Anatomy of the *parp* gene promoter of *Trypanosoma brucei*

David R. Sherman^{1,3}, Lennart Janz²,
Michael Hug² and Christine Clayton^{1,2,4}

¹The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA and ²Zentrum für Molekulare Biologie, Im Neuenheimer Feld 282, Postfach 106249, D-6900 Heidelberg, FRG

³Present address: Department of Molecular Microbiology, Washington University Medical School, 660 Euclid Avenue, St Louis, MO 63110, USA

⁴Corresponding author (at Heidelberg)

Communicated by P. Overath

While growing in the tsetse fly, *Trypanosoma brucei* expresses a major surface glycoprotein, the procyclic acidic repetitive protein (PARP). The *parp* genes are transcribed by an α -amanitin-resistant RNA polymerase. We have determined the sequence requirements for *parp* promoter activity. Studies of RNA produced from input DNA in transiently transfected trypanosomes indicate that the RNA is correctly processed by *trans*-splicing and polyadenylation. Deletion analyses show that 330 bp are sufficient for full promoter and splicing activity and that the promoter structure is complex, involving at least three elements whose mutual spacing is important. Mutagenesis pin-pointed two sequences vital for promoter activity; neither bears any resemblance to known prokaryotic or eukaryotic promoter elements.

Key words: *parp* gene/promoter activity/*trans*-splicing/*Trypanosoma brucei*

Introduction

Trypanosoma brucei, the causative agent of African sleeping sickness, is a protozoan parasite responsible for significant human and animal morbidity. Evolutionarily, it is one of the earliest and most divergent eukaryotes yet studied (Sogin *et al.*, 1986, 1989). Perhaps for this reason, the study of trypanosome molecular biology is replete with paradox. While unusual phenomena such as *trans*-splicing of mRNAs (Agabian, 1990) and post-transcriptional editing of many mitochondrial RNAs (Stuart, 1991) have been discovered and extensively characterized in trypanosomes, many basic processes remain unexplored. Central among these is trypanosome transcription initiation. To date, the precise sequence requirements for trypanosome promoter activity remain undefined.

T. brucei lives extracellularly in the blood and tissue fluids of its mammalian hosts, evading the immune response by frequently changing its glycoprotein surface coat (the variant surface glycoprotein or VSG). This antigenic variation is mediated via a combination of genetic rearrangement and transcriptional control (Donelson, 1989; Zomerdijs *et al.*, 1990; Gottesdiener *et al.*, 1991). The trypanosomes are transmitted from one mammal to the next by tsetse flies.

While in the flies, the trypanosomes (known as 'procyclic forms') lack VSG but express on their surface the procyclic acidic repetitive protein (PARP). Both PARP and VSG genes are transcribed by an RNA polymerase with the α -amanitin-resistance of RNA polymerase I and the divalent cation dependence of RNA polymerase II (Grondal *et al.*, 1989; König *et al.*, 1989; Rudenko *et al.*, 1989). Whereas the VSG genes are telomeric, lying at the end of polycistronic transcription units that can be >60 kb long (Donelson, 1989; Zomerdijs *et al.*, 1990; Gottesdiener *et al.*, 1991), the *parp* promoter probably lies within ~200 bp of the start of the mature mRNA sequence (Clayton *et al.*, 1990; Rudenko *et al.*, 1990). The *parp* and VSG promoters have very little sequence homology.

Analysis of trypanosome transcription has been hampered by numerous technical difficulties. Transcription is believed to be polycistronic, involving primary transcripts of up to at least 70 kb and the 5' ends of mRNAs are rapidly lost to *trans*-splicing; thus transcription initiation sites have been very difficult to identify. Sequence homology searches, which proved so useful in analyzing promoters of other species (Rosenberg and Court, 1979; Bucher and Trifonov, 1986), have not convincingly identified any known promoter elements upstream of trypanosome protein-coding genes (Clayton *et al.*, 1990; Zomerdijs *et al.*, 1990; Jefferies *et al.*, 1991). Finally, until recently, trypanosomes seemed refractory to all experimental forms of genetic manipulation.

We present here a detailed genetic dissection of the *parp* promoter region, indicating that novel *cis*-acting sequences are involved in the generation of *parp* transcripts in *T. brucei*.

Results

RNA analysis

CAT mRNA can be detected readily in cells transiently transfected with constructs in which transcription is driven by the *parp* promoter. Using standard Northern analysis with the gel-purified CAT gene fragment as probe, we observed a signal in RNA isolated just 1 h after transfection. This signal was most intense in RNA isolated 7 h post-transfection (data not shown), so that time-point was used in all subsequent RNA analyses. A typical Northern blot, of RNA from procyclic trypanosomes transfected with pJP25, is shown in Figure 1. The two bands generated clearly represent RNA since they are resistant to DNase (lane 5) and are susceptible to RNase (lane 6) and base (lane 8). In addition, plasmid DNA migrates to a different region of the gel (lanes 7 and 9). The CAT mRNAs are polyadenylated, as demonstrated by probing total RNA previously separated into poly(A)⁺ and poly(A)⁻ fractions (lanes 1 and 2).

While the precise 3' ends of the CAT mRNAs have not been mapped, the size of the 1.5 kb transcript is consistent with mini-exon addition at the normal *parp* splice acceptor site (part of the promoter region) and a polyadenylation event occurring within the *parp* 3' untranslated region of pJP25.

Fig. 1. Northern blot of RNA from cells transfected with pJP25. A formaldehyde gel was loaded with either 10 μ g total (unless otherwise noted) RNA (R) or 1 ng pJP25 DNA (D), blotted and hybridized with a purified CAT gene probe. Lanes are as follows: 1, poly(A)⁻ RNA from 10 μ g total RNA; 2, poly(A)⁺ RNA fraction from 10 μ g total RNA; 3, nothing; 4, RNA; 5, RNA treated with DNase; 6, RNA treated with RNase; 7, DNA treated with RNase; 8, RNA treated with base; 8, DNA treated with base.

Fig. 2. Primer extension of CAT RNA. Primer extension of 25 μ g total RNA from procyclic trypanosomes after transfection with pJP25 (lane e). The products were run on a sequencing gel next to a sequencing ladder (GATC) made using the same primer (VIB 13). Positions of major 5'-ends are indicated relative to the splice acceptor site.

The 2.8 kb transcript is probably polyadenylated within the vector sequence. Plasmid pJF6, which lacks the *parp* 3' untranslated region, generates only the larger RNA and gives 2–3 times less RNA and CAT activity. Plasmid pJP25 has a short *Leptomonas seymouri* sequence downstream of the *parp* 3' untranslated region. A derivative in which this sequence has been deleted, pJP44, produces CAT activity equivalent to pJP25 but only yields the smaller RNA species (data not shown). Thus the *L.seymouri* sequence does not

Fig. 3. Sequence of the *parp* promoter (A locus) and 5'-untranslated region up to and including the PARP initiation codon. In the expression plasmids the last two base-pairs are GC and are part of the *Hind*III site at the start of the CAT cassette. Nucleotides are numbered backwards from the *trans*-splicing acceptor site; the approximate position of transcription initiation is in bold type. Nucleotides which are different in the *parp* B locus are indicated in parentheses underneath the corresponding A locus positions. The alteration giving a *Sma*I site in pJP44 (mutating 4 bp and inserting 1 bp) is shown in italics underneath nucleotides -46 to -50; changed nucleotides are in lower case type. The region of homology with the VSG promoter (from position -432 to -410 relative to the principal VSG transcription start; non-homologous nucleotides in lower-case type) is shown in italics underneath nucleotides -236 to -258 and a possible alignment with the *T.brucei* rRNA promoter (initiation site in bold type) is indicated similarly from -107 to -81. The *Sma*I (pJP44), *Spe*I and *Nsi*I sites are indicated and the three direct repeats of the sequence PyPuTTTTPuTG are underlined. An A-rich sequence with poor resemblance to a TATA box is doubly underlined.

influence CAT expression, but does appear to reduce the efficiency of polyadenylation at the authentic *T.brucei* site.

The 5' ends of the CAT RNAs produced after transfection with pJP25 were investigated by primer extension analysis, using an oligonucleotide specific for the 5' end of the CAT coding region (Figure 2). A single predominant product was generated, mapping to 34 bases upstream of the *parp* AG dinucleotide splice acceptor site (Figure 3). We conclude that a mini-exon has been added to the 5' end of CAT RNA in the physiologically relevant position and that the different CAT RNAs have identical 5' ends. Reverse transcriptase–polymerase chain reaction (RT-PCR) analysis using CAT-specific and mini-exon oligonucleotides supported this conclusion, as a single band of the predicted size was produced in a reverse transcriptase-dependent manner (data not shown). The primer extensions reproducibly yielded an additional, weak signal at -82 relative to the splice acceptor site (Figure 2). This band could represent the 5' end of the primary transcript, although other interpretations are possible.

Promoter deletions

Plasmid pJP25 contains 623 bp of *parp* upstream sequence. In order to define the promoter region more narrowly, we created a series of 5' deletions in it. Our preliminary assays (Clayton *et al.*, 1990) were performed without regard for the background sequence of the constructs. We later discovered that the nature of the 3'-untranslated region can affect the level of CAT expression, and that (particularly

Fig. 4. Deletions and rearrangements in the *parp* promoter region. The pJP clone designation number is shown on the left. On the right is shown the CAT activity produced in transfected cells, calculated relative to the activity from the relevant control clone (either pJP25 or pJP44). The promoter region of pJP25 with four enzyme sites (*Nsi*I, *Spe*I, *Sca*I, *Sau*3A1) is shown at the top for reference, and the putative transcription initiation position indicated below the line, just upstream of the *Sca*I site. Each promoter fragment is bounded by a 5' *Kpn*I site and a 3' *Hind*III site. In pJP44 derivatives, the position of the inserted *Sma*I site is indicated as a vertical line (directly below the *Sau*3A1 site in pJP25). Large deletions are indicated by dashed lines. Mutated regions are stippled. V, VSG promoter homology region; R, repeat regions. *, splice acceptor site. For precise clone descriptions see Table I.

in the absence of an authentic trypanosome 3' end) vector orientation can also affect results (C. Clayton and D. R. Sherman, unpublished). All subsequent experiments were therefore performed using controls with identical downstream regions and vector orientation. Care was taken to leave the splice acceptor region (−55 to +33, Figure 3) (J. Huang and L. H. T. Van der Ploeg, Columbia University, personal communication) intact. The results are illustrated in Figure 4. A new vector, plasmid pJP44, incorporates three significant changes. The downstream *L. seymouri* sequence in pJP25 has been deleted, a *Sma*I site was introduced at −55, upstream of the splice acceptor, and the promoter region was deleted to position −330 (Figure 1). These changes do not alter the level of CAT expression. Deletion to position −274 (pJP51) reduces expression by no more than 30%; deletion to position −196 (pJP10;27;52) lowers activity by 80–90%. Further deletion to position −55 (pJP28) reduces CAT expression to the background levels that are usually observed when a splice acceptor site is present (Bellofatto and Cross, 1989; Laban and Wirth, 1989; Clayton *et al.*, 1990). To confirm these findings, the trypanosomatid sequences in pJP44 and each of the dele-

tions were recloned into vectors where luciferase was the reporter molecule instead of CAT. In each case equivalent results were obtained (data not shown). We conclude that the region between positions −330 and −55 is needed for maximal CAT expression but that not all of this region is essential.

The *parp* promoter region contains some intriguing sequence motifs. Especially noteworthy is the area from −191 to −135, which includes three direct repeats of the sequence PyPuTTTTPuTG (Figure 3). Any one of these repeats could potentially form an inverted repeat with a complementary sequence (−168 to −174) in the same region. To determine whether the direct/inverted repeats have any functional significance for promoter activity, several constructs were tested. Internal deletion of either the upstream portion (−191 to −155, including the inverted repeat and two direct repeats; pJP46) or the downstream portion (−155 to −55, with one direct repeat; pJP50) both abolished promoter activity (Figure 4). To analyze further the requirement for the inverted repeat region, we replaced the deletion of pJP46 with the appropriate number of randomly chosen nucleotides (pJP49). This restored CAT

Table I. Clones

pJF6	<i>parp</i> promoter (640 nt) made by PCR inserted upstream of CAT cassette. Downstream of the CAT cassette is an <i>L.seymouri</i> mini-exon terminator (there by accident). Details in Clayton <i>et al.</i> (1990).	pJP44.41	Sequence -95 to -86 replaced by AAGGCCTAAA (<i>StuI</i> site)
pJP10	<i>KpnI</i> - <i>SpeI</i> deletion of pJF6 (i.e. contains proximal promoter region up to position -196).	pJP44.335	Sequence -265 to -256 replaced by GCAGATCTGC (<i>BglII</i> site)
pJP14	<i>KpnI</i> - <i>Sau3AI</i> deletion of promoter in pJF6 (i.e. contains only promoter region +34 to -58). Made by purifying the promoter-CAT <i>Sau3AI</i> fragment (CAT has no <i>Sau3AI</i> site) and recloning into the <i>BamHI</i> site of a version of pGem which had the <i>HindIII</i> site destroyed by end-filling. Has no <i>L.seymouri</i> mini-exon terminator	pJP44.345	Sequence -255 to -246 replaced by CCCTCGACGG
pJP21	<i>parp</i> promoter and CAT (<i>KpnI</i> - <i>BamHI</i>) from pJF6 cloned in opposite direction into pGem4	pJP44.355	Sequence -245 to -236 replaced by GTCTGTGCAC (<i>SalI</i> site)
pJP25	Cloning of the <i>parp</i> 3'-untranslated region and polyadenylation site into pJF6 between <i>BamHI</i> and <i>PstI</i> sites. Contains the sequence from position 1512 to 1872 of the <i>parp</i> B α gene (Clayton <i>et al.</i> , 1990), synthesized by PCR using oligonucleotides CC1 and CC2	pJP44.365	Sequence -235 to -226 replaced by CTCGAGTGAC (<i>XhoI</i> site)
pJP25.5	Sequence -175 to -170 replaced by AGGCCT (<i>StuI</i> site)	pJP44.375	Sequence -225 to -216 replaced by GCCTCGAGGT (<i>XhoI</i> site)
pJP25.32	Sequence -185 to -176 replaced by GTCTAGAGAG (<i>XbaI</i> site)	pJP44.385	Sequence -215 to -206 replaced by GGCAGGCCTG (<i>StuI</i> site)
pJP25.33	Sequence -175 to -166 replaced by ATGTCGACGT (<i>SalI</i> site)	pJP44.575	Sequence -85 to -76 replaced by CAGTAGGCTT
pJP25.34	Sequence -165 to -156 replaced by TCGAGGCCTA (<i>StuI</i> site)	pJP45	Deletion of the 30 bp containing the VSG homology (nt -234 to -264) from pJP25. Made by PCR using oligonucleotide CC25. Resulting plasmid retains the <i>NsiI</i> site
pJP25.35	Sequence -155 to -146 replaced by TCAGAGGCCT (<i>StuI</i> site)	pJP46	Deletion of the 40 nt containing the inverted repeats from pJP25 (nt -154 to -194). Made by PCR using oligonucleotide CC26 and one further downstream; fragment cloned into <i>SpeI</i> - <i>HindIII</i> cut pJP25; <i>SpeI</i> site is retained
pJP25.36	Sequence -145 to -136 replaced by AGGAGGCCTG (<i>StuI</i> site)	pJP49	40 nt of random sequence (oligos DS108 and DS109) cloned into the <i>SpeI</i> site of pJP46
pJP26	Cloning of the aldolase 3'-untranslated region and polyadenylation site into pJF6 between <i>BamHI</i> and <i>PstI</i> sites. Contains the region between positions 3203 and 3909 of the aldolase locus (Vijayasathay <i>et al.</i> , 1990) made by PCR using oligonucleotides CC3 and CC4	pJP50	Deletion of promoter from -154 to -55 in pJP44. PCR using oligonucleotides DS110 and DS111; fragment cloned into <i>KpnI</i> - <i>SmaI</i> cut pJP44
pJP27	<i>KpnI</i> - <i>SpeI</i> promoter deletion of pJP25; contains proximal promoter region from +34 to -196	pJP51	<i>KpnI</i> (-330)- <i>NsiI</i> (-266) deletion of pJP44
pJP28	Insertion of <i>parp</i> 3'-end (from pJP25) into pJP14. Because pJP14 has no mini-exon terminator, this clone is also without it	pJP52	<i>KpnI</i> (-330)- <i>SpeI</i> (-196) deletion of pJP44
pJP32	Deletion from <i>NsiI</i> (-269) to <i>SpeI</i> (-196) in the promoter of pJP25	pJP70	The complementary clone to pJP71—the upstream promoter fragment from pJP25.34 cloned into the downstream sequences from pJP25.35. The result has 11 bp deleted and nine additional base pairs mutated
pJP33	<i>parp</i> 3'-end added to pJP21	pJP71	Upstream promoter fragment from pJP25.35 cut with <i>StuI</i> and <i>KpnI</i> , clones into 'vector' fragment made by cutting pJP25.34 with the same enzymes. The resulting plasmid is pJP25 with an 11 nt insertion at position -155 relative to the splice acceptor
bJP43	<i>KpnI</i> - <i>SmaI</i> PCR fragment, containing the promoter for pJP44 (nt -330 to -59) clones into pBSSKII+. Made by PCR using oligonucleotides CC11 and CC23	pJP72	<i>SalI</i> site of pJP25.33 cut and filled in with Klenow. The result is like pJP25.33 but with a 4 nt insertion. Relevant control is pJP 25.33 (-160 replacement)
pJP44	<i>parp</i> promoter remade by PCR to place a <i>SmaI</i> site just upstream of the <i>Sau3AI</i> site. Made by cloning this fragment (see bJP43) between the <i>KpnI</i> and <i>SmaI</i> sites in the polylinker upstream of the <i>BamHI</i> - <i>Sau3AI</i> fusion in pJP28. Sequencing shows that the <i>Sau3AI</i> - <i>BamHI</i> site has been destroyed	pJP80	<i>SpeI</i> site of pJP44 filled in with Klenow. Result has 4 nt duplication of nucleotides -193 to -196 inclusive
pJP44.30	Sequence -205 to -196 replaced by CAGGCCTGCT (<i>StuI</i> site)	pJP81	<i>NsiI</i> site of pJP44 blunted with T4 polymerase. Result has a 4 nt deletion of positions -266 to -269 inclusive
pJP44.31	Sequence -195 to -186 replaced by GCAAGGCCTA (<i>StuI</i> site)	pJP82	pJP44 cut with <i>NsiI</i> and <i>SpeI</i> . Sites filled with T4 polymerase (i.e. exonuclease on <i>NsiI</i> site, deleting 4 nt, and polymerase on <i>SpeI</i> site, adding 4 nt). 73 bp <i>NsiI</i> - <i>SpeI</i> fragment (also blunt-ended) replaced in correct orientation. 4 nt duplication of nucleotides -193 to -196 inclusive and 4 nt deletion of positions -266 to -269 inclusive. Control for pJP83
pJP44.37	Sequence -135 to -126 replaced by AAGGCCTTAG (<i>StuI</i> site)	pJP83	like pJP82 except 73 bp fragment replaced in inverse orientation
pJP44.38	Sequence -125 to -116 replaced by AGAGGCCTAG (<i>StuI</i> site)	pJP84	Synthetic 30mer placed into the <i>NsiI</i> site of pJP45. This 30mer (paired oligonucleotides CZ01 and CZ02) has the VSG homology mutated but all other nucleotides are the same as wild type
pJP44.39	Sequence -115 to -106 replaced by AACTCGAGCG (<i>XhoI</i> site)		
pJP44.40	Sequence -105 to -96 replaced by GACTCGAGAA (<i>XhoI</i> site)		

expression to ~10% of parental levels, indicating that although the sequence of this region is important, it also has some spacer function.

Another interesting motif occurs further upstream. The sequence at position -258 to -240 has strong homology (16 out of 19 nucleotides, including one gap) with a region upstream of the VSG 221 promoter (Figure 3). This sequence is not required for maximal VSG promoter activity (Zomer-dijk *et al.*, 1990). Within the *parp* promoter, however, dele-

tion of either base pairs -270 to -196 (pJP32) or just 30 bp containing the homology (-264 to -234 ; pJP45) reduced CAT expression by 80%. Further, mutating the entire VSG homology (pJP84) resulted in the same level of CAT expression as deleting it (pJP45).

Changes in the spacing between different elements

The deletion and mutation results indicate that the region from -330 to -191 has a stimulatory role and that an element or elements between -191 and -55 are essential for promoter function. Minor changes in the spacing between these different promoter elements had clear effects on the level of CAT expression. Changing the spacing between the VSG homology and the essential region by inserting 4 bp (slightly less than half a helix turn; pJP80) reduced CAT activity; so did a 4 bp deletion upstream of the VSG homology (pJP81), and the effects may be additive (pJP82; see Figure 4). Inverting the orientation of the -274 to -196 segment (pJP83) had only a minor effect—if anything less than the relevant control (pJP82).

Alterations further downstream had much more dramatic consequences. Three constructs were made using clones containing 10 bp mutations in the promoter (see below). Starting with mutant construct pJP25.33, in which bp -175 to -166 are mutated and which produces about half of parental activity, we inserted 4 bp (pJP72). Promoter activity was abolished. We then tried inserting 11 bp upstream of position -155 (pJP71); this again abolished CAT activity, although replacement of bp -165 to -156 with the same sequence (pJP25.34; see below) had a much less dramatic effect. The third construct, a more extensive replacement/deletion in this region (pJP70) has no activity, but this is to be expected as it deletes a region at -150 that is essential for promoter function (see below).

10 bp replacement analysis

Results of the foregoing experiments all pointed towards the presence of at least three elements whose relative spacing and sequence were required for maximal activity. To examine all these sequences in more detail we made a series of mutants (Table I) in which sequential 10 bp segments were completely replaced with dissimilar sequence. The effects delineated by this replacement series display a periodicity of ~ 33 bp and grow progressively weaker with increasing distance from the putative initiation site (Figure 5). Mutants in two regions had very powerful effects on promoter activity. The first extends from -125 to the putative transcription start-site at -82 . In the mutant of the region at -120 , CAT expression is indistinguishable from background and RNA is undetectable (data not shown). Mutation of the second important region, at -150 , reduces CAT activity to $\sim 5\%$ of parental levels. These results were confirmed after recloning upstream of the luciferase gene. Effects of the other replacements were less severe. In three separate areas, -195 to -176 , -225 to -206 and -255 to -246 , mutations resulted in $\sim 50\%$ loss of promoter activity.

Discussion

RNA structure and transcription initiation

The CAT RNA produced after transfection is polyadenylated and contains a mini-exon at the physiologically relevant position. Similar results were obtained by Lee and Van der

Fig. 5. Effect of 10 bp mutations on promoter activity. CAT activities (Rel. CAT Act.) are expressed relative to the activity from the relevant control clone (either pJP25 or pJP44). The centre of each 10 bp replacement is shown below the corresponding bar; for example, -80 means mutant from positions -85 to -76 relative to the splice acceptor site. The map beneath the graph indicates the VSG promoter homology (bar, V), the four repeats (filled arrows), the TATA or rRNA promoter homology (ATA) and the putative transcription start (open arrow).

Ploeg (1990) who observed α -amanitin-resistant transcription and accurate *trans*-splicing with a *parp*-CAT construct integrated into the tubulin locus. In addition, we have tentatively mapped the 5' end of the CAT primary transcript to a position approximately 82 bp upstream from the mini-exon addition site. Several arguments support this conclusion. Nuclear run-on and UV inactivation analyses have repeatedly shown that *parp* transcription initiates within about 200 bp 5' to the mini-exon splice site (Clayton *et al.*, 1990; Pays *et al.*, 1990; Rudenko *et al.*, 1990). J. Huang and L.H.T. Van der Ploeg (Columbia University, personal communication) have also obtained this primer extension band from RNA from transfected cells and Pays *et al.* (1990) observed an RNA with a similar 5'-terminus in trypanosomes treated with low levels of UV irradiation. [In such trypanosomes, RNA processing and degradation appear to be abnormal (Coquelet *et al.*, 1991).] The faint band in our primer extensions (Figure 3) is unlikely to reflect use of a secondary mini-exon addition site since that would require a splice acceptor AG dinucleotide at position -47 , where none exists. (The nearest is 10 bases from the appropriate position.) However, we have not excluded the possibility that we have located the *parp* mini-exon splicing branch site or some other processing intermediate.

Analysis of the *parp* promoter

The nature of the polymerase that transcribes the trypanosome major surface protein genes is disputed. It could be a modified form of RNA polymerase I: transcription of protein-coding genes by RNA polymerase I should be possible as the 5' cap can be provided by *trans*-splicing. Alternatively it could be a modified RNA polymerase II. It seems unlikely to be a fourth type of major polymerase subunit: attempts to purify an enzyme with the requisite properties have so far failed (Grondal *et al.*, 1989). Although there are two polymerase II genes in *T. brucei*, the three to four amino acid differences between them are neither in the region thought to be important for α -amanitin sensitivity, nor conserved between trypanosome isolates (Smith *et al.*, 1989; Evers *et al.*, 1989). Eventually promoter comparisons may

Table II. Oligonucleotide sequences

CC1	GCACTGCAGAGCGCCACATTACACGCAGC
CC2	CGGGATCCGCGGATGCAAGCGTGTAAGCGCCTC
CC3	CACTGCAGGGGAAGGAAACCAACTTGTAGTGGC
CC4	CGGGATCCTGCCCATTTAGTTGGCTTTTCCCTTG
CC11	TCGGTACCGTCATTGGGGTTAAGCGGAAAG
CC23	TCCCGGGACGAGAACATGTTAACTCAGTAC
CC25	GCTGGTACCATGCATCGTGTCTGATAGGTATCTCT
CC26	GCTGGTACCACTAGTGATGGGGTTGTTTTATGCTA
CZ01	TCATGAGTGGACTAGCAGGCCTCTCATGCA
CZ02	TGAGAGGCCTGCTAGTCCACTCATGATGCA
CZ03	TAGTAAGCTTAATGTCCTTTTCACAGTGCA
CZ04	CTGTGAAAAGGACATTAAGCTTACTATGCA
CZ012	TTTTGCTATTCCGTGTCTCTGGGTGGGCGTGCAT
CZ013	GGGGATGCACGCCCACCCAGAGACACGGAATAGC
CZ014	TTTTTCAGTCGTGTGTTGCTGCTTGTAGCGCGTG
CZ015	GGGGCACGCGCTAGCAAGCAGCAACACACGACTG
DS108	CTAGTATGAGAACTCAAGGCCTTCTCGCCAGTCGGAACCA
DS109	CTAGCTGGTTCCGACTGGCGAGAAAGGCCTTGAGTTCTCATA
DS110	ATTAGGTACCATGCATAGTGAGCTTAATGT
DS111	GGGAAGCTTGCCACAAAATGGACATTTTG
VIB13	CGATGCCATTGGGATATATCAACG

suggest an answer to this problem, although the dissimilarity between VSG and *parp* promoters is discouraging.

The *parp* promoter is contained within ~250 bp upstream of the putative initiation site (Clayton *et al.*, 1990; Rudenko *et al.*, 1990; this paper). Elements between -196 and -55 (relative to the *parp* mini-exon splice site) are necessary and sufficient for basal level transcription while sequences from -196 to -330 play a significant stimulatory role.

A series of 10 bp replacement mutations identified two regions, CAGTGATGGG centered at position -150 and GTCTCTGGGT at position -120 (68 and 38 bp respectively from the putative transcription initiation site) that were vital for CAT expression. Mutating each of these regions reduced activity by 95% or more. These sequences resemble neither the trypanosome rRNA promoter (White *et al.*, 1986) nor known polymerase II promoter elements from other eukaryotes (Struhl, 1989; Wingender, 1990; Locker and Buzard, 1990; Mitchell and Tjian, 1989). Each of these *parp* promoter elements overlaps at its 3' end a stretch of G and T residues (from -149 to -138 and -120 to -113). The *parp* B locus differs from the A gene at positions -148 and -147 interrupting the first G/T box (see Figure 3). However, compensatory mutations just downstream create another, overlapping box (-147 to -135). There are G/T-rich stretches upstream of the VSG 221 transcription start-site, but not in the region essential for transcription (Zomerdijsk *et al.*, 1990). Mutating the downstream 5 bases (TGGGT) of the element at -120 results in <5% of wild-type activity (data not shown).

Replacements of more distal sequences had less severe effects. All the effects delineated by the replacement series occur at regularly spaced intervals of 30-35 bp. The important elements of the *parp* promoter may therefore lie on the same face of the DNA, offering ample opportunity for protein-protein interactions. Preliminary gel-shift results indicate that a nuclear factor binds specifically to the -120 region (data not shown).

Ribosomal RNA promoters show almost no sequence conservation between species, although there is limited conservation of overall structure (Sollner-Webb and Mougey, 1991). There is therefore very little point in detailed

sequence comparisons between the *parp* and VSG promoter sequences and heterologous polymerase I promoters. The *T.brucei* rRNA promoter has been mapped (White *et al.*, 1986) but detailed mutational analysis is not yet available. In higher eukaryotes, the sequences that are most important for rRNA transcription usually lie within the 35 bp upstream from the initiation site, with an additional stimulatory region extending up to around -150; spacing between elements is important. Deletion and mutation analysis of the evolutionarily closer *Acanthamoeba castellanii* rRNA promoter defined the sequence from -31 to +8 as essential for transcription, the region around -29 being most strongly implicated (Kownin *et al.*, 1985, 1988).

No trypanosome promoter for α -amanitin-sensitive transcription has yet been defined. RNA polymerase II promoters generally contain three distinct elements: a start-site selector (extending from about -45 to +30 relative to the transcription initiation site), an upstream region (from about -120 to -45) and enhancer elements, whose influence is independent of orientation or position for as much as 10 kb up or downstream (Wasylyk, 1988; Lewin, 1990). The single most conserved feature of this model is the transcription factor IID binding site, the TATA box, located within the start-site selector of most eukaryotic genes. The *parp* promoter has an A-rich sequence (AAAATAG) with weak homology to a TATA box at -103 to -97, ~18 bp upstream of the putative transcription initiation site. Our 10 bp replacement in this region (-100) inhibited CAT expression by >80%. Most probably, AAAATA would not function as a TATA box in either mammalian or yeast cells or extracts (Wobbe and Struhl, 1990). A similar sequence, AAAAGTAG, is present 13 nt upstream of the trypanosome rRNA transcription start (Figure 3) (White *et al.*, 1986).

There are many examples of TATA-independent polymerase II promoters (Smale and Baltimore, 1989). However, the nucleotide composition and sequence motifs of the *parp* promoter do not resemble them (Anderson *et al.*, 1988; Biggin and Tjian, 1988; Garvin *et al.*, 1988; Perkins *et al.*, 1988; Soeller *et al.*, 1988). Examples include many housekeeping genes, which have GC-rich promoters with several potential binding sites for the transcription factor SP1

(Dyan and Tjian, 1983; Sehgal *et al.*, 1988). Other TATA-independent promoters are often tightly regulated during differentiation or development (Smale and Baltimore, 1989), as are the *parp* genes (Mowatt and Clayton, 1987).

Our results raise as many questions as they answer. Several constructs that alter spacing all suggest that the -191 to -155 region exerts a crucial spacer function. However, we were unable to identify a truly vital element upstream of -191. (We cannot rule out the possibility that some of the promoter deletions are in fact altering interactions with unidentified elements in the plasmid sequence.) Likewise, deletion or mutation of the VSG homology region (-258 to -240) was more deleterious than would be predicted from results of the 10 bp replacement series. We were unable to ascribe functions to some intriguing sequence elements: individual mutation of the direct and inverted repeats (-183 to -137) and the serum response element predicted by Dorn *et al.* (Dorn *et al.*, 1991) had little effect on promoter activity; possibly mutation of more than one of the individual repeats would be more deleterious.

Until other trypanosome promoters are analyzed in similar detail, we will be unable to say which unusual aspects of this promoter are particular to the *parp* genes and which reveal more general aspects of trypanosome transcription. Analyses of the proteins interacting with the promoter and accurate homologous *in vitro* transcription systems, will be required to answer the question of which RNA polymerase transcribes the *parp* and VSG genes.

Materials and methods

Trypanosomes

Procyclic *T. brucei* of the 427 strain were cultured as described previously (Mowatt and Clayton, 1987).

DNA cloning and characterization

Constructs were made and sequenced by standard protocols as previously described (Clayton *et al.*, 1990). Details of all clones are listed in Table I. All clones are derivatives of expression plasmid pJF6 (Clayton *et al.*, 1990). The plasmid contains a 623 bp promoter and *trans*-splicing acceptor fragment from the *parp* A α gene between the *Kpn*I site of pGEM4 (i.e. next to the T7 promoter of the vector) and the CAT gene. Downstream of the CAT gene, a fragment of the mini-exon transcription terminator region from *L. seymouri* was erroneously included. This fragment had no effect on CAT expression but did influence RNA structure (see Results). Plasmid pJP25 has the 3'-untranslated region and polyadenylation site of the *parp* B α gene inserted downstream of the CAT gene of pJF6.

Site-directed mutagenesis was performed using the *dut/ung* method of Kunkel *et al.* (1987) or by polymerase chain reaction (Nelson and Long, 1989) with a PC1 Programmable Dri-Block (Techne, Princeton, NJ) or Perkin-Elmer Cetus thermal cycler. Oligonucleotides (Table II) were from the Rockefeller University or ZMBH oligonucleotide synthesis facilities or from Oligos Etc. (Guilford, CT).

Trypanosome transfections

Initially, transfections were performed as previously described (Clayton *et al.*, 1990). Recently we have used a BTX ECM 600 electroporator (BTX, San Diego, CA) set at 1.5 kV and 24 Ω (R2 setting) to deliver a single pulse. Both protocols yielded comparable results. We generally transfected with 5 or 10 μ g plasmid DNA per cuvette for subsequent CAT enzyme assays, or with 50 μ g per cuvette for subsequent RNA analysis. CAT assays were performed as previously described (Clayton *et al.*, 1990) or by the two-phase diffusion assay (Neumann *et al.*, 1987) using [¹⁴C]butyryl CoA (Dupont-NEN) always in the linear range of the assay. Each clone was assayed at least three times and all assays were performed in duplicate or triplicate.

RNA analysis

RNA was isolated from trypanosomes either by ultracentrifugation through a CsCl cushion (Glisin *et al.*, 1974; Chirgwin *et al.*, 1979) or by acid guanidinium-phenol-chloroform extraction (Chomczynski and Sacchi,

1987). After transfection, RNA isolated by either method required extensive DNase treatment (RQ1 DNase, Promega, Madison, WI) to remove co-purifying plasmid DNA. Poly(A)⁺ RNA was isolated using Hybond-mAP (Amersham, Arlington Heights, IL) according to the manufacturer's instructions. Northern blot analyses employed standard methodologies. Control pre-treatments (on 10 μ g RNA or 1 ng DNA) were: 5 U RQ1 DNase, 200 μ g/ml RNase A or 0.2 M NaOH, each for 30 min at 37°C. Primer extension was performed as described by Geliebter (Geliebter, 1986), using 25 μ g total RNA. RT-PCR, either from whole cells or purified RNA, was performed as described previously (Sherman *et al.*, 1989).

Acknowledgements

We wish to thank Indu Tomar, Helen Murray, Tomoko Oeda, Kelly Fung and Claudia Hartmann for technical assistance; Janet Lee (Memorial Sloan Kettering Cancer Center), Martin Eilers, Achim Leutz and Thomas Wirth (ZMBH) and Jan Geliebter and all members of the George Cross laboratory (The Rockefeller University), for useful discussions; also John Boothroyd and Patricia Dorn (Stanford University), Elizabeth Wirtz (Meharry Medical School), Piet Borst, Annelore ten Asbroek and Joost Zomerdijs (Netherlands Cancer Institute) and Lex Van der Ploeg (Columbia University) for communicating unpublished results. We thank George Cross, in whose laboratory much of this work was completed, for his encouragement. This work was in part supported by NIH grant AI 22229 to C.C., now transferred to G.A.M. Cross.

References

- Agabian, N. (1990) *Cell*, **61**, 1157-1160.
- Anderson, S.J., Chou, H.S. and Loh, D.Y. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 3551-3554.
- Bellofatto, V. and Cross, G.A.M. (1989) *Science*, **244**, 1167-1169.
- Biggin, M.D. and Tjian, R. (1988) *Cell*, **53**, 699-711.
- Bucher, P. and Trifonov, E.N. (1986) *Nucleic Acids Res.*, **14**, 10009-10026.
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry*, **18**, 5294-5299.
- Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.*, **162**, 156-159.
- 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.
- Coquelet, H., Steinert, M. and Pays, E. (1991) *Mol. Biochem. Parasitol.*, **44**, 33-42.
- Donelson, J.E. (1989) In Berg and Howe (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 763-781.
- Dorn, P.L., Aman, R.A. and Boothroyd, J.C. (1991) *Mol. Biochem. Parasitol.*, **44**, 133-140.
- Dyan, W.S. and Tjian, R. (1983) *Cell*, **35**, 79-87.
- Evers, R., Hammer, A., Köck, J., Jess, W., Borst, P., Mémet, S. and Cornelissen, A.W.C.A. (1989) *Cell*, **56**, 585-597.
- Garvin, A.M., Pawar, S., Marth, J.D. and Perlmutter, R.M. (1988) *Mol. Cell. Biol.*, **8**, 3058-3064.
- Geliebter, J. (1986) *BRL Focus*, **9**, 5-8.
- Glisin, V., Crkvenjakov, R. and Byus, C. (1974) *Biochemistry*, **13**, 2633-2637.
- Gottesdiener, K., Chung, H.-M., Brown, S.D., Lee, M.G.-S. and Van der Ploeg, L.H.T. (1991) *Mol. Cell. Biol.*, **11**, 2467-2480.
- Gronlund, E.J.M., Evers, R., Kosubek, K. and Cornelissen, A.W.C.A. (1989) *EMBO J.*, **8**, 3383-3389.
- Jefferies, D., Tebabi, P. and Pays, E. (1991) *Mol. Cell. Biol.*, **11**, 338-343.
- König, E., Delius, H., Carrington, M., Williams, R.O. and Roditi, I. (1989) *Nucleic Acids Res.*, **17**, 8727-8739.
- Kownin, P., Iida, C.T., Brown-Shimer, S. and Paule, M.R. (1985) *Nucleic Acids Res.*, **13**, 6237-6248.
- Kownin, P., Bateman, E. and Paule, M.R. (1988) *Mol. Cell. Biol.*, **8**, 747-753.
- Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) *Methods Enzymol.*, **154**, 367-382.
- Laban, A. and Wirth, D.W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9119-9123.
- Lee, M.G.-S. and Van der Ploeg, L.H.T. (1990) *Science*, **250**, 1583-1587.
- Lewin, B. (1990) *Cell*, **61**, 1161-1164.
- Locker, J. and Buzard, G. (1990) *DNA Sequence*, **1**, 3-11.
- Mitchell, P.J. and Tjian, R. (1989) *Science*, **245**, 371-378.
- Mowatt, M.R. and Clayton, C.E. (1987) *Mol. Cell. Biol.*, **7**, 2838-2844.
- Nelson, R.M. and Long, G.L. (1989) *Anal. Biochem.*, **180**, 147-151.
- Neumann, J.R., Morency, C.A. and Russian, K. (1987) *Biotechniques*, **5**, 444-446.

- Pays,E., Coquelet,H., Tebabi,P., Pays,A., Jefferies,D., Steinert,M., Koenig,E., Williams,R.O. and Roditi,I. (1990) *EMBO J.*, **9**, 3145–3151.
- Perkins,K.K., Dailey,G.M. and Tjian,R. (1988) *Genes Dev.*, **2**, 1615–1626.
- Rosenberg,M. and Court,D. (1979) *Annu. Rev. Genet.*, **13**, 319–353.
- Rudenko,G., Bishop,D., Gottesdiener,K. and Van der Ploeg,L.H.T. (1989) *EMBO J.*, **8**, 4259–4263.
- Rudenko,G., Le Blanq,S., Smith,J., Lee,M.G.S., Rattray,A. and Van der Ploeg,L.H.T. (1990) *Mol. Cell. Biol.*, **10**, 3492–3504.
- Sehgal,A., Patil,N. and Chao,M. (1988) *Mol. Cell. Biol.*, **8**, 3160–3167.
- Sherman,D.R., Geliebter,J. and Cross,G.A.M.C. (1989) *Trends Genet.*, **5**, 137.
- Smale,S.T. and Baltimore,D. (1989) *Cell*, **57**, 103–113.
- Smith,J.L., Levin,J.R., Ingles,C.J. and Agabian,N. (1989) *Cell*, **56**, 815–827.
- Soeller,W.C., Poole,S.J. and Kornberg,T. (1988) *Genes Dev.*, **2**, 68–81.
- Sogin,M.L., Elwood,H.J. and Gunderson,J.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1383–1387.
- Sogin,M.L., Gunderson,J.H., Elwood,H.J., Alonso,R.A. and Peattie,D.A. (1989) *Science*, **243**, 75–77.
- Sollner-Webb,B. and Mougey,E.B. (1991) *Trends Biochem. Sci.*, **16**, 58–62.
- Stuart,K. (1991) *Trends Biochem. Sci.*, **16**, 68–72.
- Struhl,K. (1989) *Annu. Rev. Biochem.*, **38**, 1051–1077.
- Vijayasarathy,S., Ernest,I., Itzhaki,J., Sherman,D., Mowatt,M.R., Michels,P.A.M. and Clayton,C.E. (1990) *Nucleic Acids Res.*, **18**, 2967–2975.
- Wasylyk,B. (1988) *CRC Crit. Rev. Biochem.*, **23**, 77–120.
- White,T.C., Rudenko,G. and Borst,P. (1986) *Nucleic Acids Res.*, **14**, 9471–9489.
- Wingender,E. (1990) *CRC Crit. Rev. Eukaryotic Gene Exp.*, **1**, 11–48.
- Wobbe,C.R. and Struhl,K. (1990) *Mol. Cell. Biol.*, **10**, 3859–3867.
- Zomerdijk,J.C.B.M., Ouellette,M., ten Asbroek,A.L.M.A., Kieft,R., Bommer,A., Clayton,C.E. and Borst,P. (1990) *EMBO J.*, **9**, 2791–2801.

Received on July 8, 1991; revised on August 2, 1991