Published in final edited form as:

Mol Biochem Parasitol. 2007 February ; 151(2): 224–228.

A 10 base-pair sequence within Domain III of the *GPEET* procyclin promoter is essential for the autonomous replication of a plasmid in procyclic *Trypanosoma brucei*

Amin Espinal^{*}, Jedidiah Quijano^{*}, Charleen Hunt, Robert Lorenzo, Christopher Mulligan, Mario Sampson, Mark Sauchelli, and Pradeep K. Patnaik^{**}

Department of Biology, William Paterson University, 300 Pompton Road, Wayne NJ 07470

Very little is known about sequences that mediate nuclear DNA replication in *Trypanosoma brucei*. We have previously reported the construction and characterization of a panel of autonomously replicating plasmids and artificial linear mini-chromosomes for this organism [1,2]. The episome was obtained by ligating random DNA fragments from the *T. brucei* genome on to a plasmid that was otherwise incapable of stable replication in this organism. We call these inserted genomic fragments plasmid maintenance sequences (PMS). One member of this panel of episomes (pT13-11) and the corresponding artificial linear mini-chromosomes have been extensively characterized [1-4]. This plasmid, and by implication its many derivatives (such as pEV-luc; please see below), as well as the linear mini-chromosomes exist at a copy number of 1-2 per cell in procyclic *T. brucei* and demonstrate substantial mitotic stability in the absence of selection. We are using these autonomous replicons as model systems to delineate critical nuclear DNA replication control elements for this organism.

The plasmid pEV-luc (Figure 1) has been described previously [4]. In this episome, 433 base pairs (- 408 to + 25 relative to the start site of transcription) of the *GPEET* procyclin promoter (also called *PARP-A*) drives the RNA polymerase I-mediated expression of firefly luciferase and neomycin phosphotransferase (G418^R). We have previously demonstrated that the autonomous replication of pEV-luc is dependent on the PMS (as expected) and, surprisingly, on the segment bearing the *GPEET*-promoter [4]. The close association between elements controlling transcription and DNA replication and the possible interaction between factors assembling at these sites is intriguing and fuels our long-term interest in this area. We have sought to delineate the critical promoter-proximal replication control element(s) and determine the overlap (if any) between the two sets of regulatory sequences.

The *GPEET* procyclin promoter has been extensively studied [5-7]. It comprises of 4 domains which map to positions -7 to -40 (Domain I), -57 to -72 (Domain II), -90 to -143 (Domain III) and -207 to -222 (Domain IV) with respect to the start point of transcription [5,7; also see Fig 1]. Ten base pair (bp) linker-substitutions within Domains I, II, III and IV reduce *GPEET*-promoter-mediated expression of a reporter gene to approximately <0.1%, 5-10%, 30-50%, and 60-70% of wild-type levels respectively [5,7]. Two other RNA Pol I promoters (*RRNA* and *VSG*) have been characterized in *T. brucei* but demonstrate little sequence conservation [8-11]. Nevertheless, transcription mediated by the *GPEET*, *RRNA* or *VSG* promoters in

^{*}Contributed equally to this study

^{**}Corresponding author Tel: 973 720 3454 Fax: 973 720 2338 Email: patnaikp@wpunj.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

vitro is inhibited by preincubation of the cell extract with fragments corresponding to any of the 3 RNA Pol I promoters. A detailed examination of the *GPEET* promoter highlighted the critical role of Domain III in these competition experiments and the need of Domain II for full activity. On the other hand, Domain I, whose mutation has such a profound effect on *GPEET*-transcription, was dispensable for this effect [7]. The necessity of Domain III was intriguing, because linker-substitutions within this region only marginally weaken the promoter.

We have re-visited the promoter region of *GPEET* procyclin and analyzed the effect of linker substitutions within each of the 4 domains as well as several inter-domain regions on autonomous plasmid replication. Replication of pEV-Luc and the corresponding sequence-substitution derivatives were determined by a luciferase assay as described previously [4]. Briefly, a replicating plasmid carrying a gene for luciferase shows an exponential increase in luciferase levels when electroporated into *T. brucei* while a non-replicating plasmid does not. As electroporated DNA could also replicate by integrating on to a chromosome or by acquiring critical new sequences from the genome through recombination, we sought to confirm stable autonomous replication through the presence of the input plasmid in transformed cells after a period of prolonged culture (a cumulative dilution of $>10^8$ following electroporation is necessary to ensure removal of passively persisting non-replicating input DNA). To do this, DNA isolated from these long-term cultures was introduced into bacteria. Surviving episomes can transform *Escherichia coli* to ampicillin-resistance (Amp^R) while integrated constructs cannot. Subsequently, plasmids isolated from these Amp^R bacteria were examined for rearrangement or acquisition of new sequences by restriction analysis and re-sequencing of the promoter region. These methods have been used before [4].

The plasmid p Δ 237 is a derivative of pEV-luc that bears a smaller segment of the *GPEET* promoter (-237 to +25 with respect to the start of transcription). In addition, we have removed the polylinker containing a string of restriction sites immediately upstream of *KpnI*. The promoter segment was amplified from pEV-luc using a polymerase chain reaction (PCR) and appropriate oligonucleotides bearing *ApaI* and *KpnI* sites at their ends. The PCR-amplified segment was digested with *ApaI* and *KpnI* and inserted into a similarly digested pEV-luc. Please note, that in comparison to pEVluc, plasmids p Δ 237, pCJRP (see later) and all subsequent derivatives would not only have a smaller promoter segment but would have also lost ~ 1.2 kbp of the PMS as a result of the *ApaI* digestion. We have previously shown that the PMS segment removed in these construct was dispensable for autonomous replication [3]. The plasmid pCJRP (Figure 1) is similar to p Δ 237 except that it has a slightly longer promoter segment (-250 to +25) as well as a single base pair (T to G) substitution at position -169 that creates an unique *AflIII* site. In addition, it retains the 23bp polylinker that was removed in p Δ 237. The single base-pair T to G alteration was engineered by a method similar to the one described by Stemmer et. al. [12] and used by others [13]. The unique *SpeI*, *AflIII* and *ApaI* restriction-sites on pCJRP, which are located at positions -112, -171 and -251 respectively (see Figure 1), allowed us to conveniently introduce sequence-substitutions into the Domain III and Domain IV region of the *GPEET* promoter (see later). All changes have been confirmed by sequencing.

Sequence-substitution derivatives were created in one of two ways. In most cases we simply obtained two long complementary oligonucleotides bearing the desired change and appropriate restriction sites (*ApaI* and *AflIII*; or *AflIII* and *SpeI*) at their ends. These were annealed, digested with *ApaI*-*AflIII* (or *AflIII* and *SpeI*) and ligated to similarly digested pCJRP. Alternatively, where appropriate linker-substitution mutations within the *GPEET* promoter were available from other laboratories (kindly provided by Arthur Günzl and Christine Clayton) we PCR-amplified the promoter fragments from these plasmids and introduced them into *ApaI* and *KpnI* digested p Δ 237. Each of these constructs is named according to the position of the 10 bp (12 bp in one case) sequence-substitution relative to the start site of transcription mediated by

the *GPEET* promoter. All have been sequenced in the region of the promoter and the relevant sequence alterations are documented in the legend to Figure 1.

Procyclic trypanosomes (strain 427) were electroporated with closed circular plasmid DNA isolated from *E. coli*. Experiments were always done in duplicate. G418 was added ~ 18-24 hours post transfection (Day 1) to a final concentration of 15µg/ml. A lower than usual concentration of the drug was used because many of the constructs (especially pLS-32-41, which inactivates Domain I) had impaired promoters. Procyclics transfected with pLS-32-41 could withstand this level of selection. Samples were assayed for luciferase on Days 1, 3, 5, 7, and in some experiments, Days 9 and 11. Each construct has been assayed at least four times (i.e. two separate experiments each in duplicate) and some-for example the wild type plasmids (pEV-luc and pCJRP) as well as all the domain-specific mutants (pLS -32-41, pLS -62-71, pLS -133-142, and pLS-212-221), have been assayed greater than eight times. The result of one such experiment is shown (Table 1) and is similar to all others we have carried out. Longer-term survival of some of these constructs as autonomous replicons was assessed in at least 2 (and in some cases 4 or more) independently transformed *T. brucei* populations.

As expected, parasites transformed with pEV-luc, pCJRP or the deletion construct pΔ237 show an exponential increase in luciferase with time. Furthermore, the input plasmid can be re-isolated from transformed parasites after a prolonged period of culture under selection.

Linker substitution of a sequence within Domain I (pLS-32-41) severely reduces promoter strength (as expected) but luciferase levels increase exponentially and the input plasmid can be re-isolated after prolonged culture like the controls described above. We conclude that Domain I of the *GPEET*-procyclin promoter is not essential for the autonomous replication of these plasmids in procyclic *T. brucei*.

Strikingly, linker substitution of the 10 bp sequence between -133 to -142 (pLS-133-142A), which replaces the wild-type sequence ATAGGTATCT, completely compromises this plasmid's ability to replicate. The promoter is somewhat impaired (as expected - since the mutation maps to one end of Domain III) but luciferase levels remain essentially the same over the 7 day period. We have examined this mutant over longer periods (up to 11 days) and see no increase in luciferase activity. Furthermore, parasites transfected with pLS-133-142A do not yield G418^R cultures and were usually discarded after 3-4 weeks. In one instance G418^R cells did grow out after a prolonged period. However, we failed to find any episomes in this transformed culture (after requisite dilution) and assume that the construct had integrated into the genome. Linker insertions of sequences on either side of this mutation, including those mapping to slightly overlapping segments (pLS -141-150; pLS-126-135) had no significant effect on plasmid replication as measured via transient assays (luciferase) or long-term autonomous survival of these plasmid. Because it was possible that the effect demonstrated by pLS-133-142A was due to some unusual and unrecognized feature of the substituted sequence we recreated this mutant using a completely different sequence (please see Figure 1 legend). This plasmid (pLS -133-142B) demonstrated the same phenotype as pLS-133-142A (data not shown). Consequently, we conclude that the 10 bp delineated by this linker substitution defines a critical replication control element on these plasmids in procyclic trypanosomes. An alternative idea would posit that the behavior of pLS-133-142 emerges from the impaired segregation of a low-copy episome that is under strict copy-number control. However, our data demonstrates this to be untenable. Random segregation would lead to the 'birth' of a subset of transformed cells bearing plasmids below the copy-number ceiling. As these episomes would replicate we would see a slow but gradual increase in luciferase levels. This was not observed.

Another interesting finding is the behavior of the Domain II mutant pLS-62-71. This plasmid not only shows an expected decrease in luciferase when compared to its wild-type counterpart

(pΔ237) but a striking delay in the onset of amplification. Luciferase levels remain essentially constant for the first three days (five days in other experiments) and only begin to increase exponentially thereafter. Once past this delay, the plasmid displays no further impairment (we have examined luciferase levels for 11 days post-electroporation in some experiments) and can be re-isolated from transformed parasites after prolonged culture under G418 selection. We do not know the basis for the initial lag, but the delay cannot be simply related to promoter strength because the Domain I mutant with a far weaker promoter shows no such impairment.

Strikingly, PMS-less derivatives of pEV-luc show an apparently opposite phenotype to the one described for Domain II mutants (Hunt and Patnaik unpublished observations). Luciferase levels increase ~ 4 -fold over the first 3-5 days (presumably corresponding to 2 divisions of the input DNA) after which the rate of increase slows down drastically. We suspect these constructs may be sensitive to some alteration of the input DNA on introduction into *T. brucei* (such as a buildup of chromatin or the acquisition of eukaryotic specific modifications). G418^R parasites eventually did grow out from *T. brucei* electroporated with these PMS-less constructs after a long period (>4 weeks). However, we failed to re-isolate the input episome from transformed cultures and assume that the introduced DNA had ultimately integrated into the genome.

The initial amplification of the PMS-less derivatives suggests that this region is not the primary replication control element on these episomes. Rather, this role appears to be exercised by control sequences localized within Domain III and II on the *GPEET* promoter. The overlap between sequences controlling transcription and DNA replication suggests an intimate interaction between these two processes at this locus. We hypothesize that a replication-mediating factor/complex assembles in a sequence-dependent manner over the 10bp region of Domain III that we have identified in this work. This complex interacts with, and thus can titrate out a factor (or factors) mediating RNA Pol I transcription that assembles over Domain II and beyond. The interaction between the two set of complexes is mutually stabilizing but is not essential for the assembly of either. Thus, a *GPEET* promoter with a sequence alteration between -133 and -142 cannot replicate but is only slightly impaired in transcription. Alternatively, sequence substitution between -62 and -71 affects transcription much more seriously but only temporarily retards the kinetics of assembly of the replication complex which is otherwise unaffected. It is a matter of interest why this retardation only applies at the beginning, since the rate of amplification of the Domain II mutant approaches that of its wild-type parent (pΔ237) thereafter.

We do not currently know if the *GPEET*-promoter proximal element is indeed an origin of DNA replication (*ori*) in *T. brucei*. However, it is of interest to note that the ARS 1 chromosomal *ori* in the yeast *Saccharomyces cerevisiae* consists of an 11 bp essential element (Element A) as well as multiple ancillary ones (B1, B2, B3) that are collectively important for origin function [14]. One of these auxiliary elements (B3) is the binding site for a transcription factor. Other *oris* in *S. cerevisiae* have a similar multipartite structure [15] and the association of transcription factor binding sites with origins of DNA replication is commonly seen among many eukaryotic viruses [16]. However, too few eukaryotic origins have been characterized in any detail to permit broad generalizations.

Finally, despite the interchangeability of the 3 RNA Pol I promoters in certain transcription assays, our earlier experiments suggest that a *RRNA* promoter cannot functionally replace the *GPEET* promoter in supporting stable autonomous replication [4]. Constructs bearing such a replacement integrate or rearrange during long-term culture of the transformed parasites. *RRNA* loci from yeast to humans bear an *ori*, which in all case, maps to the non-transcribed spacer [for eg. see 17,18,19]. Interestingly, a single bp deletion in the *Tetrahymena* *RRNA* promoter, which maps ~750 bp away from the NTS-based *ori*, affects maintenance of the

amplified ribosomal DNA minichromosome in the *Tetrahymena* macronuclei without detectably affecting transcription [20]. Such long-range interactions may be germane to an understanding of the combined role of *GPEET* promoter and the PMS in supporting stable autonomous replication.

Acknowledgement

This work was supported by an R15 award (A1052133) from the NIH. A.E., C.H and Mario Sampson benefited from various programs (CFR, SURP, PASS) supporting undergraduate research at WPU. A.E. and Mario Sampson were also supported by a Minority Access to an Academic Career (MAC) award from the NJ Commission of Higher Education as well as a MARP summer research scholarship from NIH. We are grateful to Arthur Günzl and Christine Clayton for providing us with plasmids carrying certain mutations used in this study, to Tim Linteau for help with the program Adobe Illustrator and to George Cross and Vivian Bellofatto for critical comments on the manuscript and for allowing us the use of their laboratories where some of these experiments were carried out.

References

- [1]. Patnaik PK, Kulkarni SK, Cross GAM. Autonomously replicating single-copy episomes in *Trypanosoma brucei* show unusual stability. *EMBO J* 1993;12:2529–38. [PubMed: 8508776]
- [2]. Patnaik PK, Axelrod N, Van der Ploeg LH, Cross GAM. Artificial linear mini-chromosomes for *Trypanosoma brucei*. *Nucleic Acids Res* 1996;24:668–75. [PubMed: 8604308]
- [3]. Patnaik PK, Bellofatto V, Hartree D, Cross GAM. An episome of *Trypanosoma brucei* can exist as an extrachromosomal element in a broad range of trypanosomatids but shows different requirements for stable replication. *Mol. Biochem. Parasitol* 1994;66:153–6. [PubMed: 7984179]
- [4]. Patnaik PK, Fang X, Cross GAM. A region encompassing the procyclic acidic repetitive protein (PARP) gene promoter is essential for autonomous plasmid replication in procyclic *Trypanosoma brucei*. *Nucleic Acids Res* 1994;22:4111–8. [PubMed: 7937135]
- [5]. Sherman DR, Janz L, Hug M, Clayton C. Anatomy of the PARP gene promoter of *Trypanosoma brucei*. *EMBO J* 1991;10:3379–86. [PubMed: 1840521]
- [6]. Laufer G, Schaaf G, Bollgong S, Günzl A. In vitro analysis of alpha-amanitin-resistant transcription from the rRNA, procyclic acidic repetitive protein, and variant surface glycoprotein gene promoters in *Trypanosoma brucei*. *Mol. Cell. Biol* 1999;19:5466–73. [PubMed: 10409736]
- [7]. Laufer G, Günzl A. In-vitro competition analysis of procyclin gene and variant surface glycoprotein gene expression site transcription in *Trypanosoma brucei*. *Mol. Biochem. Parasitol* 2001;113:55–65. [PubMed: 11254954]
- [8]. Zomerdijs JCBM, Kieft R, Shiels PG, Borst P. Alpha-amanitin-resistant transcription units in trypanosomes: A comparison of promoter sequences for a VSG gene expression site and for the ribosomal RNA genes. *Nucleic Acids Res* 1991;19:5153–8. [PubMed: 1923801]
- [9]. Janz L, Clayton C. The PARP and rRNA promoters of *Trypanosoma brucei* are composed of dissimilar sequence elements that are functionally interchangeable. *Mol. Cell. Biol* 1994;14:5804–11. [PubMed: 8065315]
- [10]. Pham VP, Qi CC, Gottesdiener KM. A detailed mutational analysis of the VSG gene expression site promoter. *Mol. Biochem. Parasitol* 1996;75:241–54. [PubMed: 8992322]
- [11]. Vanhamme, L, Pays, A, Tebabi, P, Alexandre, S, Pays, E. Specific binding of proteins to the non-coding strand of a crucial element of the variant surface glycoprotein, procyclin, and ribosomal promoters of *Trypanosoma brucei*. *Mol. Cell. Biol* 1995;15:5598–606. [PubMed: 7565711]
- [12]. Stemmer WP, Cramer A, Ha KD, Brennan TM, Heyneker HL. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxynucleotides. *Gene* 1995;164:49–53. [PubMed: 7590320]
- [13]. Withers-Martinez C, Carpenter EP, Hackett F, Ely B, Sajid M, Grainger M, Blackman MJ. PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome. *Protein Eng* 1999;12:1113–20. [PubMed: 10611405]
- [14]. Marahrens Y, Stillman B. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* 1992;255:817–23. [PubMed: 1536007]
- [15]. Rao H, Marahrens Y, Stillman B. Functional conservation of multiple elements in yeast chromosomal replicators. *Mol. Cell. Biol* 1994;14:7643–51. [PubMed: 7935478]

- [16]. Murakami Y, Ito Y. Transcription factors in DNA replication. *Front Biosci* 1999;4:D824–33. [PubMed: 10577391]
- [17]. Linskens MH, Huberman JA. Organization of replication of ribosomal DNA in *Saccharomyces cerevisiae*. *Mol. Cell. Biol* 1988;8:4927–35. [PubMed: 3062373]
- [18]. Hernandez P, Martin-Parras L, Martinez-Robles ML, Schvartzman JB. Conserved features in the mode of replication of eukaryotic ribosomal RNA genes. *EMBO J* 1993;12:1475–85. [PubMed: 8467799]
- [19]. Coffman FD, He M, Diaz ML, Cohen S. Multiple initiation sites within the human ribosomal RNA genes. *Cell Cycle* 2006;5:1223–33. [PubMed: 16721058]
- [20]. Gallagher RC, Blackburn EH. A promoter region mutation affecting replication of the *Tetrahymena* ribosomal DNA minichromosome. *Mol. Cell. Bio* 1998;18:3021–33. [PubMed: 9566921]

Figure 1.

Top: Schematic diagram showing the plasmids pEV-luc and its derivative pCJRP (please see description in the text). **Middle:** A representation of the GPEET promoter in pCJRP showing the position of various promoter domains. The figure is adapted from Laufer and Günzl (7). **Bottom:** Schematic representation of wild-type (wt) and sequence substitution mutants used in this study. Dark boxes represent the 10bp substituted region and their position has been roughly aligned with the promoter domains drawn above. Deletion and linker substitution mutants used in this study were constructed using pEV-luc, pΔ237 or pCJRP and differ in the sequence of their corresponding GPEET-procyclicin promoter as described below and in the text. As many of the mutants showed the additional loss of a T at position - 108 (perhaps due

to an aberrant ligation event at the *SpeI* site utilized in their construction), we often used a variant form of pCJRP (pCJRP*) that had an identical alteration as a control. This variant had been isolated during the initial screen for pCJRP. Both pCJRP and pCJRP* behaved similarly in our assays and have been used interchangeably. **Summary of sequence changes:**

Substitution mutants are named according to the position of the 10bp alteration relative to the start site of GPEET transcription (6). Constructs pLS-32-41 and pLS-62-71 are derived from pJP44.39 and pJP25.26 respectively (5). In these, the sequence AGAGGCCTAG replaces the corresponding wt sequence in pΔ237. In pLS-114-125 the 12bp sequence AAGGATCCAGCA replaces the corresponding wt sequence in pCJRP. In pLS-126-135 and pLS-141-150, GAGGATCCAG replaces the corresponding wt sequence in pCJRP* and pCJRP respectively. Construct pLS-133-142-A is derived from plasmid 'T55' obtained from A. Günzl. In this, CGCTAGCGAG replaces the corresponding wt sequence in pΔ237. In pLS-133-142-B, TGCTCGAGTA replaces the corresponding wt sequence in pCJRP*. Construct pLS-152-161 is derived from pJP44.355 (5). In this, GTCTGTCGAC replaces the corresponding wt sequence in pΔ237. In pLS-159-168, pLS-202-211 and pLS-212-221, AAGGATCCAG replaces the corresponding wt sequence in pCJRP. The latter two constructs also show a 1 and 4 bp alteration at -254 and between -252 to -255 respectively that damages their *ApaI* site. Replication is assayed as described in the text and in Table 1.

Luminometer readings correspond to 1/2000 volume of transformed procyclic cultures (strain 427) except in the case of pLS-32-41 where 1/1000 volume was used. Assays were carried out on Days 1, 3, 5 and 7 post-electroporations using a Turner 20/e luminometer and reagents from a PROMEGA luciferase assay kit as previously described [4]. Briefly, cells corresponding to 1/10 volume of the transfected-culture suspension were harvested and lysed with 100 µl of cell lysis buffer. 5 µl of this cell-lysate (10µl in case of pLS-32-41) was used in the assays. Results shown have been normalized for electroporation with 10 µg of DNA in each case (except the no DNA control) and are the average of two independent transfections for each construct carried out on the same day. Luminometer readings corresponding to the two electroporations with each sample varied by 0 -25%. NT=not tested.

Table 1

CONSTRUCT	LUMINOMETER READING				PROLONGED SURVIVAL AS AN EPISOME
	DAY 1	DAY 3	DAY 5	DAY 7	
pEV-luc	20.7	61.5	454	2346	Yes
pΔ237	31.4	76.5	313	1358	Yes
pCJP	28.6	71.3	425	1679	Yes
pLS-32-41	0.02	0.07	0.30	1.85	Yes
pLS-62-71	1.97	1.86	3.70	15.7	Yes
pLS-114-125	18.8	46.6	163	808	NT
pLS-126-135	31.8	60.5	145	690	Yes
pLS-133-142 A	2.44	2.24	2.40	2.37	No
pLS-141-150	18.5	39.8	102	647	Yes
pLS-152-161	24.7	70.4	230	1306	NT
pLS-159-168	28.3	52.1	207	1132	NT
pLS-202-211	8.88	21.0	81.0	538	NT
pLS-212-221	14.5	37.3	169	830	Yes
No DNA	0.002	-0.008	0.005	0.001	-