyadenylated at the same position as the endogenous procyclin mRNA. In order to analyze the effects of the two conserved domains on expression and polyadenylation, a nested set of deletion mutants was generated from the wild-type construct by treatment with exonuclease III. Each clone was designated by the initials PIG followed by the number of base pairs that remained downstream of the polyadenylation site. Procyclic forms of *T. brucei* were transiently transfected with the different constructs and then assayed for CAT activity. Figure 2 shows a schematic representation of the constructs and levels of CAT activity relative to the wild-type level. Deletion of 74 bp from the wild-type construct, to yield PIG 187, did not reduce the level of CAT activity compared with that obtained with the wild-type construct. Somewhat surprisingly, however, despite the deletion of the more conserved downstream domain, transfection with PIG 126 still resulted in 79% of the wild-type CAT activity. In contrast, a deletion extending into the upstream conserved domain (PIG 73) led to a fivefold decrease in CAT activity, and no further reduction was observed when an additional 50 bp of the intergenic region was removed.

To examine whether the differences in CAT activity obtained with the different constructs reflected changes in polyadenylation, sequences corresponding to the 3' ends of CAT mRNAs were specifically amplified by RACE (rapid amplification of cDNA ends)-PCR and cloned (see Materials and Methods). Individual cDNA clones were then sequenced to determine the precise site of poly(A) addition (Fig. 2 and 3). All six clones derived from PIG 187 were polyadenylated at the same position as procyclin cDNA clones (8, 18). Of nine clones derived from PIG 126, six were polyadenylated at the correct site, but the remaining three were polyadenylated further upstream within the 3' UTR. In contrast, all of the clones derived from either PIG 73 or PIG 23 were polyadenylated within the 3' UTR, and seven of these were at the same

A



B

	10	20	30	40	50	60
PRO-PRO	1	AAAATTATCATTGGTGCCTGTGT--TATTGTGCGTGCCTGCGTGTGAATGTGGCGCTCTG--CCT				
PRO-PAG 1	10	ATCCACTGCATTTGCGGTCTTCTCCTAT--GTCTCGTTTCTGCTGC---AATGGTTGACAGCAACCTC				
	70	80	90	100	110	120
PRO-PRO	61	TTTAATTTGTTGGATGAGCTATTTTCATT-----AATTTTTTTGCCCTTCT-CTTCTTTTGGGTTCGA				
PRO-PAG 1	70	CTCCATTTGTGAC--TGAACTTTTTTTTTTTCGAACTTTTT--GCCTTGCTCTTTTTTCGTACTGGG				
	130	140	150	160	170	180
PRO-PRO	121	ATAATAGTTCCTTCTAAACCTTCAGGCCAGAAATGGGAAACAAGTGTAGACCGGCCAAC'TTGGGCCGA				
PRO-PAG 1	132	AAAATAGTTCCTTCTAAACCTTCAGGCTAGAAATGGGAGACAAATGACTTTTGTGACTCAAGCGAT				

FIG. 1. (a) Organization of the Pro A locus in *T. brucei* 427. The Pro A locus is equivalent to the procyclic acidic repetitive protein B2 locus characterized by Mowatt and Clayton (18). Black boxes represent tandemly linked procyclin genes. The stippled box represents PAG 1, which is the next gene in the transcription unit (13). P, *PvuII* sites. Conserved regions downstream of the procyclin genes are marked by black bars. The wild-type CAT vector (9) was constructed by replacing the procyclin α gene coding sequence with that of the CAT gene. (b) Conserved sequences are found at a similar distance downstream of the procyclin genes. The sequence marked PRO-PRO is from the intergenic region between the procyclin α and β genes (12). PRO-PAG 1 extends from the intergenic region downstream of the procyclin β gene into PAG 1 (13). In each case, numbering begins with the first adenine residue after the end of the transcript (determined from cDNA clones [18, 27]). Identical residues are marked by colons. AG dinucleotides marked in boldface type can be used as splice acceptor sites (13, 35). Conserved pyrimidine-rich regions are underlined. The endpoints of clones generated by exonuclease III deletion (PIG clones) are marked by arrows.

position. As we have previously observed for CAT transcripts derived from a construct in which the 3' UTR and intergenic region had been reversed (9), polyadenylation occurred preferentially at small stretches of adenine residues.

The data presented above provide a strong indication that polyadenylation is determined by an element downstream of the procyclin genes. It might be argued, however, that this effect is independent of a specific sequence in the intergenic region but is rather due to the influence of Bluescript vector sequences which have now been juxtaposed with the polyadenylation site. To exclude this possibility, we initially constructed the plasmid Δ PIG which contained a 65-bp deletion from positions 76 to 140 of the intergenic region (Fig. 2). In this construct, the distance between the wild-type polyadenylation site and vector sequences differs from that of PIG 187 by only 9 bp. In marked contrast, however, all CAT transcripts derived from Δ PIG were incorrectly polyadenylated (Fig. 3), whereas those from PIG 187 were all correctly processed. These results confirm that accurate polyadenylation depends on an element within the intergenic region itself. To further refine this element, the plasmid Δ PYR, with a deletion of the pyrimidine-rich region from positions 79 to 112, was constructed (Fig. 1 and 2). In this case, seven of eight CAT cDNAs

derived from Δ PYR were incorrectly processed (Fig. 2 and 3), establishing that this sequence is a major determinant of polyadenylation.

Sequences in the procyclin 3' UTR are not essential for accurate polyadenylation. Although downstream sequence elements can also influence polyadenylation in higher eukaryotes, the principal polyadenylation signal is usually found within the last 40 bases of the 3' UTR. Since all the PIG constructs contained an intact 3' UTR, a new set of deletion mutants was generated in order to examine the effect of this region. One of these constructs, Δ UTR, is devoid of all but the last 7 bp of the 3' UTR but still contains the wild-type polyadenylation site (Fig. 4). Two additional constructs, Δ 27 and Δ 28, contain 115 and 41 bp of the 3' UTR, respectively, but lack the wild-type polyadenylation site (Fig. 4). The deletion in Δ 28 extends 16 bp beyond the polyadenylation site, while that in Δ 27 ends 16 bp further downstream. Despite these large deletions, each of the three mutants gave rise to higher levels of CAT activity than the wild-type construct when assayed by transient transfection (Fig. 4). Poly(A) addition sites were also determined for individual CAT cDNA clones as

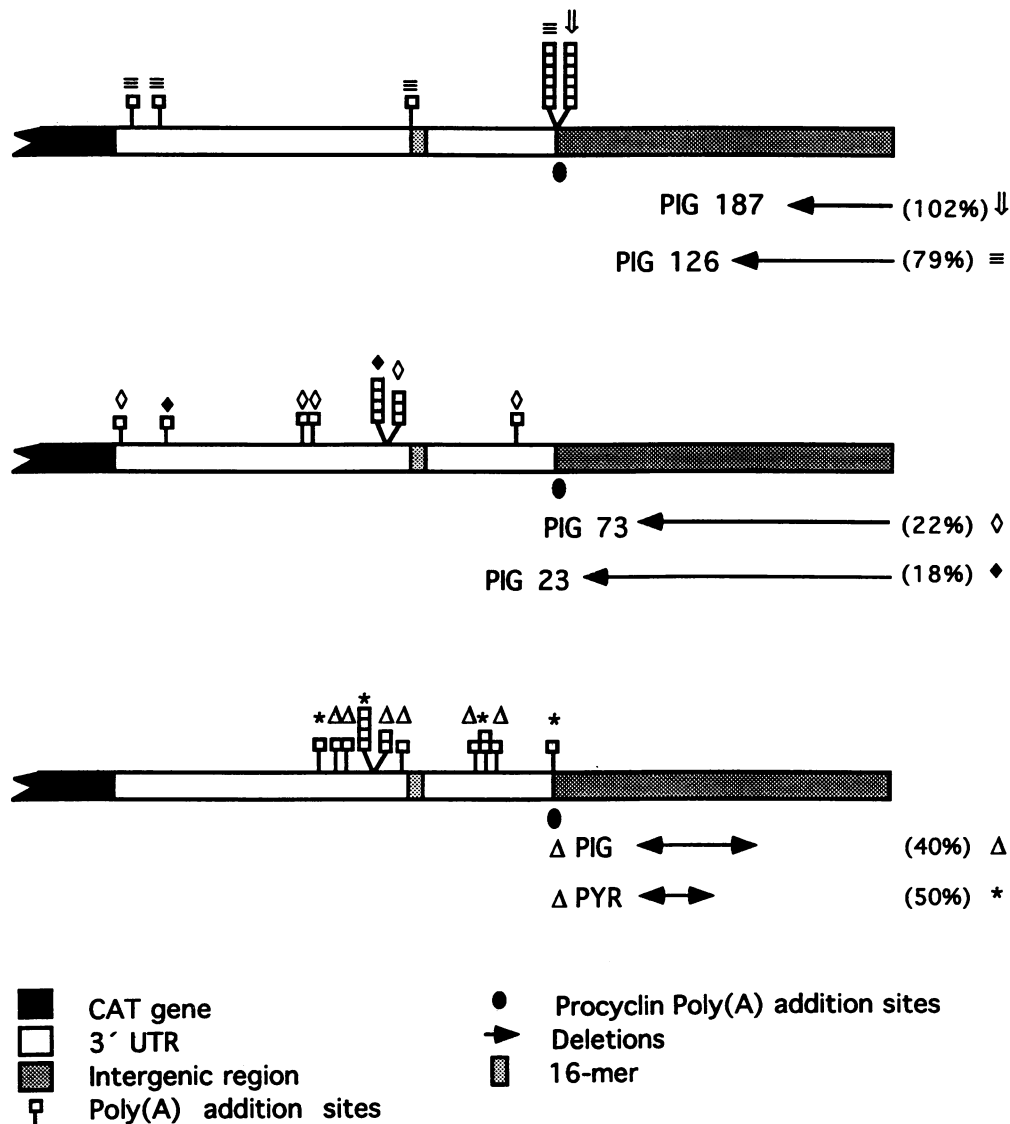


FIG. 2. Schematic representation of poly(A) addition sites in cDNA clones derived from different PIG constructs. Boxes indicate the number of clones that were mapped to a given site. CAT activities relative to that of the wild type (100%) are shown in parentheses. The symbols used for the different constructs are given in the legend at the bottom of Fig. 3.

described above. All eight clones derived from Δ UTR were polyadenylated at the correct position, indicating that no essential signals are provided by upstream sequences in the 3' UTR (Fig. 3 and 4). Of a total of 12 clones originating from Δ 27 and Δ 28, 11 were polyadenylated within the remaining portion of the 3' UTR (Fig. 4), and only 1 clone extended into the intergenic region. The poly(A) addition sites in all six clones derived from Δ 28 mapped to two positions which were only 4 bases apart. These sites were not used in cDNA clones derived from Δ 27, although they are present; instead, poly(A) addition occurred at a cluster of sites 20 to 40 bases further downstream (Fig. 3 and 4). Thus, in all cases, a similar distance is maintained between the poly(A) addition sites and the downstream polyadenylation signal. These results suggest that in the absence of the wild-type poly(A) addition site, alternative sites are chosen by progressively scanning upstream at a certain distance from the polyadenylation signal.

Effect of deletions on the levels of steady-state mRNA. To

ascertain the effect of different polyadenylation sites on the steady-state levels of CAT mRNA, Northern blot analysis was performed with total RNA which had been isolated from transiently transfected cells. Filters were first hybridized with a CAT-specific DNA probe and subsequently rehybridized with a procyclin probe (Fig. 5). In general, the relative amounts of CAT mRNA reflect the CAT activities obtained with the various constructs, with the lowest levels in trypanosomes transfected with PIG 23 and PIG 73. The one exception to this is Δ UTR, which gives rise to less steady-state mRNA but nevertheless a higher level of CAT activity than the wild-type construct. In all cases, the sizes of the transcripts are fully consistent with the results obtained by mapping of polyadenylation sites.

DISCUSSION

A conserved sequence which is located downstream of procyclin genes contains a 34-bp pyrimidine-rich region which

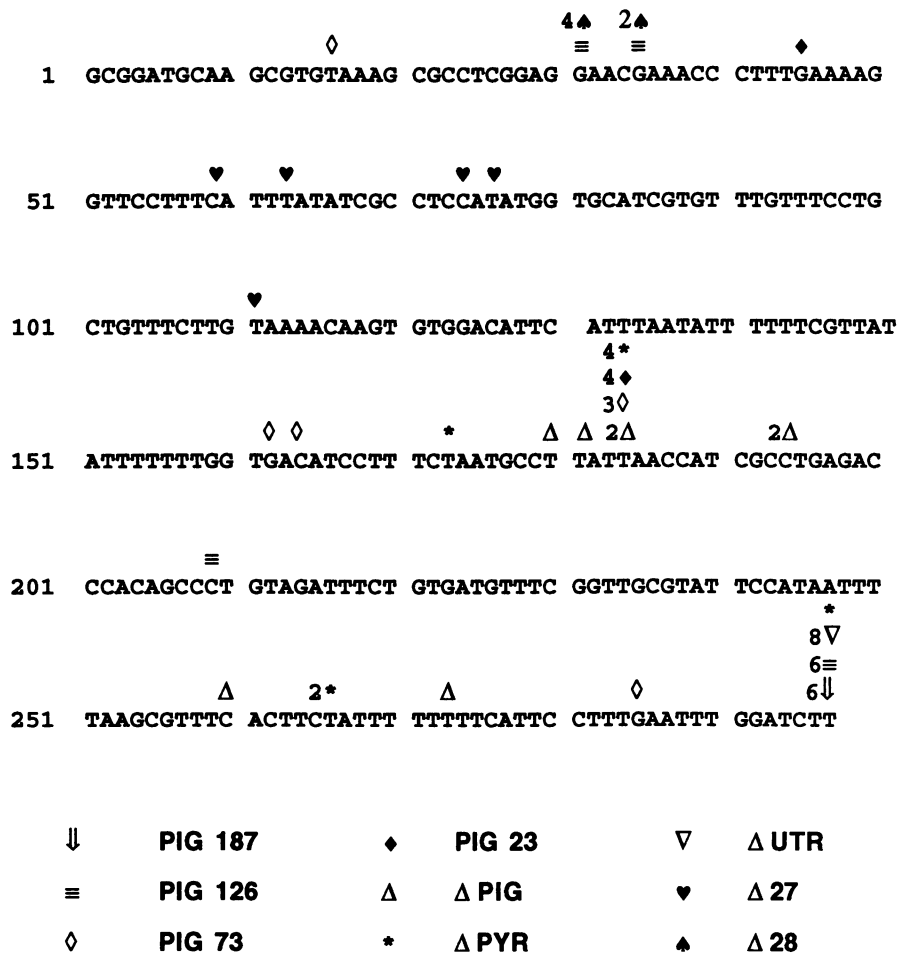


FIG. 3. Sequence of the procyclin 3' UTR (12) and the precise location of polyadenylation sites determined from 61 cDNA clones. The last nucleotide in the sequence corresponds to the wild-type polyadenylation site.

is of major importance in determining the correct polyadenylation site in procyclic-form trypanosomes. With one exception, deletion of this element shifted polyadenylation from the wild-type position to internal sequences within the 3' UTR. The choice of these sites probably reflects the use of alternative polypyrimidine tracts that are located upstream from the wild-type element. As we have observed previously, poly(A) addition occurs preferentially at small stretches of adenine

residues (9, 13), but apart from this there are no obvious sequence motifs in common. Indeed, although certain sites are used more efficiently than others, the requirements cannot be particularly stringent, since 23 alternative poly(A) addition sites were identified from an analysis of 41 independent cDNA clones that were incorrectly polyadenylated (Fig. 3). Our results also indicate that there is no simple correlation between expression and correct polyadenylation, since the endogenous

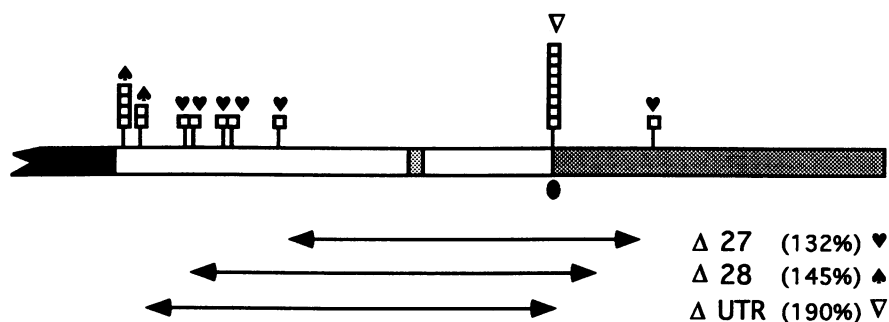


FIG. 4. Schematic representation of poly(A) addition sites in cDNA clones derived from constructs with deletions in the 3' UTR. The symbols used are the same as those in Fig. 2 and 3.

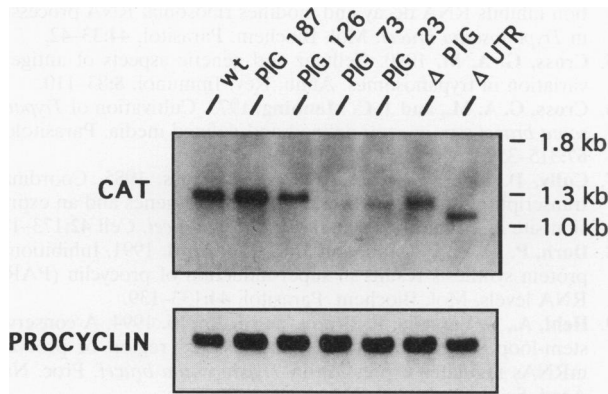


FIG. 5. Analysis of the levels of steady-state RNA. Total RNA was isolated 7 h after transient transfection with individual CAT constructs. The filter was first hybridized with a probe corresponding to the CAT coding region and then normalized by hybridization with a procyclin probe. wt, wild type.

polyadenylation site was deleted from Δ27 and Δ28, yet these constructs consistently resulted in higher levels of CAT activity than the wild type.

From our analysis, it seems unlikely that there are any polyadenylation signals within the procyclin 3' UTR itself. The construct ΔUTR contains only 7 bp of the 3' UTR, yet all eight CAT cDNA clones derived from this construct were polyadenylated at the wild-type position. Although the 3' UTR does not appear to play a role in polyadenylation, it is nevertheless clear that it can modulate expression in procyclic forms in other ways. We have previously demonstrated that deletion of

a conserved 16-base element from the 3' UTR reduces CAT activity more than 10-fold, although the relative amount of steady-state CAT mRNA is unaltered and polyadenylation still occurs at the correct position (9). In contrast, much larger deletions of the 3' UTR, which span the 16-mer, lead to an increase in CAT activity relative to that of the wild-type construct, suggesting that a negative element has now been removed from the 3' UTR. We have also shown that transfection with the constructs PIG 23 and PIG 73 results in approximately seven times less CAT activity than that with Δ27 and Δ28. Since the majority of transcripts derived from PIG 23 and PIG 73 contain a portion of the 3' UTR from positions 111 to 161 that is absent from transcripts derived from Δ27 and Δ28 (Fig. 3), this would also be consistent with a negative element in this region.

Inspection of the nucleotide sequences of other intergenic regions in the procyclin and VSG expression sites for potential processing signals revealed related motifs at similar distances downstream from the polyadenylation sites of seven genes for which data were available (Fig. 6). Alignment of these sequences indicates that there might be four elements that are involved in the choice of a polyadenylation site in these transcription units: (i) a small stretch of three to six adenine residues at the poly(A) addition site itself, (ii) an intervening sequence of approximately 60 to 90 bp, (iii) a pyrimidine-rich tract followed by the trinucleotide YAG, and (iv) a short distance further downstream, a second pyrimidine tract again followed by one or two copies of the trinucleotide YAG. It is worth noting that this last element was deleted from PIG 126 but was fortuitously replaced by similar vector sequences (Fig. 6), which may explain the predominance of the wild-type polyadenylation site in cDNA clones derived from this construct.

	Distance from Poly(A) site to		
	Pyr	AG	
PRO α	<u>TTTTTTTGCCTTCTCTCTTTTGGGTTGGAATAAT</u> AG <u>TTCCTTCTAAACCTTCAGGCCAGAAATGGGAA</u>	79	126
PRO β	<u>CTTTTTGCCTTGTCTTTTTTCGTAAGGAAAAT</u> AG <u>TTCCTTCTAAAACCTTCAGGCTAGAAATGGAG</u>	87	137
PAG 1	<u>GTGGGTTCGAGGTGCCAGACATCTTTTGTTCCTTC</u> AG <u>CAAAATGCCAAGAACAAGTGGCATTTTTAGTAGT</u>	63	78
MVAT 5	<u>TTTTTTTTTCTCTCTTTTCCCTGTTTTTTTTTGTGG</u> AG <u>TTTTTGTAGGTAGGAATGGGGGGTTAGTAGG</u>	80	114
ESAG 1	<u>ACAAATTTTCTTTTTTATTTTCCGCCCTCTTTC</u> AG <u>GACAATTCCTCTTGTAGGCTATAAAAGTGAAA</u>	67	98
ESAG 5	<u>TGAGTGCAGTTTTTCTGGTAACCTATTTCTTTTCAT</u> AG <u>TTTTCTTTCATAACCTCTCTATAATTTTAAAG</u>	71	97
ESAG 7	<u>TAAC TGCTTATCATGGCGTCGAATTCACCTCTCC</u> AG <u>AATCTTTATTTTAGTGCGCTATAGATGATAGT</u>	79	89
PIG 126	<u>TTTTTTTGCCTTCTCTCTTTTGGGTTGGAATAAT</u> Ag <u>gtgagctcaattcgcctatag</u>		

Consensus sequence

(A₃₋₆) ≈60-90bp YYYYYYYY . . YAG . . YYY . . YAG

FIG. 6. Related motifs are found downstream of other genes in the procyclin and VSG expression sites. Sequences are aligned at an AG dinucleotide (shown in boldface type). Pyrimidines are underlined. In some cases, the pyrimidine-rich regions begin upstream of the sequences shown in the figure. PIG 126 shows the juxtaposition of sequences from the procyclin intergenic region (uppercase) with Bluescript vector sequences (lowercase). Sequence data were obtained from the following sources: procyclin α gene (PRO α), reference 12; procyclin β gene (PRO β), reference 13; metacyclic VSG gene (MVAT 5), reference 16; PAG 1, reference 17; ESAG 1, reference 7; and ESAG 5 and ESAG 7 (23). The same sequences are found downstream of ESAG 6 and ESAG 7 (23).

Pyrimidine-rich regions followed by the dinucleotide AG function as splice acceptor sites in trypanosomes (10), and it has been shown for the DHFR-TS locus of *L. major* that one element regulates both polyadenylation and *trans* splicing (15). This may also be true for a number of transcription units in *T. brucei*, particularly in cases in which the intergenic regions are small and the first polypyrimidine tract beyond a given gene is the splice acceptor site for the next gene (32, 37). In the procyclin and VSG transcription units, however, the (putative) signals for polyadenylation do not correspond to the major splice acceptor sites of downstream genes. This is particularly striking in the case of ESAG 1, in which the potential polyadenylation signal is separated from the downstream VSG splice acceptor site by several kilobases of barren region sequence (7). There is also a pyrimidine-rich region downstream of the VSG gene (16), even though it is the last gene in the transcription unit and there is theoretically no need for further *trans* splicing to occur. It is worth noting, however, that at least three of the elements shown in Fig. 6 are also used for *trans* splicing in vivo. The pyrimidine tract downstream of the procyclin α gene determines the splice acceptor site for a minor population of mRNAs derived from the procyclin β gene (35), and the element downstream of the procyclin β gene is used to generate the 2.7-kb transcript from PAG 1 (13). The pyrimidine-rich region 3' of ESAG 7 is also used to produce an alternatively spliced form of ESAG 6 mRNA (4). In each case, however, the use of this site results in an mRNA in which the major open reading frame is preceded by other small reading frames. Although these mRNAs are found within the polysomal fraction (35), it is not known whether they are actually translated.

The regulation of polyadenylation in trypanosomes may provide yet another control point for stage-specific gene expression. Transcripts from several mitochondrial genes are polyadenylated to different extents in bloodstream and procyclic forms (1). It has also recently been shown that polyadenylated forms of minixon-derived RNA accumulate in short, stumpy bloodstream forms but not in long, slender bloodstream forms or in procyclic forms of the parasite (24). In contrast to all the nuclear genes we have examined, the minixon repeat does not contain a recognizable polyadenylation signal. It is possible, however, that polyadenylation of minixon-derived RNAs differs from that of normal mRNAs and is dictated by other elements.

The finding that polyadenylation in trypanosomes is dictated by downstream sequences casts the intergenic regions in a new light. They should no longer be regarded as inert and therefore, dispensable spacer sequences between adjacent genes but rather should be explored for further elements which might regulate gene expression.

ACKNOWLEDGMENTS

We thank Christine Clayton and Michael Hug for exchanging data before publication.

This work was supported by the Swiss National Science Foundation and the Roche Foundation.

REFERENCES

- Bhat, G. J., A. E. Souza, J. E. Feagin, and K. Stuart. 1992. Transcript-specific developmental regulation of polyadenylation in *Trypanosoma brucei* mitochondria. *Mol. Biochem. Parasitol.* **52**: 231–240.
- Brun, R., and M. Schoenenberger. 1979. Cultivation and *in vitro* cloning of procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. *Acta Trop.* **36**:289–292.
- Clayton, C. 1992. Developmental regulation of nuclear gene expression in *Trypanosoma brucei*. *Progr. Nucleic Acids Res.* **43**:37–66.
- Coquelet, H., M. Steinert, and E. Pays. 1991. Ultraviolet irradiation inhibits RNA decay and modifies ribosomal RNA processing in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **44**:33–42.
- Cross, G. A. M. 1990. Cellular and genetic aspects of antigenic variation in trypanosomes. *Annu. Rev. Immunol.* **8**:83–110.
- Cross, G. A. M., and J. C. Manning. 1973. Cultivation of *Trypanosoma brucei* ssp. in semi-defined and defined media. *Parasitology* **67**:315–331.
- Cully, D. F., H. S. Ip, and G. A. M. Cross. 1985. Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene in *Trypanosoma brucei*. *Cell* **42**:173–182.
- Dorn, P. L., R. A. Aman, and J. C. Boothroyd. 1991. Inhibition of protein synthesis results in super-induction of procyclin (PARP) RNA levels. *Mol. Biochem. Parasitol.* **44**:133–139.
- Hehl, A., E. Vassella, R. Braun, and I. Roditi. 1994. A conserved stem-loop structure in the 3' untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. USA* **91**:370–374.
- Huang, J., and L. H. T. van der Ploeg. 1991. Requirement of a polypyrimidine tract for *trans*-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3' splice acceptor site. *EMBO J.* **10**:3877–3885.
- Hug, M., and C. Clayton. Hierarchies of RNA processing signals in a trypanosome surface antigen precursor. Submitted for publication.
- Koenig, E., H. Delius, M. Carrington, R. O. Williams, and I. Roditi. 1989. Duplication and transcription of procyclin genes in *Trypanosoma brucei*. *Nucleic Acids Res.* **17**:8727–8739.
- Koenig-Martin, E., M. Yamage, and I. Roditi. 1992. A procyclin-associated gene in *Trypanosoma brucei* encodes a polypeptide related to ESAG 6 and 7. *Mol. Biochem. Parasitol.* **55**:135–146.
- Kooter, J. M., and P. Borst. 1984. α -Amanitin-insensitive transcription of VSG genes provides further evidence for discontinuous transcription in trypanosomes. *Nucleic Acids Res.* **12**:9457–9472.
- 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.
- Lu, Y., T. Hall, L. S. Gay, and J. Donelson. 1993. Point mutations are associated with a gene duplication leading to the bloodstream reexpression of a trypanosome metacyclic VSG. *Cell* **72**:397–406.
- Martin, E. 1992. Ph.D. thesis. University of Tübingen, Tübingen, Germany.
- Mowatt, M. R., and C. E. Clayton. 1987. Developmental regulation of a novel repetitive protein of *Trypanosoma brucei*. *Mol. Cell. Biol.* **7**:2833–2844.
- Mowatt, M. R., and C. E. Clayton. 1988. Polymorphisms in the procyclic acidic repetitive protein (PARP) gene family of *Trypanosoma brucei*. *Mol. Cell. Biol.* **8**:4055–4062.
- 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.
- Murphy, W. J., K. P. Watkins, and N. Agabian. 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. *Cell* **47**:517–525.
- 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 promoters. *EMBO J.* **9**:3145–3151.
- Pays, E., P. Tebabi, A. Pays, H. Coquelet, P. Revelard, D. Salmon, and M. Steinert. 1989. The genes and transcripts of an antigen gene expression site from *T. brucei*. *Cell* **57**:835–845.
- Pellé, R., and N. B. Murphy. 1993. Stage-specific differential polyadenylation of mini-exon derived RNA in African trypanosomes. *Mol. Biochem. Parasitol.* **59**:277–286.
- Revelard, P., S. Lips, and E. Pays. 1990. A gene from the VSG expression site of *Trypanosoma brucei* encodes a protein with both leucine-rich repeats and a putative zinc finger. *Nucleic Acids Res.* **18**:7299–7304.
- Richardson, J. P., R. P. Beecroft, D. L. Tolson, M. K. Liu, and T. W. Pearson. 1988. Procyclin: an unusual immunodominant glycoprotein surface antigen from the procyclic stage of African trypanosomes. *Mol. Biochem. Parasitol.* **31**:203–216.

27. **Roditi, I., M. Carrington, and M. Turner.** 1987. Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*. *Nature (London)* **325**:272-274.
28. **Roditi, I., and T. W. Pearson.** 1990. The procyclin coat of African trypanosomes. *Parasitol. Today* **6**:79-82.
29. **Roditi, I., H. Schwarz, T. W. Pearson, R. P. Beecroft, M. K. Liu, J. P. Richardson, H.-J. Bühring, J. Pleiss, R. Bülow, R. O. Williams, and P. Overath.** 1989. Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of *Trypanosoma brucei*. *J. Cell Biol.* **108**:737-746.
30. **Rudenko, G., D. Bishop, K. Gottesdiener, and L. H. T. van der Ploeg.** 1989. Alpha-amanitin resistant transcription of protein coding genes in insect and bloodstream form *Trypanosoma brucei*. *EMBO J.* **8**:4259-4263.
31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning: a laboratory manual* 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Schlaeppli, K., J. Deflorin, and T. Seebeck.** 1989. The major component of the paraflagellar rod of *Trypanosoma brucei* is a highly helical protein which is encoded by two identical, tandemly clustered genes. *J. Cell Biol.* **109**:1695-1709.
33. **Sutton, R. E., and J. C. Boothroyd.** 1986. Evidence for trans-splicing in trypanosomes. *Cell* **47**:527-535.
34. **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.
35. **Vassella, E., R. Braun, and I. Roditi.** 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.*, in press.
36. **Wickens, M.** 1990. How the messenger RNA got its tail: addition of poly(A) in the nucleus. *Trends Biochem. Sci.* **15**:277-281.
37. **Wong, S., T. H. Morales, J. E. Neigel, and D. A. Campbell.** 1993. Genomic and transcriptional linkage of the genes for calmodulin, EF-Hand 5 protein and ubiquitin extension protein 52 in *Trypanosoma brucei*. *Mol. Cell. Biol.* **13**:207-216.
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. The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *EMBO J.* **9**:2791-2801.