Nucleic Acids Research, 1993, Vol. 21, No. 17 **4067–4072**

Differential response to RNA *trans*-splicing signals within the phosphoglycerate kinase gene cluster in *Trypanosoma brucei*

Nikitas Kapotas* and Vivian Bellofatto
Laboratory of Molecular Parasitology, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

Received May 10, 1993; Revised and Accepted July 12, 1993

ABSTRACT

In trypanosomatids, nuclear pre-mRNA splicing is exclusively a *trans*-splicing reaction in which a capped, 39 nt exon, the mini-exon, is positioned 5′ to an open reading frame. Differential RNA splicing might reflect specific mini-exon and 3′ splice site interactions. To test this hypothesis, we compared the efficiency of mini-exon addition to three natural 3′ splice acceptor sites (SASs) located within a single pre-mRNA transcript. In *Trypanosoma brucei*, the phosphoglycerate kinase A, B and C genes (PGK A, B and C) are co-expressed as three consecutive sequences on a polycistronic pre-mRNA. This pre-mRNA gives rise to unequal amounts of PGK A, B and C mRNAs. When the SAS from each gene was placed upstream of the luciferase open reading frame and the resultant constructs transiently transfected into *T. brucei* procyclic cells, luciferase activity levels indicated differential SAS utilization. Enzyme activity was low when the SAS from the A gene was present. Levels were indistinguishable when the B and C SASs were compared. After replacing luciferase with chloramphenicol acetyl transferase in the test constructs, enzyme activities were shown to directly correlate with mRNA amounts. Thus, poor splicing efficiency accounts for the differential expression of the PGK A mRNA during PGK pre-mRNA maturation. This reaction appears to reflect the polypyrimidine pattern within the 3′ splice acceptor site.

INTRODUCTION

Trans-splicing is a key RNA processing event that occurs in an eclectic set of eukaryotic organisms including trypanosomatids, trematode and nematode worms, and euglenoids (1–4). Within the organisms of the trypanosomatidae family, mRNAs are carved from pre-mRNA transcripts by an intermolecular *trans*-splicing reaction that docks a 39 nt, capped 5′ exon (the mini-exon), which is initially the 5′ extremity of a short, non-polyadenylated medRNA (mini-exon derived RNA), upstream of a translation initiation site; and by a polyadenylation reaction, that assembles a polyadenosine tail 3′ to a translation stop site. In trypanosomes, multiple protein coding regions within polycistronic pre-mRNAs (5) are punctuated by trans-splicing and polyadenylation sites and it appears that the *trans*-addition of 5′ capped mini-exons permits the formation of multiple mRNAs from a single pre-mRNA.

Although trans-splicing appears to be a co-transcriptional process (6), little is known about how the mini-exon chooses a splice site within pre-mRNA and even less is known about differential splicing efficiencies or if such a phenomenon influences polycistronic pre-mRNA processing. To address this problem, we have compared the efficiency of 3′ splice sites utilization within three related genes that are expressed as part of one polycistronic pre-mRNA. By positioning each 3′ splice site between a common transcriptional promoter and reporter gene flanked by a polyadenylation site, we were able to focus on splice site usage in a genetic environment in which RNA transcription, polyadenylation and turnover remained constant.

What RNA signals direct splicing in trypanosomes? Certainly, sequence elements that are common to all medRNA donor molecules and to all 3′ splice site regions probably contribute to trans-spliceosome recognition of substrate RNAs. For example, medRNAs isolated from numerous members of the trypanosomatid family all contain an identical UUG/GUA sequence at the 39 nt exon/intron boundary (7, 8). In addition, a 3′ splice site within a pre-mRNAs is always an AG dinucleotide, and one or several AGs which precede each open reading frame on a pre-mRNA can serve as the mini-exon acceptor site (9–15). In the cases of *T. brucei* tubulin genes, in which the medRNA-pre-mRNA branch points have been mapped, the 2′-5′ phosphodiester bond-containing nucleotides are a subset of A residues located 42 to 58 nt upstream from the AG splice sites (16). In addition to these simple sequence elements, splicing efficiency is probably governed by additional, proximal sequences that surround the AG dinucleotide; we refer to this region as the 'splice acceptor site' (SAS). The role of an SAS in affecting splice site utilization was initially addressed in the analysis of Varient Surface Glycoprotein (VSG) gene mRNA processing in *Trypanosoma equiperdum* (9). The authors demonstrated that

* To whom correspondence should be addressed at: Box 300, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

mini-exon addition occurred at several AG dinucleotides within a short region 5′ to the translational start site of the VSG mRNA, and concluded that differential mRNA accumulation indicated preferential mini-exon addition to AG sites that followed polypyrimidine tracts. A recent, detailed description of the procyclic acidic repetitive protein (PARP) SAS region in *Trypanosoma brucei* (17) and the substitution of a 222 bp α-tubulin SAS region with a synthetic polypyrimidine tract in *Leishmania enriettii* (18), highlight a *trans*-splicing requirement for pyrimidine-rich sequences upstream of the 3′ splice site.

Based on these data, we postulate that each SAS region within pre-mRNAs is recognized by medRNA with a specific efficiency during RNA processing. If this is true, then we should be able to find natural SAS regions that are differentially recognized during pre-mRNA maturation. Such SAS regions probably reside within the *T. brucei* phosphoglycerate kinase (PGK) locus that is transcribed as a single pre-mRNA which gives rise to unequal amounts of three mRNAs. The PGK locus contains a compact cluster of three similar genes, called A, B and C: the B and C genes code for PGK proteins that localize to the cytoplasm and glycosome, respectively; the A gene codes for a PGK-like protein that resides within the glycosome but has a yet undetermined function (15, 19, 20). The expression of the B and C genes is regulated during the trypanosome life cycle (19, 21). PGK B mRNA is highly expressed in insect forms (procyclics) of the parasite; PGK C mRNA is highly expressed in blood forms where the PGK C protein is concentrated in the well-developed glycosome characteristic of this stage of the life cycle. The levels of PGK A mRNA are extremely low in both forms (15, 19). It is possible that unequal amounts of A, B and C mRNAs arise from the PGK multicistronic primary transcript during differential RNA processing and therefore the SAS regions are utilized with dissimilar efficiencies during the mini-exon addition step of mRNA formation. In this work we investigated the activities of the three SAS regions associated with the PGK A, B and C genes. We determined that splicing was dependent upon a 186−220 nt SAS region in each case and that the A SAS region functions at one-tenth the efficiency compared to that of B or C. In contrast, the B and C SAS regions function equally well in procyclic trypanosomes.

MATERIAL AND METHODS

Parasites

Procyclic *T.brucei*, strain 427, were cultured as described previously (22). Transfections were performed using a BTX ECM 600 electroporator (BTX, San Diego, CA), set at 1.5 kV and 24 Ohms (R2 setting), to deliver a single pulse. Log-phase cultures were washed in ZM (23), resuspended to a concentration of 6×10^7 cells/ml, and 400 μl aliquots were mixed with 5 μg of CsCl-purified plasmid DNA in a 4 mm cuvette and electroporated at room temperature. Cells were resuspended in 10 ml SDM-79 (24), incubated for 18 hrs and harvested for enzyme and RNA assays. Co-transfections were done, using an additional 5 μg of a β-galactosidase-containing plasmid (described below), to standardize transfection efficiencies.

Plasmid constructs

All constructs are derivatives of pLSB1, which was generated by cloning PCR products containing the PARP promoter, the SAS of PGK B, the luciferase coding region and the PARP polyadenylation signal region into the plasmid cloning vector pGEM3Z(+) from Promega Corp. (Madison, WI). A 283 bp fragment containing the PARP promoter was cloned in the *Xba*I site of pGEM3Zf(+) after PCR amplification of bases −271 to +12 (as numbered in Brown et al. (25)) of the PARP gene using *Xba*I-restriction site-flanked primers 5′-GGTGCAGTCTA-GATTTGTTGTCAT-3′ and 5′-CTCTGTTAATCTAGATACT-CACCC-3′ and pJP 44 DNA (26). This region is sufficient for promoter activity, and sequences downstream from +15 (17, 25) fail to influence promoter utilization. The SAS of PGK B was obtained by PCR amplification of a 220 bp fragment, which contains the initial 220 bp upstream from the PGK B protein coding region, beginning 10 bp 5′ to the ATG, from genomic DNA using primers NK1 and NK2 (Table I), which added *Sal*I restriction sites to the ends of the product. This fragment was cloned in the forward orientation and downstream from the PARP promoter by using the *Sal*I site within pGEM3Zf(+). A 2.2 kb fragment containing the luciferase coding region followed by the PARP polyadenylation signal region were PCR-amplified using pHD1 DNA (C. Clayton, personal communication) and primers 5′-GGGCCCTGATCAGCAGTGACATTAGCATTC-3′ and 5′-CCCGGGAAGCTTAGCACGCACAATAACACAAGG-3′ to add *Bcl*I−*Hind*III restriction sites to the fragment's termini. The PARP polyadenylation signal region is defined as a 335 bp sequence that immediately follows the translational stop of the PARP gene and includes the site of polyadenylation and the signals necessary for this function (22, 27). The 2.2 kb fragment was cloned downstream of the SAS of PGK B in unique *Bcl*I-*Hind*III sites within the promoter and SAS-containing pGEM3Zf(+) derivative. In this construct the SAS region is a *Sal*I-*Bcl*I cassette. pLSA1, which contains the SAS of PGK A in place of the SAS B region in pLSB1, was constructed by replacing the SAS B region with a PCR fragment produced by using primers NK3 and NK4 (Table I) and genomic DNA to amplify a 192 bp fragment that immediately precedes the translational start site of the PGK A gene. Similarly, pLSC1 was constructed by replacing the SAS B region with the analogous region from the PGK C gene using primers NK5 and NK6 (Table I) to amplify a 186 SAS region. pLSB2 is identical to pLSB1 but lacks the sequence (94 bp) between the 3′ splice site within the SAS B and the luciferase translational start site. pLSB0

Table I. Oligonucleotide sequences

NK1 (−115 to −95)	5′-GGGCCCGTCGACGTGAACGTTGTTTGAGAAGTA-3′
NK2 (+84 to +105)	5′-CCCGGGGTCGACTGATCAGATAAGTATATATCG-3′
NK3 (−130 to −106)	5′-GGGCCCGTCGACTCATAGAACTGCTTTCCAGCA-3′
NK4 (+42 to +62)	5′-CCCGGGTGATCAGATCCCGCTTGTTTTTCTTAT-3′
NK5 (−110 to −88)	5′-GGGCCCGTCGACGTAATATTTCAATCGCTATGA-3′
NK6 (+49 to +76)	5′-CCCGGGTGATCAGTATCACAATTACTTATTAGG-3′

Numbers indicate nucleotide position either (−) upstream or (+) downstream from the canonical AG for either the PGK A, B or C gene, allele 2.

contains no SAS and was produced from pLSB1 by a deletion of the SAS. pLSB1P- lacks the PARP promoter and was constructed from pLSB1 by deleting this region. The PARP promoter and SAS A, B and C regions, as well as all junction sites, were sequenced to assess sequence fidelity after PCR amplification and cloning procedures.

The CAT-containing vectors pCATA1, pCATB1, pCATC1 and pCATB2 were constructed by replacing the luciferase gene in pLSA1, pLSB1, pLSC1 and pLSB2, respectively, with the CAT translated region that was produced using primers 5'-GGGCCCGGATCCATGGAGAAAAAAATCACTGGA-3' and 5'-CCCGGGCCGCGGTTAAAAAAATTACGCCCCGCC-3' in a PCR to produce a 665 bp fragment that was bounded by *Bam*HI−*Sac*II restriction sites.

The β-galactosidase-expressing construct is a derivative of pJP6 (26) and contains the PARP promoter followed by the *lacZ* gene in place of the *VSG117* gene (Hon Ip, personal communication).

Enzyme assays

Cells were harvested 18 hrs post-transfection for luciferase and β-galactosidase assays. β-galactosidase quantities, which served as an internal control for transfection efficiency, were determined using the protocol recommended by GIBCO BRL in their 'Non-Isotopic Immunoassay Kit' for β-galactosidase. For these assay we used 1.4×10^7 cells and calculated β-galactosidase protein concentration after developing a standard curve. For luciferase enzyme assays, 2×10^6 cells were lysed after collecting cells by centrifugation (16,000×g, 30 sec) and resuspending cell pellets in 60 μl of lysis buffer included in a 'Luciferase Assay Kit' (Promega Corp., Madison,WI). The supplier's recommended protocol was followed, although assays were done within 0.5 min of preparing lysate and contained 20 μl of lysate and 80 μl of substrate-containing buffer. For each sample, luciferase activity units were calculated by extrapolating from a standard curve generated using purified enzyme (specific activity of 10^7 arbitrary units/mg, based on bioluminesense assay, Boehringer Mannheim, Indianapolis, IN). Bioluminesense was quantified on a Turner TD-20e Luminometer.

RNA analysis

For mRNA isolation, cells were transiently transfected with pCATA1, pCATB1 or pCATC1 and, 18 hrs post- transfection, 3×10^8 cells were collected by centrifugation and mRNA was extracted using the 'Fast Track mRNA Isolation Kit' (InVitrogen, San Diego, CA). To analyze transcription from the CAT-containing constructs using RT-PCR, an oligonucleotide (5'-CCCGGGCCGCGGTTAAAAAAATTACGCCCCGCC-3') from the 3' end of the CAT coding region was used as a primer to reverse transcribe CAT-containing mRNAs. The products were amplified by PCR for 30 cycles (each cycle: 94°C for 1 min, 45°C for 1.2 min and 72°C for 1 min) and using the CAT primer and a *T. brucei* mini-exon oligonucleotide (5'-CGCTATTATTA-GAACAGTTTCTGTACTATATTG-3') as the 5' primer. PCR products were visualized on 3% Nue-Sieve-1% Seakem agarose gels (SeaKem, Rockland, ME). Gels were transfered to nylon membranes and probed with ^{32}P-labeled CAT sequences that were labeled using the standard random-hexamer method. To assess that the PCR product amounts reflect RTase-product amounts, we amplified several different concentrations of pCATA1 DNA, ranging from 5 to 200 ng/μl, using the same PCR conditions and primers 5' GGGCCCGGATCCATGGAG-AAAAAAATCACTGGA-3' and 5'-CCCGGGCCGCGGTTAA-AAAAATTACGCCCCGCC-3'.

RESULTS

In procyclic *T. brucei* strain 427, the PGK locus is transcribed as a multigenic pre-mRNA that gives rise to low levels of gene A mRNA, high levels of gene B mRNA and intermediate levels of gene C mRNA. We therefore examined the SAS regions that precede the gene A, B and C protein coding region for their relative activity. We constructed plasmids in which the maturation of a reporter gene mRNA was dependent upon mini-exon addition to either the PGK A, B or C SAS (Figure 1). SAS B and C regions are defined as the sequence between the upstream polyadenylation site and the translational start site of their corresponding gene. For the SAS A sequence, we chose a 190 bp fragment that borders the PGK A gene translational start site and is approximately the same size as are the B (220 bp) and C (186 bp) SAS regions. To ensure quantitative analysis of reporter gene expression, we used the firefly luciferase gene since luciferase enzyme assays are sensitive and linear over several orders of magnitude. To maintain constant promoter and polyadenylation activities in a background of varying splice acceptor regions, we chose the well-characterized PARP promoter (25, 26, 28, 29) as well as sequences including and surrounding the PARP polyadenylation site. We limited the size of the PARP promoter fragment to include either 9 or 12 bases

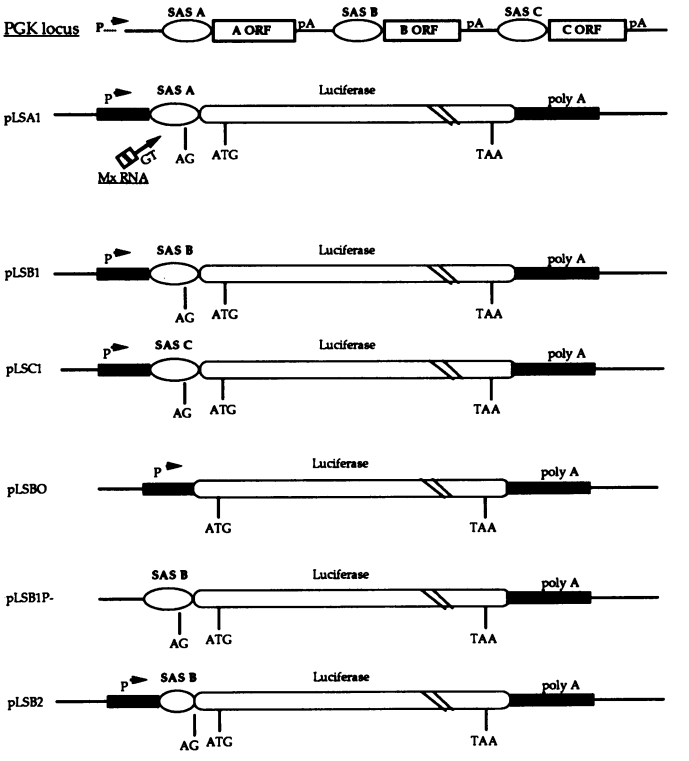

Figure 1. Summary of PGK locus and derivative constructs. The PGK locus contains an undefined promoter (P...), at least three open reading frames (boxed) which encode the PGK A, B and C genes. The SAS regions which precede each open reading frame are drawn as oval. The polyadenylation regions are marked as 'pA'. pLS constructs contain the PARP promoter and polyadenylation region (filled boxes), the luciferase coding region (elongated oval), an SAS region from either the A, B or C PGK gene and bacterial plasmid sequence (indicated as a thin line). The luciferase translational start and stop sites are shown. 3' splice acceptor sites (AG dinucleotides) and the medRNA, in which the mini-exon portion is shown (hatched box) and the intron portion (indicated by its 5' GT sequence) are depicted.

Table II. Effect of different splice sites on luciferase activity.

Plasmid	Luminometer reading(1)	U of luciferase/10^6 cells	U of luciferase/ng β-gal(2)
pLSA1	11.6 ± 1.2	3.5 ± 0.5	19.4 ± 2.8
pLSB1	200.0 ± 70.5	54.3 ± 19.0	301.7 ± 105.5
pLSC1	200.0 ± 65.8	54.3 ± 18.0	301.7 ± 100.0
pLSB0	0.0	0.0	0.0
pLSB1P−	0.1 ± 0.1	0.0	0.0
pLSB2	61.3 ± 15.4	16.9 ± 4.1	93.9 ± 22.8

(1) Mean values of 10 transfections
(2) The mean value of β-galactosidase was 0.18 ± 0.02 ng/ml for all transfections

beyond the PARP transcription initiation site (as described in ref. 25 and 26); this sequence has been shown to influence only transcription and not RNA splicing. The polyadenylation site-containing fragment following the luciferase gene has been shown (C. Clayton, personal communication) to function as such when arranged in this context. The different constructs were transfected into procyclic trypanosomes and luciferase activity was measured.

The relative luciferase activity indicated that the cells containing the pLSB1 and pLSC1 constructs produced the highest levels of enzyme activity (Table II). Although PGK B and PGK C mRNA levels normally found in the cell are different, these data indicate that the SAS B and C regions function with similar efficiencies. Translatable mRNA in these cases is dependent upon constructs complete with both a promoter and splice acceptor site, since constructs lacking either component (pLSB0 and pLSB1P−) do not produce luciferase activity. Replacement of the SAS B or C region with the SAS A one generated a chimera that produced low levels of luciferase activity in transient assays. Thus, the SAS A region appears to be inefficiently utilized in the production of luciferase mRNA in this construct. To test the importance of the size of the SAS region as well as the integrity of the sequence between the AG dinucleotide mini-exon addition site and the translational start site on luciferase enzyme levels, we transfected cells with a construct that does not contain the sequence between the AG and the translational start site of the PGK B gene. Cells transfected with this plasmid (pLSB2) showed a decreased amount of luciferase in comparison to cells transfected with the parent plasmid pLSB1. With this control plasmid we showed that the absence of the sequence 3′ to the AG splice site affects the production of luciferase. This result is due to either a decrease in splicing efficiency, possibly as a result of the loss of secondary AG splice sites present in the deleted region, or to a decrease in the translational capacity of the altered mRNA.

In all cases, we monitored transfection efficiency by co-transfecting each test construct with a β-galactosidase gene that was expressed from a PARP promoter and PARP SAS signals. We found little variation in transfection efficiency; thus, the values in the right column of Table II, where luciferase units are expressed as a function of β-gal activity, directly reflect the values in the middle column, where unstandardized luciferase activities are indicated.

Although it is convenient to score splice site utilization using transient transfections and enzyme assays as we have described, it is essential to ensure that differential luciferase activities reflect different mature mRNA levels. To do this, we performed reverse transcriptase-polymerase chain reactions (RT-PCR) using mRNA from cells transfected with either the pLSA1, B1 or C1 constructs. It was extremely difficult to obtain sufficient amounts of luciferase

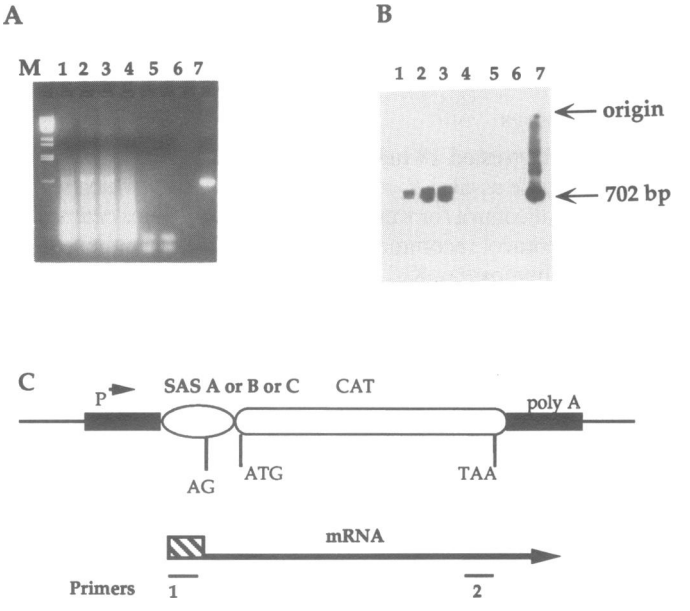

Figure 2. RT-PCR analysis of CAT-containing mRNA present in transformants. **A.** An ethidium bromide-stained 3% Nue-Sieve-1% Seakem agarose gel containing RT-PCR products using 1 μg poly A+ RNA from procyclic cells transfected with pCAT A1 (lane 1), pCAT B1 (lane 2), pCAT C1 (lane3), or no plasmid (lane 4). Lane 5 contains RT-PCR products using mRNA, but no MMLV reverse transcriptase, from cells transfected with pCAT B1. Lane 6 contains RT-PCR products when *T. brucei* genomic DNA was used in place of RNA. Lane 7 shows PCR products from the CAT coding region and is included as a control for size analysis and hybridization specificity. 'M' indicates the lane containing lambda BstE II markers. **B.** Southern blot analysis of RT-PCR products within the gel shown in panel A. The probe was a ^{32}P-labeled restriction fragment containing the entire coding sequence of CAT. Lanes correspond to those in panel A. **C.** Schematic of the pCAT A1-C1 constructs indicating the mRNA products and the oligonucleotide primers used for the RT-PCR. The mini-exon is shown as a hatched box and the direction of RNA is indicated by the arrow.

mRNA for this procedure for unknown reasons. To overcome this problem, we replaced the luciferase gene with the CAT gene in the pLSA1, B1 and C1 constructs. Transient transfection assays using these CAT-plasmids showed that relative CAT levels reflected luciferase levels (data not shown). RT-PCR was done using a primer from the 3′ end of the CAT open reading frame to generate the first strand cDNA and a mini-exon primer to ensure the amplification of only mature CAT-containing mRNA. The data from this experiment are shown in Figure 2. Predominant bands of the expected sizes (~ 750 bp) are seen in Figure 2A, lanes 1, 2 and 3, which correspond to the amplified RNA from mature CAT mRNA. These bands were identified

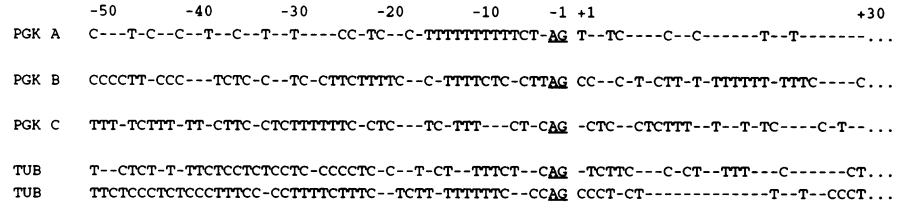

Figure 3. Comparison of the −50 to + 30 sequences within the PGK and tubulin SAS regions in *T.brucei*. To highlight the pyrimidines, all purines are expressed as dashed lines. The −1 marks the AG 3′ splice site.

as CAT-containing by Southern transfer and hybridization to radiolabeled CAT sequences (Figure 2B). A control containing mRNA, but no reverse transcriptase, one containing untransfected cellular mRNA and one containing untransfected cellular DNA (Figure 2A and B, lanes 4, 5 and 6) all demonstrate that the RT-PCR is RNA-dependent and that the CAT primer only hybridizes to RNA transcribed in transfected cells. Therefore, the RT-PCR assays demonstrated that mRNA quantities reflected reporter gene enzyme activities. Since the constructs differ only in their SAS regions, these data show that different splice site usage, during mRNA maturation, results in different CAT-containing mRNAs amounts produced in cells transfected with a plasmid containing either the PGK SAS A, B or C regions.

To discern a possible correlation between pyrimidine content and frequency of splice site utilization, we compared the 3′ SAS regions between nt position −50 and +30 of the three PGK genes (Figure 3). A dearth of pyrimidines centers around the −30 region of the PGK A sequence. In the PARP splice acceptor site, the −30 region must be pyrimidine rich for normal PARP mRNA maturation (17). Work of Curotto, et al (18) showed that, by arranging polypyrimidine tracts, each punctuated with an AG dinucleotide at the 3′ end, in tandem, and upstream from a reporter gene, an increased amount of mRNA could be obtained. In these constructs, the pyrimidine content 3′ to the initial AG dinucleotide was greater than in the parent plasmid. We compared the pyrimidine levels downstream from the 3′ splice site in the PGK genes to each other and with the α- and β-tubulin genes (Figure 3). Again, the PGK A gene is remarkable for its lack of pyrimidines in this region. This analysis supports the contention that native 3′ splice acceptor site selection is influenced by the pyrimidine content of sequences proximal to the 3′ splice site and that the PGK A gene's purine rich splice acceptor site is infrequently used for *trans*-splicing.

DISCUSSION

In this paper, we investigated whether the differential production of the PGK A, B and C mRNAs in *T. brucei* procyclic cells is caused by differential mini-exon recognition of each gene's SAS region. First, we showed that splicing required only the 186−220 nt that precede the translational start site of each gene. Then, by substituting each SAS between a transcriptional promoter and a reporter gene, we were able to assess the unique role of the SAS in affecting mRNA maturation. Thus, by altering only the SAS sequence within each construct, we were able to isolate and study the activity of only this sequence within a constant background. This is important since we know very little about the specific signals in trypanosome RNA that direct transcriptional, post-transcriptional and translational activities. Our results show that the SAS region within the PGK A gene is less efficiently utilized than are the SAS regions that precede the PGK B and C genes. Our comparative analysis of three natural splice sites that normally reside within a single pre-mRNA transcript demonstrates that the efficiency with which a mini-exon is trans-spliced onto a specific SAS is a crucial post-transcriptional regulatory step in RNA maturation in trypanosomes.

Transcriptional analysis of the PGK gene locus in both procyclic and bloodstream-form *T. brucei* cells indicated that post-transcriptional processes probably regulate the different quantities of A, B and C PGK mRNAs that are found in both these cell types. The levels of A mRNA are low in both developmental stages of the parasite, indicating a constitutive low level of pre-mRNA processing associated with this mRNA. Although B and C mRNA levels are always greater than those of A mRNA, there appears to be a difference in absolute amounts of the C mRNA between bloodstream forms and procyclics (15, 21). Since nascent transcription rates are the same in both developmental stages, there probably is developmentally-regulated mRNA processing of at least the C, and possibly the B, mRNA (15, 19, 21). Low levels of A mRNA reflect inefficient mini-exon addition to the A SAS site, e.g. inefficient *trans*-splicing. Therefore, constitutive inefficient splicing accounts for the poor synthesis of A mRNA. However, we do not observe differential splicing associated with PGK B or C mRNA maturation. It seems that developmental regulation of B and C mRNA levels is controlled not by genetic signals that reside within the 200 base pair sequences that immediately precede the open reading frame of either B or C mRNA, but by a more complex set of processes.

Analysis of VSG and PARP gene transcription initiation indicates that transcription occurs in both bloodstream-form and procyclic stages of the parasite life cycle and thus differential expression of these mRNA levels within a single life cycle stage probably is a function of dissimilar rates of transciption, pre-mRNA maturation and/or differential mRNA stability (29). By analogy, it is possible that differential mRNA stability may govern the characteristic steady state levels of B and C mRNA found in procyclic cells. Regulation of mRNA amounts in this more complex way is consistent with our data indicating that differential *trans*-splicing does not occur during the formation of these specific mRNAs. Although intact B and C mRNAs are very homologous (30, 31), the sequence between the mini-exon and the translational start site, as well as the 3′ translated region, vary considerably. These sequences, either directly or as a result of a possible interaction with the mini-exon, could affect translational efficiency and, in turn, mRNA turnover. Alternatively, differential expression of the B and C genes may be influenced by non-identical rates of polyadenylation to each mRNA during pre-mRNA maturation. This possibility is currently being investigated.

By defining short sequences that function as SAS sites we were able to highlight the role of the polypyrimidine region within the 3′ splice acceptor site as a possible effector of differential mini-exon addition. Although it has been hypothesized that this region effects mini-exon addition to a particular site, we have demonstrated that this sequence is naturally used to discriminate between SAS sites during mRNA maturation.

During the multi-step process of RNA cis-splicing in other eukaryotes, pre-spliceosome formation is characterized by the binding of the U2 snRNP to the branchpoint sequence within the pre-mRNA. This binding is preceded, as well as enhanced, by the binding of a U2 auxiliary factor (U2AF) to the polypyrimidine tract that lies between the branch point and 3′ splice site (32–34). In yeast and mammalian cells, length and pyrimidine content of the polypyrimidine tracts that occur between the 3′ splice site and the branching nucleotide have a great effect upon splicing (34). It seems probable that U2AF mediates these effects. A similar process can be envisioned to occur in trypanosomes and can account for the correlation between polypyrimidine levels and splicing efficiencies. The identification of a U2AF analog in these organisms awaits further work.

MedRNA-SAS complementarity has been proposed to mediate the efficiency of SAS usage (9). Although we have searched the PGK A, B and C SASs for non-equal amounts of medRNA-SAS complementarity, we have found the same small number of three contiguous bases, which could pair with the 'medRNA-comp site' conserved region, in all three genes (9). Therefore, it is probable that U2AF interaction with a 3′ SAS site influences splicing efficiencies among different genes and then a medRNA-SAS interaction determines the particular AG acceptor site used within a particular SAS region.

ACKNOWLEDGEMENTS

We wish to thank all members of the Cross lab for their helpful suggestions, Enrique Medina-Acosta and David Hartree for critical reading of the manuscript and G.A.M. Cross for his interest in this project. This work has been supported by an International Research Fellowship from the Fogarty Foundation (to N.K.) and by NIH grant AI29478 (to V.B.)

REFERENCES

1. Nilsen, T.W. (1989) Exp. Parasitol., **69**, 413–416.
2. Agabian, N. (1990) Cell, **61**, 1157–1160.
3. Rajkovic, A., Davis, R.E., Simonsen, J.M., and Rottman, F.M. (1990) Proc. Natl. Acad. Sci. USA, **87**, 8879–8883.
4. Tessier, L.-H., Keller, M., Chan, R.L., Fournier, R., Weil, J.-H., and Imbault, P. (1991) EMBO J., **10**, 2621–2625.
5. Clayton, C.E. (1992) Progress Nuc. Acid Res., **43**, 37–66.
6. Ullu, E., Matthews, K.R., and Tschudi, C. (1993) Mol Cell Biol., **13**, 720–725.
7. Agami, R. and Shapira, M. (1992) Nucleic Acids Res., **20**, 1804–1804.
8. Bellofatto, V., Cooper, R., and Cross, G.A.M. (1988) Nucleic Acids Res., **16**, 7437–7456.
9. Layden, R.E. and Eisen, H. (1988) Mol Cell Biol., **8**, 1352–1360.
10. Tschudi, C., Young, A.S., Ruben, L., Patton, C.L., and Richards, F.F. (1985) Proc. Natl. Acad. Sci. USA, **82**, 3998–400.
11. Sather, A. and Agabian, N. (1985) Proc. Natl. Acad. Sci. USA, **82**, 5695–5699.
12. Lee, M.G.-S. and Van der Ploeg, L.H.T. (1990) Mol Biochem Parasitol., **41**,221–232.
13. Vijayasarathy, S., Ernest, I., Itzhaki, J.E., Sherman, D.R., Mowatt, M.R., Michels, P.A.M., and Clayton, C.E. (1990) Nucl. Acids Res., **18**, 2967–2975.
14. Kapler, G.M., Zhang, K., and Beverley, S.M.,(1987) Nucleic Acids Res., **15**, 3369–3383.
15. Gibson, W.C., Swinkels, B.W., and Borst, P. (1988) J. Mol. Biol., **201**, 315–325.
16. Patzelt, E., Perry, K.L., and Agabian, N. (1989) Mol Cell. Biol., **9**, 4291–4297.
17. Huang, J. and Van der Ploeg, L.H.T. (1991) EMBO J, **10**, 3877–3885.
18. Curotto de Lafaille, M.A., Laban, A., and Wirth, D.F. (1992) Proc. Natl. Acad. Sci. USA, **89**, 2703–270
19. Osinga, K.A., *et al.* (1985) EMBO J., **4**, 3811–3817.
20. Alexander, K. and Parsons, M. (1991) Mol Biochem Parasitol., **46**, 1–10.
21. Parsons, M. and Hill, T. (1989) Mol Biochem Parasitol., **33**, 215–228.
22. Mowatt, M.R. and Clayton, C.E. (1987) Mol. Cell Biol., **7**, 2838–2844.
23. Bellofatto, V. and Cross, G.A.M. (1989) Science, **244**, 1167–1169.
24. Brun, R. and Schonenberger, M. (1979) Acta Trop., **36**, 289–292.
25. Brown, S., Huang, J., and Van der Ploeg, L.H.T. (1992) Mol. Cell. Biol., **12**, 2644–2652.
26. Sherman, D.R., Janz, L., Hug, M., and Clayton, C. (1991) EMBO J., **10**, 3379–3386.
27. Clayton, C.E., Fueri, J.P., Itzhaki, J.E., Bellofatto, V., Wisdom, G.S., Vijayasarathy, S., and Mowatt, M.R. (1990) Mol. Cell. Biol., **10**, 3036–3047.
28. Rudenko, G., Le Blanco, S., Smith, J., Lee, M.G.S., Rattray, A., and Van der Ploeg, L.H.T. (1990) **10**, 3492–3504.
29. Pays, E., *et al.* (1990) EMBO J., **9**, 3145–3151.
30. Le Blancq, S.M., Swinkels, B.W., Gibson, W.C., and Borst, P.(1988) J. Mol. Biol., **200**, 439–447.
31. Swinkels, B.W., Evers, R., and Borst, P. (1988) EMBO J., **7**,:1159–1165.
32. Ruskin, B., Zamore, P.D., and Green, M.R. (1988) Cell, **52**, 207–219.
33. Zamore, P.D., Patton, J.G., and Green, M.R. (1992) Nature, **355**, 609–611.
34. Green, M.R. (1991) Annu. Rev. Cell. Biol., **7**, 559–599.