Bidirectional silencing of RNA polymerase I transcription by a strand switch region in *Trypanosoma brucei*

Simon Haenni, Erwin Studer, Gabriela Schumann Burkard and Isabel Roditi*

Institute of Cell Biology, University of Bern, Bern, Switzerland

ABSTRACT

The procyclin genes in *Trypanosoma brucei* are transcribed by RNA polymerase I as part of 5–10 kb long polycistronic transcription units on chromosomes VI and X. Each procyclin locus begins with two procyclin genes followed by at least one *procyclin-associated gene (PAG)*. In procyclic (insect midgut) form trypanosomes, *PAG* mRNA levels are about 100-fold lower than those of procyclins. We show that deletion of *PAG1*, *PAG2* or *PAG3* results in increased mRNA levels from downstream genes in the same transcription unit. Nascent RNA analysis revealed that most of the effects are due to increased transcription elongation in the knockouts. Furthermore, transient and stable transfections showed that sequence elements on both strands of *PAG1* can inhibit Pol I transcription. Finally, by database mining we identified 30 additional *PAG*-related sequences that are located almost exclusively at strand switch regions and/or at sites where a change of RNA polymerase type is likely to occur.

INTRODUCTION

African trypanosomes (*Trypanosoma brucei* spp.) are protozoan parasites causing human African sleeping sickness and the disease Nagana in cattle. The nuclear genome of *T. brucei* contains 11 megabase chromosomes which are organized into large polycistronic transcription units (1,2). The transcription units are often separated by strand switch regions (changes of the coding strand). A comparison of the genomes of *T. brucei, Trypanosoma cruzi* and *Leishmania major* [the Tritryps, (1)] revealed that they are highly syntenic (i.e. genes are at the same relative positions). Interestingly, 25% of synteny breaks in *T. brucei* correlate with strand switch regions (H. Renauld, personal communication). *T. brucei* contains about 120 strand switches (counted manually on the chromosome maps from GeneDB). The eight-strand switch regions on *T. brucei* chromosome I have been analyzed, but did not reveal specific features, except for a lower number of single nucleotide polymorphisms, that allowed putative elements to be identified (3).

The major surface glycoproteins of procyclic (insect midgut) form trypanosomes, classified as EP and GPEET procyclins according to internal peptide repeats (4), are encoded by pairs of genes that are located on chromosomes VI and X (Figure 1). They are followed by one or more *procyclin-associated genes* (*PAGs*) that form part of the same transcription unit and potentially also encode membrane proteins (5–8). Procyclin genes and *PAGs* occur at a break in synteny between *T. brucei, Leishmania major* and *Trypanosoma cruzi* (1). Moreover, in some strains at least (5,8,9), the loci on allelic copies of chromosome X are polymorphic (Figure 1). The procyclin expression sites are transcribed by RNA polymerase I (Pol I) (5,10–12) and, in the case of the EP/PAG1 and EP/PAG2 loci on the two copies of chromosome X, overlap with a transcription unit on the opposite strand (11). It has been shown previously that *PAG4* is followed by a 'T region'; this gives rise to mature transcripts that overlap by 700 bases with mRNAs from the *GU2* gene in an antisense transcription unit (11). Transcription of the procyclin unit extends about 2.5 kb downstream of the 'T region' (in strain 427), and overlapping transcription from the opposite strand can be detected, with the strongest antisense signals between *PAG2* and *PAG4* (11). The procyclin transcription units on the two copies of chromosome VI consist of only four genes (Figure 1) (6,7,9,13). About 2–3 kb downstream of *GRESAG2.1*, the last gene in the unit, there are three elements that act synergistically to terminate transcription by Pol I (6). However, no clear

*To whom correspondence should be addressed. Tel: +41 31 6314647; Fax: +41 31 6314684; Email: isabel.roditi@izb.unibe.ch
Present address:
Simon Haenni, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, UK.

sequence motif for transcription termination could be identified, and no related sequences can be recognized downstream of the procyclin transcription units on chromosome X.

In common with the messenger RNAs of most protein-coding genes in *T. brucei*, procyclin mRNAs and *PAG* mRNAs are produced from primary polycistronic transcripts by *trans*-splicing at the 5′ end and polyadenylation at the 3′ end (14,15). *Trans*-splicing results in the addition of an ectopic spliced leader (SL) RNA at the first AG dinucleotide following a polypyrimidine tract of 8–25 nt in the intergenic region (14–16). This process is coupled to polyadenylation of transcripts covering the upstream gene and occurs at a certain distance (usually between 60–100 bases) 5′ of the SL addition site (16–18).

The steady-state mRNA levels of the *PAGs* are 100–500 times lower than those of procyclins (9,17), but the underlying regulatory mechanisms are not yet understood. In contrast to mRNAs of procyclin genes, which have most of their regulatory elements in their 3′ untranslated regions (UTRs), *PAG* mRNAs have extremely short 3′UTRs which are unlikely to play such a role [for instance, the stop codons of the major ORFs of *PAG1* and *PAG2* mRNAs are the beginning of their poly(A) tails]. However, *PAG1, PAG2* and *PAG3* mRNAs contain unusually long and conserved 5′UTRs that represent good candidates for regulatory elements. The first 640 bp of the three genes are almost identical, diverging at approximately the position where the major ORF of *PAG3* starts (7,9). *PAG1* and *PAG2* share almost complete sequence identity up to nucleotide 1721 into the middle of their major ORFs, which begin at position 1264. The 5′UTRs contain several small ORFs of unknown function. Furthermore, *PAG1* transcripts are alternatively *trans*-spliced to produce mono-, bi- and polycistronic mRNAs (17). However, it is not known what their function is or whether they are produced sequentially or independently.

In this study we show that RNA interference (RNAi) or an RNAi-related pathway depending on the argonaute protein TbAGO1 is not involved in regulating *PAG* mRNA levels. However, we demonstrate that *PAG1* contains sequence elements that promote termination of transcription by Pol I. Furthermore, in addition to the six annotated (putative) *PAG*-like genes in the genome of *T. brucei* TREU 927/4, more than 20 *PAG*-like sequences without annotation could be identified by Blast searches of contigs. Interestingly, almost all of the sequences are located at strand switch regions and/or at sites where there is likely to be a change of RNA polymerase type.

MATERIALS AND METHODS

Trypanosomes

The culture conditions for AnTat1.1 and mutants derived from it were described previously (8). *T. brucei rhodesiense* YTat 1.1 (19), and the derivatives Ago$^{-/-}$ [deficient in RNA interference (RNAi)] and TAD26 (complemented with *TbAGO1*) (20) were cultivated in SDM-79 containing 10% fetal bovine serum (FBS) and 20 mM glycerol.

Polymerase chain reactions (PCR) and primers

PCR was performed by standard procedures (21). Primer list (restriction sites are underlined and indicated in the third column):

640down	5′-TAT<u>AGATCT</u>TTCGATAAAGCCAAA-3′	BglII
640up	5′-ATA<u>TCTAGA</u>GCGGCAGTCGAC-3′	XbaI
640up2	5′-ATA<u>AAGCTT</u>GGCGCGCAGTCGAC-3′	HindIII
1240up	5′-TGA<u>TCTAGA</u>AGACCGCGGCAAGTT-3′	XbaI
1240up3	5′-ATA<u>CTCGAG</u>AAGACCGCGGCAAGTT-3′	XhoI
1240down	5′-TAT<u>GGATCC</u>TTGCACCTTCCC-3′	BamHI
PAG1down2	5′-AGC<u>AAGCTT</u>CCTTCTAAAACCTCAG-3′	HindIII
PAG1down3	5′-AAT<u>TCTAGA</u>TTCCTTCTAAAACCTCAG-3′	XbaI
PAG1up	5′-ATA<u>GCGGCCGC</u>GAGGGTCCTC-3′	NotI
PAG1up2	5′-GTG<u>TCTAGA</u>GGGTCCTCCTGAG-3′	XbaI
PAG1up3	5′-ATA<u>AAGCTT</u>GAGGGTCCTCCTGAG-3′	HindIII
PAGmid	5′-ATA<u>TCTAGA</u>GCTTTAAATTCTAGCTGC-3′	XbaI
pBSXba	5′-TAT<u>TCTAGA</u>GGTCGACGGTATCG-3′	XbaI

Constructs for stable and transient transfections

The numbering of sequences referring to *PAG1* starts at the first base after the SL addition site immediately downstream of EP2 (9). The plasmids pPAG1ko-Neo, pPAG2ko-Hygro, pPAG3ko-Phleo and pPAG3ko-Hygro to knock out whole *PAGs* were described previously (8). The construct pPAG1-Δ640-Neo for deletion of the first 640 bp of the *PAG1* 5′ untranslated region (UTR) was made as follows: PCR was performed with primers 640down and PAG1up on the genomic DNA clone λPAG2-711 [derived from the EP/PAG1 locus in AnTat1.1 (8)]. The product was digested with BglII and NotI and cloned between the BamHI and NotI sites of pPAG1ko-Neo. For the plasmid pPAG1-Δ1240-Neo, to delete the first 1240 bp of the *PAG1* 5′UTR, PCR was performed with primers 1240down and PAG1up on the genomic DNA clone λPAG2-711. The product was digested with BamHI and NotI and cloned between the corresponding sites from pPAG1ko-Neo.

Constructs for chloramphenicol acetyltransferase (CAT) assays. The shortened names of the constructs, which are used in the main text, are indicated in parentheses.

pG-mcs-CAT/EP2 (mcs): the multiple cloning site from pG-mcs-ΔLII (22) was excised with HindIII and XbaI and cloned into the corresponding sites of pGAPRONE-ble/EP1Δ164 CAT/EP2 (23). pG-neo-CAT/EP2 (neo): the neomycin resistance gene was cut out from pBS-Neo (24) with HindIII and XbaI and cloned between the corresponding sites of pG-mcs-CAT/EP2. pG-640-CAT/EP2 (640$_F$): the first 640 bp of *PAG1* were amplified by PCR on the genomic DNA clone λPAG2-711 using the primers PAG1down2 and 640up. The product was digested with HindIII and XbaI and cloned into the corresponding sites of pG-mcs-CAT/EP2. pG-640rev-CAT/EP2 (640$_R$): the first 640 bp of *PAG1* were amplified by PCR using pG-640-CAT/EP2 as template and the primers PAG1down3 and 640up2. The product was digested with HindIII and XbaI and cloned between the corresponding sites of pG-mcs-CAT/EP2. pG-640-neo-CAT/EP2 (640$_F$-neo): the neomycin resistance gene was amplified by PCR from pBS-Neo using the primers M13 (Invitrogen) and pBSXba. The product was digested with XbaI and SpeI (part of the amplified pBluescript backbone, compatible

with XbaI) and cloned into the XbaI site of pG-640-CAT/EP2. pG-1240-CAT/EP2 (1240$_F$): the first 1240 bp of *PAG1* were amplified by PCR with the primers PAG1down2 and 1240up on the genomic DNA clone λPAG2-711 and the product cloned into pCR®2.1-TOPO (Invitrogen). The insert was excised with XhoI (cuts within the vector backbone 5′ to the *PAG1* insert) and XbaI and cloned into the XhoI and XbaI sites of pG-mcs-CAT/EP2. pG-1240rev-CAT/EP2 (1240$_R$): the first 1240 bp of *PAG1* were amplified by PCR from pG-1240-CAT/EP2 with the primers PAG1down3 and 1240up3 and cloned into the XhoI and XbaI sites of pG-mcs-CAT/EP2. pG-1240-neo-CAT/EP2 (1240$_F$ -neo): the same strategy was used as for pG-640-neo-CAT/EP2, however the neo fragment was cloned into XbaI site of pG-1240-CAT/EP2. pG-PAG1end-CAT/EP2 (ORF$_F$): to obtain the fragment with the *PAG1* sequence from its internal HindIII site at 1087 to position 2458, a PCR was performed with the template pBS-711-HN (8) and the primers M13r (Invitrogen) and PAG1up2. The product was digested with HindIII and XbaI and cloned into the corresponding sites of pG-mcs-CAT/EP2. pG-PAG1end-rev-CAT/EP2 (ORF$_R$): the same strategy was used as for pG-PAG1end-CAT/EP2 using the primers PAGmid and PAG1up3. pG-662_1240-CAT/EP2 (Mid$_F$): the *PAG1* sequence from 662 to 1240 was amplified by PCR with the primers 640down and 1240up on template pG-1240-CAT/EP2. The product was digested with BglII and XbaI and cloned into the BamHI and XbaI sites of pG-mcs-CAT/EP2. pG-662_1240rev-CAT/EP2 (Mid$_R$): the construct pG-mcs-CAT/EP2 was digested with BamHI and XbaI. Then fill-in reactions with Klenow were performed with the vector backbone and with the digested PCR product which was used for pG-662_1240-CAT/EP2. The products were ligated and the correct orientation of the insert was confirmed by restriction analysis and sequencing. To convert CAT plasmids into versions that could be used for stable transformation, these were digested with HindIII, repaired with Klenow, and digested with KpnI. The plasmid backbones were gel-purified. The plasmid pGAPRONE-ble/EP1Δ164 CAT/EP2 (23) was digested with XbaI, repaired with Klenow then digested with KpnI. A fragment of 1.7 kb, containing the EP1 promoter and bleomycin resistance gene was gel-purified and ligated to the backbones described above. The veracity of the constructs was confirmed by sequencing.

Stable and transient transfection of procyclic form trypanosomes

Transient transfections and CAT assays were performed as described previously (25,26) using 10 μg CsCl-purified plasmid DNA. Stable transfections were performed as described in (8). *PAG1/PAG2* combination knockouts (PAG1ko/PAG2ko clones 1 and 2) were obtained by successive transfection of the plasmids pPAG1ko-Neo and pPAG2ko-Hygro. The plasmid pPAG1-Δ640-Neo and pPAG1-Δ1240-Neo were digested with KpnI and NotI before transfection and selected clones were named PAG1-Δ640-Neo and pPAG1-Δ1240-Neo, respectively. The pGAPRONE-ΔEP2-Neo clone in which *EP2* in the

EP/PAG1 locus is replaced by neo was obtained using the plasmid pGAPRONE-GARP/EP1-Neo (27). It was digested with XbaI and NotI before stable transfection. For stable transformation with CAT plasmids, the DNA was digested with KpnI and NotI and used to transform AnTat1.1. Transformants were selected with 1.5 μg phleomycin ml^{-1}.

RNA isolation and northern blot analysis

Total RNA isolation, enrichment of poly(A)$^+$ RNA and northern blot analysis, and the fragments for the *PAG1*, *PAG2*, *PAG3*, *PAG4* and α-tubulin probes were described previously (8). In all experiments, antibiotics were omitted from the medium for at least one week before RNA extraction. Probes were labeled with the Megaprime™ DNA labelling kit (Amersham Biosciences). The neo fragment was excised from pBS-Neo (24) with HindIII and BamHI. The GU2 probe corresponds to a full length cDNA of *GU2* (11). The GRESAG2.1 probe derives from pPAG3ko-Phleo and was produced by PCR with the primers PAG3down and GRESAG21up (8). A gene-specific probe for PAG5 was obtained by excision of the XhoI-SalI fragment from pBS-711-HN4 (8). To detect EP2 transcripts on northern blots, the oligonucleotide EP-G3 (5′-CGTATATGCAAGTGTCCTGTCGCC-3′,) specific for the *EP2* 3′UTR (27) was radioactively labeled by polynucleotide kinase following standard procedures (21). Hybridization with the 5′-labeled probe was performed overnight at 37°C; the washing conditions were 2 × 30 min at 55°C with 1× SSC, 0.05% SDS.

Nascent RNA analysis

M13 constructs and slot blotting of single-stranded DNA (ssDNA) probes. Single-stranded DNA probes were produced in M13mp18 (28) according to standard procedures (21). The probes were used for nuclear run-on analysis of nascent transcripts from both strands of the *EP/PAG1* and *EP/PAG2* loci (11). The nomenclature from (11) was adopted in this study: clones with a subscript L derive from the lower strand (with reference to the procyclins) and detect transcription from the procyclin transcription unit. Clones followed by the subscript U contain upper strand DNA and detect transcription from the antisense strand. The probes V (empty M13mp18 vector), M$_L$ (*MARP2*), E$_L$ (*EP1* procyclin) and 3$_L$/3$_U$ (*GU2*) have been described previously (11). Probe P5$_L$ is a *PAG5*-specific XhoI-SalI fragment from pBS-711-HN4 (8) which contains the sequence from 175 to 889 of the *PAG5* ORF. Single-stranded probes for *BLE* and *CAT* were produced by cloning the complete coding regions into M13mp18. For slot blotting of the ssDNA probes onto membranes, 1 μg of ssDNA was diluted in 10 μl H$_2$O, mixed with 20 μl of denaturation solution (2 M NaCl, 0.1 M NaOH) and boiled for 2 min. After quenching on ice for ~2 min, 180 μl 6× SSC were added. A piece of positively charged Nylon membrane (Roche) was soaked in 6× SSC for 5–10 min and the denatured samples were blotted onto the membrane using a Slot Blot Filtration Manifold PR 600 (Hoefer Scientific Instruments). Each slot was washed once with 200 μl

6× SSC, the membrane air-dried and the DNA UV-cross-linked to the membrane (Stratalinker: 120 mJ).

Nuclear run-on analysis. Trypanosome cultures in logarithmic phase were prepared to obtain 10^9 cells per reaction. Nuclear run-on assays were performed as described previously (29). For hybridization, the nuclear run-on transcripts were denatured for 3 min at 95°C, quenched on ice and hybridized overnight at 60°C (1–2 × 10^6 cpm/ml). The filters were washed 2× for 15 min at 60°C in prewarmed 0.2× SSC, 0.1% SDS. Signals were detected after 3–5 days exposure using a PhosphorImager (Molecular Dynamics, Sunnydale, CA, USA).

RESULTS

PAG steady-state mRNA levels are not regulated by TbAGO1

As shown in Figure 1, the procyclin transcription units on chromosome X overlap with antisense transcription units. Since *T. brucei* has a functional RNAi pathway (30), overlapping sense and antisense transcription of *PAG*s might lead to degradation of the mRNAs. This was tested using an RNAi-resistant deletion mutant (Ago$^{-/-}$) that is deficient for the argonaute gene TbAGO1, and the addback control TAD26, in which RNAi was restored (20). Figure S1 shows the northern blot analyses with total RNA from procyclic forms using *PAG1, PAG2* and *PAG3* probes from the procyclin transcription unit and the GU2 probe from the antisense transcription unit. No signals could be detected with the

PAG4 and *PAG5* probes (data not shown). Neither the *PAG*s nor *GU2* appeared to be strongly influenced by the loss of *TbAGO1*, although the relative amounts of the different spliced products of *PAG1* and *PAG2* varied between the wild-type YTat 1.1 and the two mutants (panels A and B). Thus, despite strain differences in the endogenous levels of *PAG* transcripts, we conclude that the very low levels are not due to the activity of TbAGO1.

Deletion of *PAGs* increases steady-state mRNA levels of downstream genes

It was shown previously that deletion of individual *PAG*s had no detectable phenotype (8). To investigate whether there might be compensatory changes in the steady-state mRNA levels of the remaining *PAG*s, we performed northern blot analyses on *PAG1, PAG2* and *PAG3* knockouts in strain AnTat1.1. The *PAG1* knockout (PAG1ko), *PAG2* knockout (PAG2ko), *PAG3* single knockout (PAG3ko1) and *PAG3* double knockout (PAG3ko2) have the first *PAG* of the corresponding locus deleted (8). For *PAG1* and *PAG2* combination knockouts (PAG1ko/PAG2ko), two different clones were obtained: in PAG1ko/PAG2ko clone 1, the genes *PAG1, PAG5* and *PAG2** in the EP/PAG1 locus are removed (Δ1/Δ2*, Figure 1). In PAG1ko/PAG2ko clone 2, *PAG1* in the EP/PAG1 locus and *PAG2* in the EP/PAG2 locus are deleted (Δ1/Δ2). As described previously (8), four alternatively spliced *PAG1* transcripts of about 4.0, 2.6, 1.8 and 1.4 kb can be detected in late procyclic forms of wild-type AnTat1.1 (Figure 2A). As expected, the transcripts disappear in all mutants from which *PAG1* is
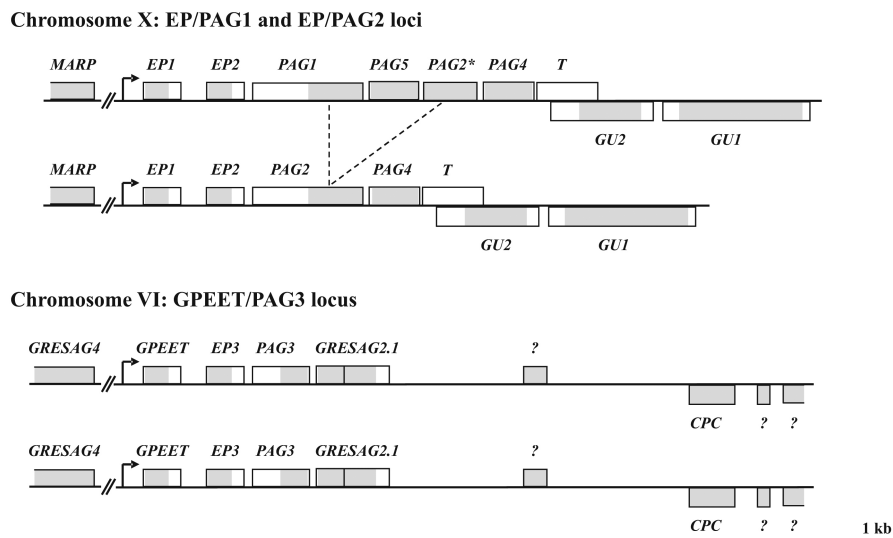

Chromosome X: EP/PAG1 and EP/PAG2 loci

Chromosome VI: GPEET/PAG3 locus

Figure 1. Genomic context of the procyclin loci in AnTat1.1. The EP/PAG1, EP/PAG2 and GPEET/PAG3 loci are named according to the first procyclin and *PAG* of each unit. Small arrows indicate transcription start sites in the promoter regions. Grey boxes represent major open reading frames (ORFs) of the corresponding genes. Untranslated regions (UTRs) are shown as white boxes (some are so small that they cannot be seen in the outline). The beginning of *PAG1* and *PAG2* and the end of *PAG2** and *PAG2* are almost identical, as indicated by a dotted line. The flanking transcription units of the GPEET/PAG3 loci were taken from GeneDB (strain 927/4, release 4) and are only partially characterized in AnTat1.1. *MARP*: gene encoding a microtubule-associated repetitive protein; *EP*: gene encoding a procyclin with internal dipeptide (EP) repeats; *GPEET*: gene encoding a procyclin with internal pentapeptide (GPEET) repeats; *PAG*: *procyclin-associated gene*; *T*: 'T region' encoding transcripts containing small ORFs of ≤240 bp (not shown); *GU*: gene of unknown function; *GRESAG*: *gene related to ESAG (expression site associated gene)*; *CPC*: gene encoding a cysteine peptidase C; *?*: gene encoding a hypothetical protein. For references see text. The figure is drawn to scale.
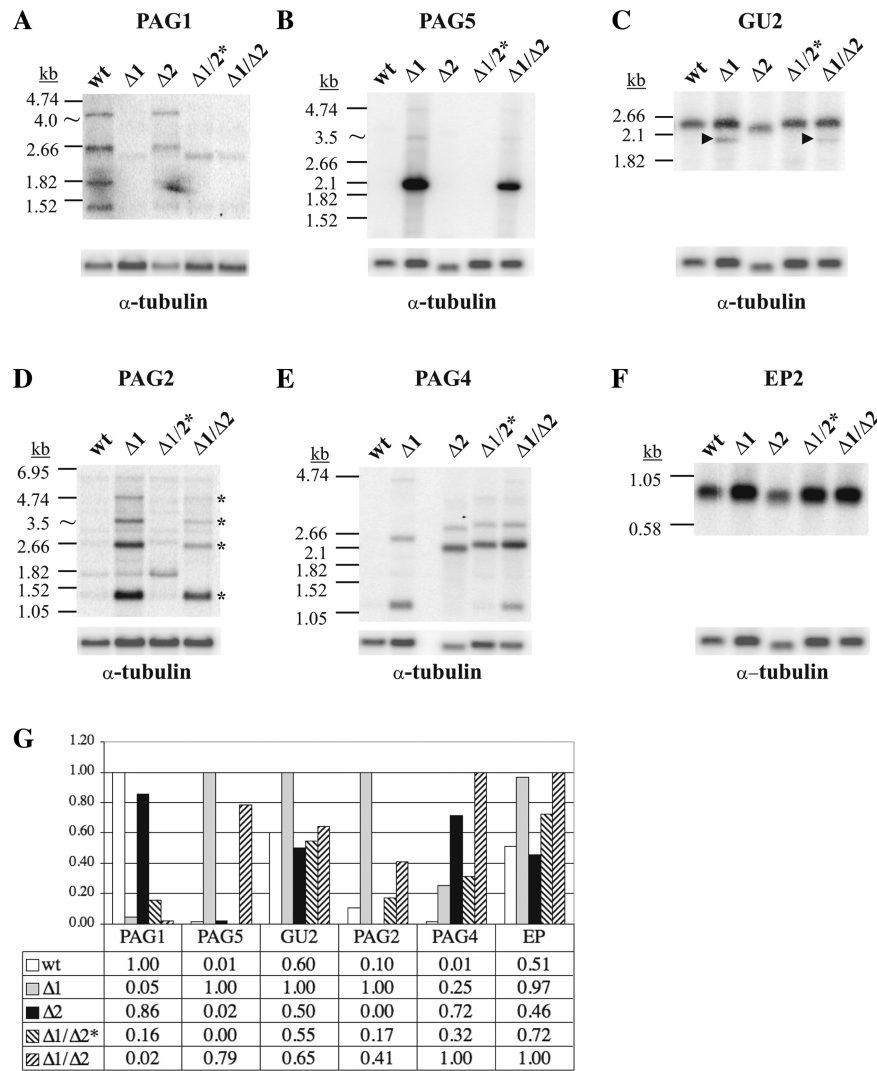
Figure 2. Northern blot analysis of *PAG* deletion mutants. Five micrograms poly(A)$^+$-enriched RNA from late procyclic cultures were loaded per lane. As a loading control, the blots were hybridized with an α-tubulin probe (lower image of each panel). (**A–E**) Megaprime-labeled probes were used as indicated above and below the images. PhosphorImager exposure times: 14 days (PAG1, PAG4, PAG5 probes), 6 days (PAG2 probe), 2 days (GU2 probe) or 5 h (α-tubulin probe). The blot in panel B was sequentially used for C and F. Deduced sizes of transcripts are marked by a (∼). (**C**) The shorter transcripts at ∼2.1 kb in the PAG1ko and PAG1ko/PAG2ko clone 2 (marked with an arrowhead) most likely represent residual PAG5 signals from the previous hybridization. (**F**) Signals were detected with a 5′-labeled oligonucleotide hybridizing to the *EP2* 3′UTR (27). PhosphorImager exposure time was 2 days. Lanes: wt: AnTat1.1wt. Δ1: PAG1ko. Δ2: PAG2ko. Δ1/2*: PAG1ko/PAG2ko clone 1. Δ1/Δ2: PAG1ko/PAG2ko clone 2. The bands marked with * in panel D indicate stable transcripts deriving from *PAG2*. (**G**) Chart showing normalized signals for the different probes as sum of the intensities of all bands. The highest value per probe was set as 1. Experiments were also performed with RNA from early procyclic cultures and the same effects were observed.

deleted (Δ1, Δ1/Δ2* and Δ1/Δ2). However, a single band at about 2.3 kb appears, which most likely derives from an ectopic *PAG1*-like sequence which had been described previously (17). This suggests that *PAG1* and *PAG1*-related sequences might weakly influence each other at the RNA level.

A surprising result was obtained when a PAG5 probe was used: although *PAG5* cannot be detected in wild-type cells, deletion of *PAG1* led to the appearance of a strongly hybridizing band at ∼2.1 kb (Figure 2B) that is a bicistronic transcript encompassing *neo* and *PAG5* (Supplementary Figure S2). In contrast, no transcripts were detectable in PAG2ko (Δ2), indicating that the phenomenon is restricted to the EP/PAG1 locus. In PAG1ko/

PAG2ko clone 1 (Δ1/2*), in which *PAG5* is also deleted, no bands were detected, confirming that the signals in the PAG1ko do not derive from ectopic *PAG5*-like sequences. For the PAG1ko/PAG2ko clone 2 (Δ1/Δ2), the same pattern was seen as in the *PAG1* knockout. This suggests an inhibitory effect of the *PAG1* sequence on *PAG5* transcription or mRNA maturation.

It has been shown previously that the 'T region' downstream of *PAG4* overlaps with the *GU2* gene in the antisense transcription unit (11). To test whether a general increase in transcription from this region might be the reason for higher levels of *PAG5* mRNA in the PAG1ko, the same northern blot was hybridized with a *GU2* probe (Figure 2C). Compared to the wild-type,

however, *GU2* signals were not significantly affected in the individual *PAG* mutants (Figure 2C), nor in a mutant from which all PAGs were deleted from the procyclin loci [allPAGko, (8) data not shown]. We therefore conclude that the increased *PAG5* signals in the *PAG1* mutants are independent of a general increase in transcription of this region.

A *PAG2* probe which detects transcripts from both *PAG2* and *PAG2** was used to further test the hypothesis that deletion of *PAG1* leads to increased levels of mRNA from downstream genes (Figure 2D). Indeed, *PAG2* signals were increased in all three *PAG1* mutants (Figure 2G). This result supports the hypothesis that deletion of *PAG*s enhances the expression of downstream genes. At least five alternatively spliced *PAG2* transcripts were detected at ~4.7, 3.5, 2.7, 1.8 and 1.4 kb in the PAG1ko (Δ1). Four of the transcripts, marked with * in lane Δ1/Δ2, derive from *PAG2** in the EP/PAG1 locus, whereas only the ~1.8 kb transcript can be attributed to *PAG2* in the EP/PAG2 locus. This is surprising since *PAG2* transcripts, but not those of *PAG2**, have a long 5′UTR containing additional putative *trans*-splicing signals. The ~1.4 kb *PAG2** transcript most likely represents a monocistronic mRNA (the *PAG2** ORF is 1257 bases long), whereas the longer *PAG2** transcripts probably also contain upstream *PAG5* and/or downstream *PAG4* sequences.

To complete this part of the analysis, *PAG4* transcripts were compared in the wild-type and deletion mutants (Figure 2E). *PAG4* is undetectable in wild-type cells, but several transcripts are visible in the four *PAG* deletion mutants. Once again, different alternatively spliced products were detected, but we were unable to determine unequivocally whether the individual transcripts derive from the EP/PAG1 locus or the EP/PAG2 locus, because *PAG4* deletion mutants were not available.

To ascertain whether upstream genes were also affected by deletion of *PAG1*, we monitored the steady-state mRNA level of EP2 procyclin in the different mutants. The procyclin transcripts were processed normally and at most, a 2-fold increase was observed (Figure 2F).

We then investigated whether the effects observed also applied to the GPEET/PAG3 locus on chromosome VI (Figure 1). For this purpose, we analyzed *GRESAG2.1* mRNA levels in *PAG3* single and double knockouts (Figure 3). No *GRESAG2.1* signals were detected in the wild-type (wt). In contrast, a faint band at ~2.3 kb appeared in the *PAG3* single knockout (Δ3/+), and two transcripts were weakly detectable in the double knockout (Δ3/Δ3). This suggests that the effect on downstream genes also holds true for this expression site. The transcripts in the *PAG3* mutants were longer than the expected size of ~1.8 kb (6). Once again it is possible that they represent bicistronic mRNAs produced by the use of alternative *trans*-splice sites upstream of the integrated antibiotic resistance genes.

Effect of deleting the *PAG1* 5′ UTR

In the previous section it was shown that the replacement of *PAG1* by the neomycin resistance gene (neo) led to
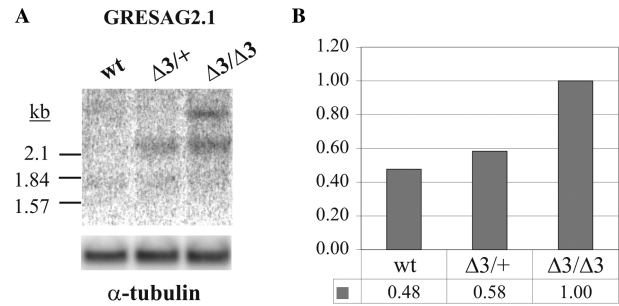

Figure 3. Effect of *PAG3* knockouts on *GRESAG2.1* mRNA levels. (**A**) Five micrograms poly(A)$^+$-enriched RNA, isolated from late procyclic cultures, were loaded per lane. The blot was hybridized with a GRESAG2.1 probe (upper panel). As a loading control, an α-tubulin probe was used (lower panel). PhosphorImager exposures: 14 days (GRESAG2.1 probe) or 5 h (α-tubulin probe). Lanes: wt: AnTat wt. Δ3/+: PAG3ko1. Δ3/Δ3: PAG3ko2. (**B**) Chart showing normalized signals as sum of the intensities of all bands, the highest value was set as 1.
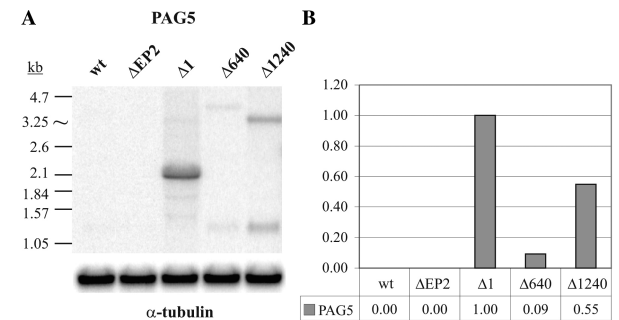

Figure 4. Effect of *PAG1* deletion on *PAG5* mRNA levels. Ten micrograms of total RNA were loaded per lane. The samples were extracted from early procyclic cultures. G418 was omitted from the medium after selection of the clones. (**A**) Northern blot with Megaprime-labeled PAG5 probe. The deduced size of the transcript is marked by a (~). The α-tubulin probe served as a loading control (lower image). Signals were detected using a PhosphorImager. Exposure times were 5 days (PAG5 probe) or 2 h (α-tubulin probe). Lanes: wt: AnTat1.1 wt. ΔEP2: GAPRONE-ΔEP2-Neo. Δ1: PAG1ko-Neo. Δ640: PAG1-Δ640-Neo. Δ1240: PAG1-Δ1240-Neo. (**B**) Chart with the normalized values as sum of all bands and with the signal of the PAG1ko set as 1. The experiment was performed twice.

a dramatic increase in the steady-state level of *PAG5* mRNA. To determine whether the long and conserved 5′UTR of *PAG1* is involved in down-regulating *PAG5* mRNA levels, the first 640 bp (shared between *PAG1*, *PAG2* and *PAG3*) and the first 1240 bp (almost the entire 5′UTR of *PAG1* and *PAG2*) were replaced by neo to generate clones Δ640 and Δ1240, respectively. In all cases, constructs were designed to leave the splice acceptor site intact. As a control, a mutant in which neo had replaced the *EP2* gene (ΔEP2) was used. When a northern blot was hybridized with the PAG5 probe (Figure 4), no stable transcripts could be detected in ΔEP2. Compared to the PAG1ko (Δ1), PAG5 signals reached ~9% and ~55% in Δ640 and Δ1240, respectively. Monocistronic *PAG5* mRNAs would have an expected size of ~1.3 kb.
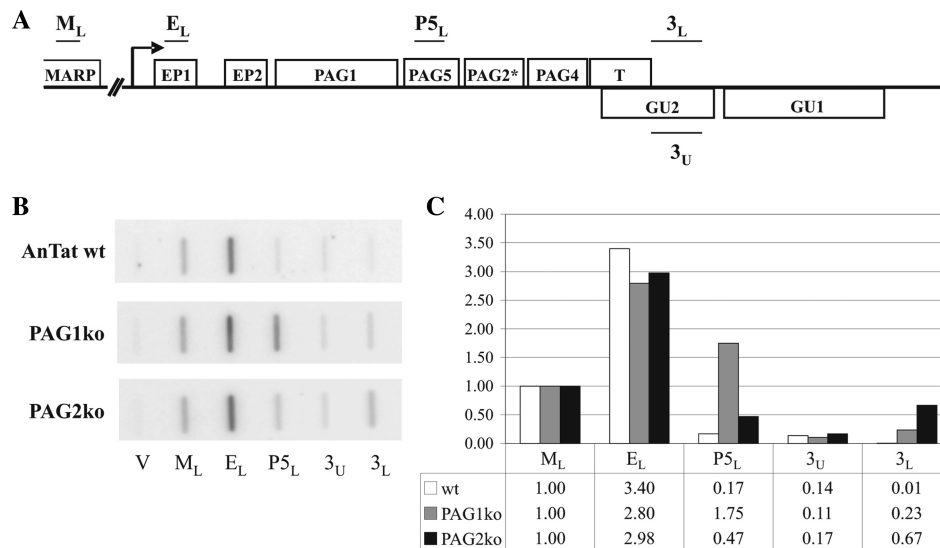
Figure 5. *PAG5* primary transcripts are increased in *PAG1* knockouts. Single-stranded genomic fragments cloned into M13mp18 were used for the detection of nascent transcripts by nuclear run-on analysis. The subscripts L and U indicate the sequence of the single-stranded probes as lower and upper strand with reference to the procyclin genes, respectively (i.e. a probe with label L detects transcripts from the sense strand of the procyclin unit). V: empty M13mp18 vector. M_L: *MARP*. E_L: *EP1*. $P5_L$: *PAG5*. $3_L/3_U$: *GU2*. (**A**) Schematic depiction of the EP/PAG1 locus with the probes used for the analysis. (**B**) Analysis of nascent transcripts from AnTat wt, PAG1ko and PAG2ko cells. (**C**) Chart with normalized signals: The *MARP* signals minus background levels (V lane) were set as 1 for each cell line. The experiment shown is representative of three independent experiments.

Interestingly, such a band was only detected in the 5′UTR mutants, suggesting that *PAG1* ORF sequences influence downstream splice site usage.

Primary transcription of *PAG5* is increased in *PAG1* knockouts

To find out whether the inhibitory effect of the *PAG1* sequence on *PAG5* mRNA production occurred co- or post-transcriptionally, nascent transcription was analyzed in wild-type cells and in PAG1ko and PAG2ko mutants (Figure 5). To detect the labeled transcripts, gene-specific single-stranded DNA probes were used as outlined in Figure 5A. Transcription of *MARP* (M_L) from the upstream Pol II transcription unit was used as an internal standard (Figure 5C). Transcription of *PAG5* ($P5_L$) was increased 10-fold in PAG1ko compared to the wild-type, while the procyclin genes (E_L) were unaffected. Increased transcription of the procyclin sense strand reached as far as *GU2* (3_L), but was weaker than for *PAG5*. However, no increase in transcription of the opposite strand was observed for *GU2* (3_U). PAG2ko showed increased transcription of the procyclin expression site in the *GU2* region (3_L), and the effect was stronger than in PAG1ko. Once again, there was no alteration in the transcription of the procyclin genes or of *GU2*. Taken together, these results suggest that deletion of *PAG1* results in loss of one or more transcription terminator elements.

Bidirectional effects of *PAG1* sequences on expression driven by Pol I

In order to define which regions of *PAG1* affect transcription we used a reporter plasmid based on the *EP/PAG1* expression site. The basic plasmid contains the *EP1*

promoter followed by a multiple cloning site (mcs), the *EP1* ΔLII 3′UTR and intergenic region (31), and a chloramphenicol acetyltransferase (CAT) gene flanked by functional *trans*-splice and polyadenylation signals from EP2 (Figure 6A). Different sequences were inserted between the *EP1* promoter and the *CAT* gene, whereby the processing signals for *CAT* transcripts were left intact. CAT activity should therefore provide a measure for the processivity of RNA polymerase I through the upstream sequences (Figure 6B). In each case, correct splicing of the *CAT* mRNA was confirmed by reverse transcription-PCR using a primer within the CAT-coding region and a primer corresponding to the SL (data not shown). CAT activity obtained with the mcs construct was set at 100%. Insertion of the ~800 bp sequence of neo led to 25% more CAT activity than the control. Interestingly, the first 640 bp of the *PAG1* 5′UTR (640_F) stimulated CAT activity by ~88% compared to the mcs construct. In contrast, introducing the first 1240 bp of *PAG1* (1240_F) reduced CAT activity to ~26% and a similar reduction was observed for the sequence containing most of the *PAG1* major ORF (ORF_F). To test if the size of the insert had an effect on CAT activity, the neo fragment was cloned immediately downstream of the 640_F and 1240_F fragments. For 640_F-neo, CAT activity was 101% compared to the mcs control. For 1240_F-neo, CAT activity was also slightly lower than for 1240_F (~21% compared to ~26%). However, since the chimeric insert 640_F-neo is longer than the 1240_F fragment, yet CAT activity was still 5-fold higher, the length of the insert does not appear to play a role in the reduction of CAT activity.

To test if the second half of the *PAG1* 5′ UTR was sufficient to reduce CAT activity; the fragment from 662
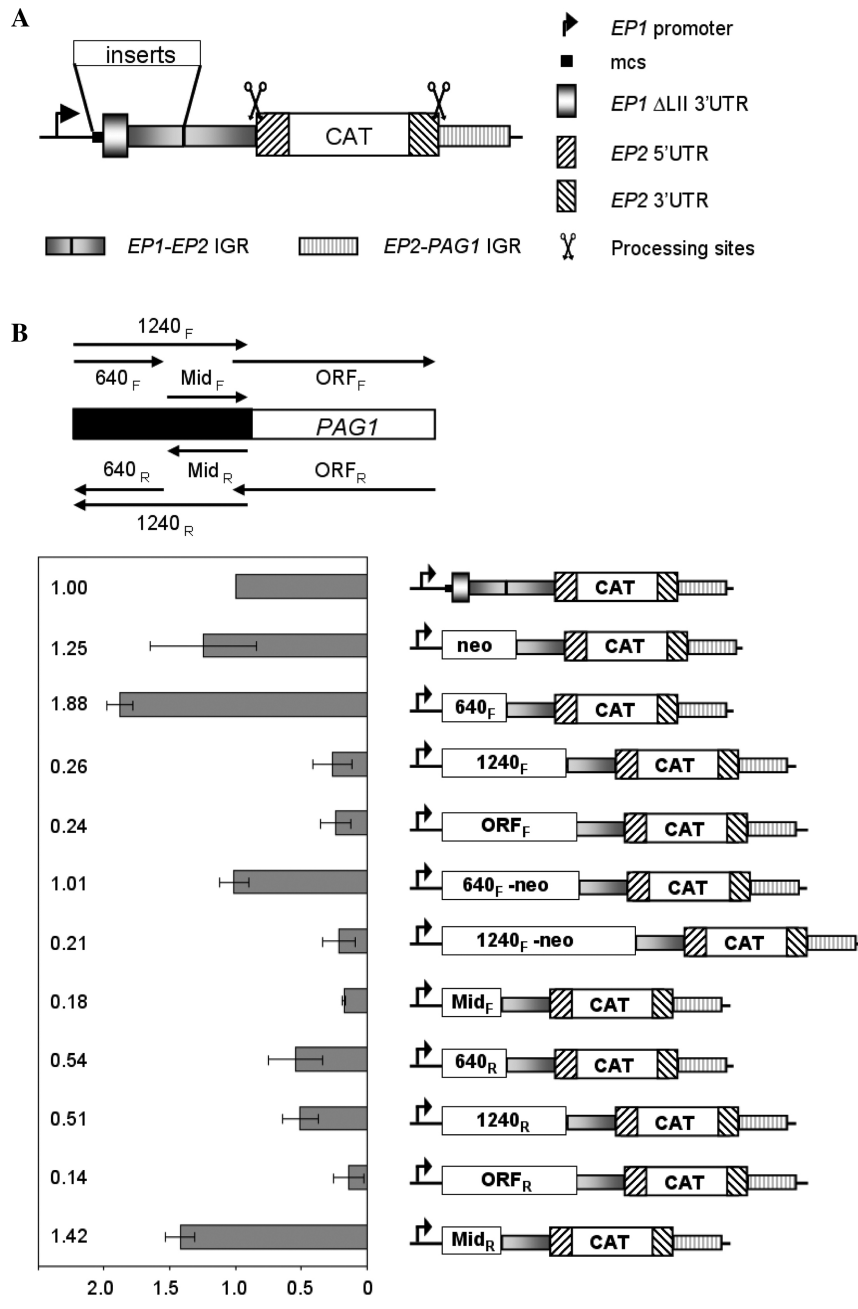
Figure 6. Effect of different regions of *PAG1* on expression of a downstream reporter gene. (**A**) Outline of the basic construct for the CAT assays [mcs construct with the *EP1* ΔLII 3′UTR (31)]. Different inserts were cloned as indicated. The CAT ORF is flanked by a 5′ splice acceptor site and 3′ polyadenylation signal derived from EP2 procyclin (indicated by scissors) and can be processed independently of the inserts. IGR: intergenic region; mcs: multiple cloning site; CAT: chloramphenicol acetyltransferase; see also legend to Figure 1. (**B**) Upper panel: overview of the *PAG1* fragments which were inserted into the mcs construct. The subscripts $_F$ and $_R$ refer to the orientation in which the inserts were cloned (forward and reverse, respectively, also indicated by the direction of the arrows). The black box represents the 1263 bp sequence of the *PAG1* 5′ UTR, the white box indicates the *PAG1* major ORF. Lower panel: chart showing mean CAT activities for the different constructs. To calculate the mean (numbers on the left side of the bars), three to four independent CAT assays were performed for each plasmid and the values normalized to those from the mcs construct.

to 1240 was inserted in the vector (Mid$_F$). The results showed a >5-fold reduction in CAT activity compared to the mcs construct, and >10-fold reduction compared to the 640$_F$ fragment. The different *PAG1* fragments were also assayed in the reverse orientation (subscripts $_R$): for 640$_R$ and 1240$_R$ CAT activities were reduced to ~54%

and ~51%, respectively. Furthermore, the ORF$_R$ fragment led to a strong reduction of CAT activity to ~14%. In contrast, the second half of the *PAG1* 5′ UTR (Mid$_R$) increased CAT activity by ~42%. Taken together, when inserted in front of the *CAT* gene, most of the *PAG1* sequences tested inhibited *CAT* expression from a Pol I
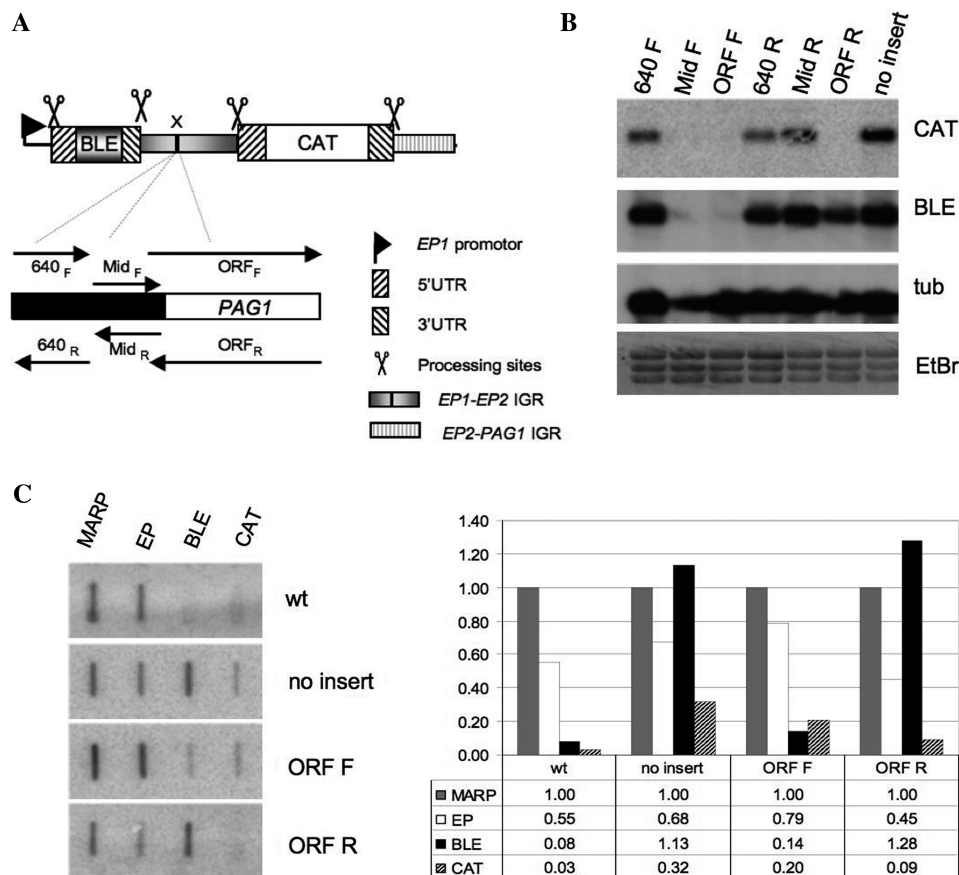
Figure 7. Stable transfection confirms that *PAG* sequences reduce transcription and steady-state levels of mRNAs from genes in the vicinity. (**A**) Schematic diagram of a bicistronic construct based on the two procyclin genes in the EP/PAG1 locus. Fragments derived from *PAG1* were inserted into the unique XbaI site (X) in the forward or reverse orientation. (**B**) Northern blots were hybridized sequentially with probes corresponding to the open reading frames of CAT and BLE, then normalized with α-tubulin. (**C**) Nuclear run-on analysis of *BLE* and *CAT* in selected clones. Other single-stranded probes are as described in the legend to Figure 5. The *MARP* signals minus background levels were set as 1 for each clone.

promoter. The two exceptions were 640_F and Mid_R, both of which stimulated CAT activity slightly.

To analyse the effects in a natural context, we tested the influence of a subset of fragments on the steady-state levels of mRNA when they were stably integrated into the genome (Figure 7A and B). For these experiments we used bicistronic constructs in which the open reading frames of EP1 and EP2 were replaced by those for phleo-myin-resistance (*BLE*) and *CAT*, respectively. In contrast to most other genes in *T. brucei*, adjacent procyclin genes have independent polyadenylation and *trans*-splicing signals (17); the *PAG* fragments were inserted at a unique site between them in the construct. Southern blot analysis confirmed that the clones used in the subsequent analysis had integrated both *BLE* and *CAT* (data not shown). Northern blot analysis corroborated the predictions from the transient transfections that Mid_F, ORF_F and ORF_R reduced expression of the gene downstream, while 640_F, 640_R and Mid_R had little or no effect. In addition, however, Mid_F and ORF_F also reduced the steady-state levels of mRNA from the upstream *BLE* gene, suggesting that different gene silencing mechanisms could be operating on the two strands. To address this, nuclear run-on analysis was performed with wild-type

procyclic forms and with clones containing either no insert between *BLE* and *CAT*, or the inserts ORF_F and ORF_R, respectively (Figure 7C). Nascent transcript levels were normalized against MARP. Consistent with the northern blot analysis, both inserts inhibited *CAT* transcription (~2- to 4-fold), and transcription of the upstream *BLE* gene was reduced ~5-fold by ORF_F. In contrast, *BLE* was transcribed at similar levels in the clone without an insert and that containing ORF_R.

The *Trypanosoma brucei* genome contains multiple *PAG*-like sequences

Since some *PAG* sequences in the procyclin loci (7,9) are involved in gene silencing, we checked whether there were similar sequences located elsewhere in the *T. brucei* genome. Blast searches in the genome database with different *PAG* sequences revealed a surprising number of additional *PAG*-related sequences in the genome. Six of them were annotated as (putative) *PAG* genes, but the others could only be found by Blast searches in unannotated contigs. Interestingly, *PAG* sequences appear to be specific to *Trypanosoma* spp. since they could not be found in the genome databases currently available for other organisms. This suggests that they have a parasite-specific

role. Furthermore, we noted that almost all *PAG*-related sequences are located at strand switch regions and/or regions where transcription changes from one polymerase to another (for instance, from rRNA or tRNA genes to transcription units containing protein-coding genes). The most significant Blast results are summarized in Supplementary Table S1 and Figure S3. Supplementary Table S1 lists all currently annotated *PAGs* and *PAG*-related genes. Interestingly, they are all located at convergent strand switch regions. However, this was not a uniform characteristic for the additional *PAG*-associated strand switch regions we identified, as some of these were divergent (not shown). The following query sequences were used for the Blast searches: the conserved first half of the *PAG1* 5′UTR (0-640) which is shared between *PAG1*, *PAG2* and *PAG3*; the beginning of the *PAG1* major open reading frame (1240–1920) which is shared with *PAG2* up to position 1721; a *PAG3*-specific sequence spanning its major ORF (540–1260). The list could not be extended by querying the second half of the *PAG1* 5′UTR (640–1240) or the end of *PAG1* (1920–2480). In addition, sequences from other *PAGs* did not lengthen the list. In total we have identified about 30 such regions. However, since we estimate the total number of strand switch regions in *T. brucei* to be about 100–120, the *PAG*-associated strand switch regions might constitute a functional subgroup.

DISCUSSION

The steady-state mRNA levels of *PAGs* are about two orders of magnitude lower than those of procyclins from the same polycistronic transcription units (9,17). Several post-transcriptional processes have been proposed to be responsible for these differences, such as processing efficiency, mRNA stability and export from the nucleus (7). The discovery that *PAGs* are located at regions of overlapping transcription (11) raised the possibility that this might lead to the formation of double-stranded RNAs and result in their silencing by an RNAi-related mechanism (32,33). To test this hypothesis we analyzed the steady-state mRNA levels of the *PAGs* in cells deficient in *TbAGO1* (20), but observed only minor changes in the mRNA levels and splice patterns. It can therefore be concluded that RNAi(-related) mechanisms that depend on *TbAGO1* are not involved in the down-regulation of *PAG* transcripts.

Deletion of the first *PAG* in each procyclin locus led to a significant increase in mRNA derived from the downstream genes. This effect was locus-specific and restricted to genes transcribed from the same strand. In northern blot analyses with *PAG* mutants, complex patterns of alternative splicing, sometimes involving the production of long, bi- or polycistronic steady-state transcripts, were often observed. Some of these splice variants are very likely to contain sequences transcribed from the integrated antibiotic resistance genes and therefore do not represent natural mRNAs. Interestingly, at least part of *PAG1* appears to be essential for the production of monocistronic *PAG5* mRNAs. This suggests that the *PAG1* major

ORF harbors one or more enhancer elements for its polyadenylation and/or the *trans*-splicing of *PAG5* transcripts. This is reminiscent of exonic splicing enhancer sequences found in some eukaryotic genes (34).

A comparison of nascent transcription in wild-type cells and deletion mutants showed that the gene silencing effects occurred co-transcriptionally. This provides a new and additional mechanism for regulating the mRNA levels of genes from the same transcription unit. Only a single region of transcription termination has been documented previously in trypanosomes (6) and this occurs downstream of the GPEET/PAG3 locus rather than within (potential) coding regions. Deletion analysis implicated two regions of *PAG1*, positions 640–1240 in the 5′ UTR ('Mid') and the ORF, as being responsible for lowering the expression of downstream genes. This was corroborated by transient transfection assays in which each of these sequences reduced Pol I-driven expression of a *CAT* reporter gene ∼4-fold. The ORF also reduced expression 7-fold when inserted in the reverse orientation. These effects were independent of length of the inserted fragments and the distance of *CAT* from the promoter. The same fragments inhibited *CAT* expression when the construct was stably integrated, but there were differences in effects on the upstream gene, depending on the orientation. In the forward orientation relative to the procyclin promoter, Mid_F and ORF_F sequences caused a reduction in transcripts from the upstream gene (*BLE*) as well as the downstream *CAT* gene. In contrast, the Mid_R region did not affect transcript levels when it was inserted in the reverse orientation, and ORF_R reduced the steady-state level of transcripts from the *CAT* gene, but not the *BLE* gene. At present we do not know if these sequences also modulate expression of protein-coding genes by Pol II as such promoters have not been identified unequivocally in trypanosomes (35).

The fact that Mid_F and ORF_F affect the transcription and steady-state levels of downstream and upstream genes, while ORF_R only inhibits expression of the downstream gene (Figure 7) indicates that more than one mechanism is at work. One possibility is that some *PAG* sequences interfere with efficient *trans*-splicing of transcripts from the downstream gene and/or with cleavage and polyadenylation of transcripts from the upstream gene. Such interference could lead to the degradation of the initially cleaved primary transcripts by exoribonucleases before they can acquire a protective cap structure and/or poly(A) tail. The degradation of unprotected transcripts by exoribonucleases could also provide an explanation for how the actual gene silencing process is initiated since the recognition of functional poly(A) sites and the recruitment of termination factors such as exoribonucleases are critical events in the current transcription termination models for other eukaryotes (36–39). An additional possibility is that the polymerase is physically impeded by the chromatin structure (see below) or by proteins binding directly to the DNA. However, trypanosomes do not encode a recognizable homologue of TTF-I/Reb1p, a factor involved in transcription termination by RNA Pol I in mammals and yeast, respectively (40). Intriguingly, it has recently been demonstrated that

knockdown of the RNA-binding protein TbDRDB3/
PTB1 results in increased levels of *PAG4* mRNA without
altering the half-life of the message (41). An increase in
PAG5 mRNA was observed independently by Stern *et al.*
(42) and steady-state levels of other *PAG* mRNAs also
increased when TbDRDB3/PTB1 was depleted (Antonio
Estévez, personal communication). It could be of interest
to establish whether this protein interacts with any of the
elements that we have identified.

Analysis of the *T. brucei* database revealed at least
six *PAG*-like genes and about 30 additional *PAG*-like
sequences that are located almost exclusively at strand
switch regions and/or sites where the transcribing RNA
polymerase changes (Supplementary Table S1 and
Supplementary Figure S3), where they may be important
for preventing interference between different transcription
units. They are not universal, however, as *PAG*-like
sequences are not detected in the other members of the
Tritryps, nor do they occur at the borders of most strand
switch regions in *T. brucei*. It is possible that the *PAG*-
associated strand switch regions consist of various sub-
groups that fulfil different functions and that the *PAG*-
like sequences at divergent and convergent strand switch
regions have separate roles. It has recently been shown
that histone H4K10ac is enriched at strand switch regions
harboring putative transcription start sites, while histone
variants H3V and H4V accumulate at putative termina-
tion sites (43). The strand switch region that we have
characterized on chromosome X and all *PAGs* listed in
Supplementary Table S1 are located in regions that are
enriched for histone H3V (Nicolai Siegel, personal com-
munication). In contrast, *PAG*-like sequences at divergent
strand switch regions map to sites shown to be enriched
for H4K10ac (43). Interestingly, *PAG*-like sequences that
appear to be in the middle of a polycistronic transcription
unit on chromosome IX (marked as 'not assigned'
Supplementary Figure S3) coincide with an H3V binding
site. This is followed by a site enriched for H4K10ac,
suggesting that transcription termination and (re)initia-
tion might occur in this region.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. El-Sayed,N.M., Myler,P.J., Blandin,G., Berriman,M., Crabtree,J., Aggarwal,G., Caler,E., Renauld,H., Worthey,E.A., Hertz-Fowler,C. *et al.* (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science*, **309**, 404–409.
2. Berriman,M., Ghedin,E., Hertz-Fowler,C., Blandin,G. and Renauld,H. (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science*, **309**, 416–422.
3. Hall,N., Berriman,M., Lennard,N.J., Harris,B.R., Hertz-Fowler,C., Bart-Delabesse,E.N., Gerrard,C.S., Atkin,R.J., Barron,A.J., Bowman,S. *et al.* (2003) The DNA sequence of chromosome I of an African trypanosome: gene content, chromosome organisation, recombination and polymorphism. *Nucleic Acids Res.*, **31**, 4864–4873.
4. Roditi,I. and Clayton,C. (1999) An unambiguous nomenclature for the major surface glycoproteins of the procyclic form of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **103**, 99–100.
5. König,E., Delius,H., Carrington,M., Williams,R.O. and Roditi,I. (1989) Duplication and transcription of procyclin genes in *Trypanosoma brucei*. *Nucleic Acids Res.*, **17**, 8727–8739.
6. Berberof,M., Pays,A., Lips,S., Tebabi,P. and Pays,E. (1996) Characterization of a transcription terminator of the procyclin PARP A unit of *Trypanosoma brucei*. *Mol. Cell Biol.*, **16**, 914–924.
7. Roditi,I., Furger,A., Ruepp,S., Schürch,N. and Bütikofer,P. (1998) Unravelling the procyclin coat of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **91**, 117–130.
8. Haenni,S., Renggli,C.K., Fragoso,C.M., Oberle,M. and Roditi,I. (2006) The procyclin-associated genes of *Trypanosoma brucei* are not essential for cyclical transmission by tsetse. *Mol. Biochem. Parasitol.*, **150**, 144–156.
9. Koenig-Martin,E., Yamage,M. and Roditi,I. (1992) A procyclin-associated gene in *Trypanosoma brucei* encodes a polypeptide related to ESAG 6 and 7 proteins. *Mol. Biochem. Parasitol.*, **55**, 135–145.
10. Pays,E., Coquelet,H., Tebabi,P., Pays,A., Jefferies,D., Steinert,M., Koenig,E., Williams,R.O. and Roditi,I. (1990) *Trypanosoma brucei*: constitutive activity of the VSG and procyclin gene promoters. *EMBO J.*, **9**, 3145–3151.
11. Liniger,M., Bodenmüller,K., Pays,E., Gallati,S. and Roditi,I. (2001) Overlapping sense and antisense transcription units in *Trypanosoma brucei*. *Mol. Microbiol.*, **40**, 869–878.
12. Günzl,A., Bruderer,T., Laufer,G., Schimanski,B., Tu,L.C., Chung,H.M., Lee,P.T. and Lee,M.G. (2003) RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot. Cell*, **2**, 542–551.
13. Berberof,M., Pays,A. and Pays,E. (1991) A similar gene is shared by both the variant surface glycoprotein and procyclin gene transcription units of *Trypanosoma brucei*. *Mol. Cell Biol.*, **11**, 1473–1479.
14. Huang,J. and Van der Ploeg,L.H. (1991) Maturation of polycis-tronic 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.
15. Huang,J. and Van der Ploeg,L.H. (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.
16. Benz,C., Nilsson,D., Andersson,B., Clayton,C. and Guilbride,D.L. (2005) Messenger RNA processing sites in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **143**, 125–134.
17. Vassella,E., Braun,R. and Roditi,I. (1994) Control of polyadenyla-tion 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.
18. Schürch,N., Hehl,A., Vassella,E., Braun,R. and Roditi,I. (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.

19. Ruben,L., Egwuagu,C. and Patton,C.L. (1983) African trypanosomes contain calmodulin which is distinct from host calmodulin. *Biochim. Biophys. Acta*, **758**, 104–113.

20. Shi,H., Djikeng,A., Tschudi,C. and Ullu,E. (2004) Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance. *Mol. Cell Biol.*, **24**, 420–427.

21. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

22. Furger,A., Jungi,T.W., Salomone,J.Y., Weynants,V. and Roditi,I. (2001) Stable expression of biologically active recombinant bovine interleukin- 4 in *Trypanosoma brucei*. *FEBS Lett.*, **508**, 90–94.

23. Vassella,E., Acosta-Serrano,A., Studer,E., Lee,S.H., Englund,P.T. and Roditi,I. (2001) Multiple procyclin isoforms are expressed differentially during the development of insect forms of *Trypanosoma brucei*. *J. Mol. Biol.*, **312**, 597–607.

24. Ruepp,S., Furger,A., Kurath,U., Renggli,C.K., Hemphill,A., Brun,R. and Roditi,I. (1997) Survival of *Trypanosoma brucei* in the tsetse fly is enhanced by the expression of specific forms of procyclin. *J. Cell Biol.*, **137**, 1369–1379.

25. Hehl,A., Vassella,E., Braun,R. and Roditi,I. (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.

26. Zomerdijk,J.C., Ouellette,M., ten Asbroek,A.L., Kieft,R., Bommer,A.M., Clayton,C.E. and Borst,P. (1990) The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *EMBO J.*, **9**, 2791–2801.

27. Urwyler,S., Vassella,E., Van den Abbeele,J., Renggli,C.K., Blundell,P., Barry,J.D. and Roditi,I. (2005) Expression of procyclin mRNAs during cyclical transmission of *Trypanosoma brucei*. *PLoS Pathog.*, **1**, e22.

28. Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103–119.

29. Amiguet-Vercher,A., Pérez-Morga,D., Pays,A., Poelvoorde,P., Van,X.H., Tebabi,P., Vanhamme,L. and Pays,E. (2004) Loss of the mono-allelic control of the VSG expression sites during the development of *Trypanosoma brucei* in the bloodstream. *Mol. Microbiol.*, **51**, 1577–1588.

30. Ngo,H., Tschudi,C., Gull,K. and Ullu,E. (1998) Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc. Natl Acad. Sci. USA*, **95**, 14687–14692.

31. Furger,A., Schürch,N., Kurath,U. and Roditi,I. (1997) Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation. *Mol. Cell Biol.*, **17**, 4372–4380.

32. Matzke,M.A. and Birchler,J.A. (2005) RNAi-mediated pathways in the nucleus. *Nat. Rev. Genet.*, **6**, 24–35.

33. Zamore,P.D. and Haley,B. (2005) Ribo-gnome: the big world of small RNAs. *Science*, **309**, 1519–1524.

34. Blencowe,B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.*, **25**, 106–110.

35. Ben Amar,M.F., Jefferies,D., Pays,A., Bakalara,N., Kendall,G. and Pays,E. (1991) The actin gene promoter of *Trypanosoma brucei*. *Nucleic Acids Res.*, **19**, 5857–5862.

36. Connelly,S. and Manley,J.L. (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.*, **2**, 440–452.

37. Kawauchi,J., Mischo,H., Braglia,P., Rondon,A. and Proudfoot,N.J. (2008) Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev.*, **22**, 1082–1092.

38. Logan,J., Falck-Pedersen,E., Darnell,J.E. Jr. and Shenk,T. (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc. Natl Acad. Sci. USA*, **84**, 8306–8310.

39. Luo,W., Johnson,A.W. and Bentley,D.L. (2006) The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev.*, **20**, 954–965.

40. Jansa,P. and Grummt,I. (1999) Mechanism of transcription termination: PTRF interacts with the largest subunit of RNA polymerase I and dissociates paused transcription complexes from yeast and mouse. *Mol. Gen. Genet.*, **262**, 508–514.

41. Estévez,A.M. (2008) The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. *Nucleic Acids Res.*, **36**, 4573–4586.

42. Stern,M.Z., Gupta,S.K., Salmon-Divon,M., Haham,T., Barda,O., Levi,S., Wachtel,C., Nilsen,T.W. and Michaeli,S. (2009) Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism. *RNA*, **15**, 648–665.

43. Siegel,T.N., Hekstra,D.R., Kemp,L.E., Figueiredo,L.M., Lowell,J.E., Fenyo,D., Wang,X., Dewell,S. and Cross,G.A. (2009) Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev.*, **23**, 1063–1076.