© 1990 Oxford University Press

The *Trypanosoma brucei* protein phosphatase gene: polycistronic transcription with the RNA polymerase II largest subunit gene

Raymond Evers and Albert W.C.A.Cornelissen*
Max-Planck-Institut für Biologie, Molecular Parasitology Unit, Spemannstrasse 34, 7400 Tübingen, FRG

Received June 11, 1990; Accepted July 30, 1990

EMBL accession nos X52745, X52746

ABSTRACT

We have previously described the trypanosomal gene encoding the largest subunit of RNA polymerase II (RNAP II) and found that two almost identical genes are encoded within the Trypanosoma brucei genome. Here we show by Southern analyses that the 5′ breakpoint between both loci is located approximately 7.5 kb upstream of the RNAP II genes. Northern analyses revealed that the 5′ duplicated segment contains at least four other genes, which are transcribed in both bloodstream and procyclic trypanosomes. The gene located immediately upstream of the RNAP II gene in both loci was characterized by sequence analyses. The deduced amino acid sequences show a high degree of similarity to the catalytic subunit of protein phosphatase class 1 (PP1) genes. S1 mapping provided strong evidence in support of the fact that the PP1 and RNAP II genes belong to a single transcription unit.

INTRODUCTION

African Trypanosomes are protozoan parasites, that alternate between an insect vector, the tse-tse fly, and a mammalian host. They evade the immune response of the mammal by periodically switching their surface antigen, the variant surface glycoprotein or VSG. This process is known as antigenic variation (1–3).

In vitro transcription assays with *Trypanosoma brucei* nuclei (run-ons) showed an unusual property of the VSG transcription unit, namely that the RNA polymerase (RNAP) transcribing this unit is completely resistant to amanitin. All other protein-coding genes are transcribed by an RNAP that is sensitive to amanitin (4,5). Recently, it was shown that the surface antigen (PARP or procyclin) genes of procyclic trypanosomes (6–8) and telomeric sequences (9) are also transcribed by an amanitin-resistant RNAP. Although very likely, it has not yet been formally proven that the same RNAP is involved in these cases. In order to obtain a possible explanation for the amanitin-resistant transcription in *T. brucei*, the trypanosomal RNAPs have recently been characterised by a molecular analysis of the genes encoding the largest subunits and by a biochemical analysis of the enzymes themselves.

The normal set of eukaryotic RNAPs were identified by separating the RNAP I, II and III activities by chromatography (10). Moreover, the genes encoding the largest subunit of RNAP I–III were isolated and identified by sequence analyses. The largest subunit of both RNAP I and III are encoded by single genes, but two genes encode for the largest subunit of RNAP II (11–15). Both gene copies differ in three (13) or four (15) amino acids, depending on the strain analysed, and are found on different chromosomes (13). The second copy of RNAP II is only present in kinetoplastid species performing antigenic variation, suggesting that the second RNAP II gene might generate the amanitin-resistant enzyme (13,16,17). Evidence in support of this hypothesis comes from run-on experiments. A modulation of the Mn^{2+} concentration in the run-ons resulted in a class-specific elongation pattern, which enabled the discrimination between RNAP I–III activities. These experiments showed that the elongation pattern of RNA encoded by the transcription unit of VSG genes was similar to that of the tubulin genes, strongly suggesting that the former unit is transcribed by an RNAP II-like enzyme (10).

An alternative explanation for the presence of the additional RNAP II locus was given by Smith et al. (15). These authors suggested that the duplication event giving rise to the two RNAP II genes happened more recently than the emergence of the salivarian trypanosomes and the process of antigenic variation. The presence of two RNAP II copies and the amino acid differences between both RNAP II largest subunit genes might result in two enzymes with different catalytic properties, nescessary for the parasite to adapt to different environments encountered during the life-cycle.

An analysis of genes that are physically linked to both RNAP II genes might provide some insights into the occurrence of the RNAP II duplication. We therefore decided to characterize the 5′ duplicated segment in more detail with the following aims: (i) to localize the 5′ breakpoint between both RNAP II loci more precisely, (ii) to analyse one of the genes present within the

* To whom correspondence should be addressed

duplicated segment, preferably close to the RNAP II gene, and (iii) to determine whether these genes belong to the same or different transcription units.

RESULTS

Mapping of the 5′ breakpoint between the 4.8 and 5.9 locus

Our initial mapping data (13), suggested that the point of divergence between the Trp4.8 and Trp5.9 loci, encoding the RNAP II genes, maps 1.5 kb upstream of the RNAP II genes. This conclusion was based on the presence of polymorphic restriction sites upstream of the 5′ Eco RI site (marked with an asterisk in Figure 1). For our present analysis of the breakpoint, we cloned the 14 kb and 15 kb Bgl II fragments of both loci (see Materials and Methods). To identify the appropriate clones, the DNA of positive clones was isolated, digested with Hind III/Eco RI and individual Southern blots of these digests were probed with either the 0.8 kb Hind III/Eco RI fragment of pTrp4.8 or with the 2.1 kb Hind III/Eco RI fragment of pTrp5.9. The maps of the phage clones identified in this way are presented in Figure 1.

To map the point of divergence between both loci, Southern blots were prepared of *T.brucei* genomic DNA digested with Bgl II or Cla I. The blots were hybridized with fragments covering 8.7 kb 5′ of the marked Eco RI site of the Trp4.8 locus (Figures 1 and 2). All probes, except probes number 10 and 11, hybridize to the expected Bgl II fragments (14 and 15 kb in size; Figure 1) of both the Trp4.8 and Trp5.9 loci (Figure 2). Probes 10 and 11, however, hybridize only to the Bgl II fragment of the 4.8 locus, even under conditions of reduced stringency (3×SSC, 65°C; data not shown). Similar data were obtained with Cla I digests (Figures 1 and 2). The fact that probe 11 recognizes two large genomic fragments (>20 kb), suggests that the newly mapped Cla I site of the Trp4.8 locus is polymorphic.

We previously suggested that the breakpoint, based on the presence of polymorphic restriction sites, was localized at 1.5 kb upstream of both RNAP II genes. Our present Southern analyses of the cloned Bgl II fragments of both loci strongly indicate that the 5′ boundary of the breakpoint of the Trp4.8 and Trp5.9 loci is located approximately 7.5 kb upstream of both RNAP II genes.

Mapping the breakpoint of the duplicated fragment by northern blotting analyses

To verify the Southern analyses of the breakpoint described above, we have analysed the transcripts derived from this region by northern blotting analysis.

Poly(A)$^+$ RNA from both bloodstream and procyclic trypanosomes was isolated, separated on agarose gels, blotted and hybridized with the probes derived from the Trp4.8 locus used for the Southern analysis. The duplicated fragment contains at least five RNAs that have approximate sizes between 1.5 and 3 kb, covering most of the 5′ area (Figure 2B). In addition to these transcripts additional faint bands can be seen, which most likely represent precursor RNAs. All RNAs are present in both bloodstream and procyclic trypanosomes. It is interesting to note here that the signal obtained with probe 11 is weaker in procyclic trypanosomes (Figure 2B). Since equal amounts of poly(A)$^+$ RNA were loaded, the difference in intensity suggests that the RNA species recognized by probe 11 might be differentially expressed. As expected, all homologous and cross-hybridizing fragments of the Trp5.9 locus, that are covered by probes 1, 8 and 9 of the Trp4.8 locus detect the same RNAs (data not shown). However, the 3.8 kb Nru I/Pvu II fragment of the Trp5.9 locus (probe 5.9 in Figure 2B), which does not cross-hybridize to the corresponding area in the Trp4.8 locus (see above) recognizes, at least, one other RNA (Figure 2B and data not shown). It is of interest that the RNA recognized by this probe gives a slightly larger RNA in procyclics and is apparently also differentially expressed. This is in full agreement with the Southern mapping data, which localizes the breakpoint between the Trp4.8 and Trp5.9 loci in this area.

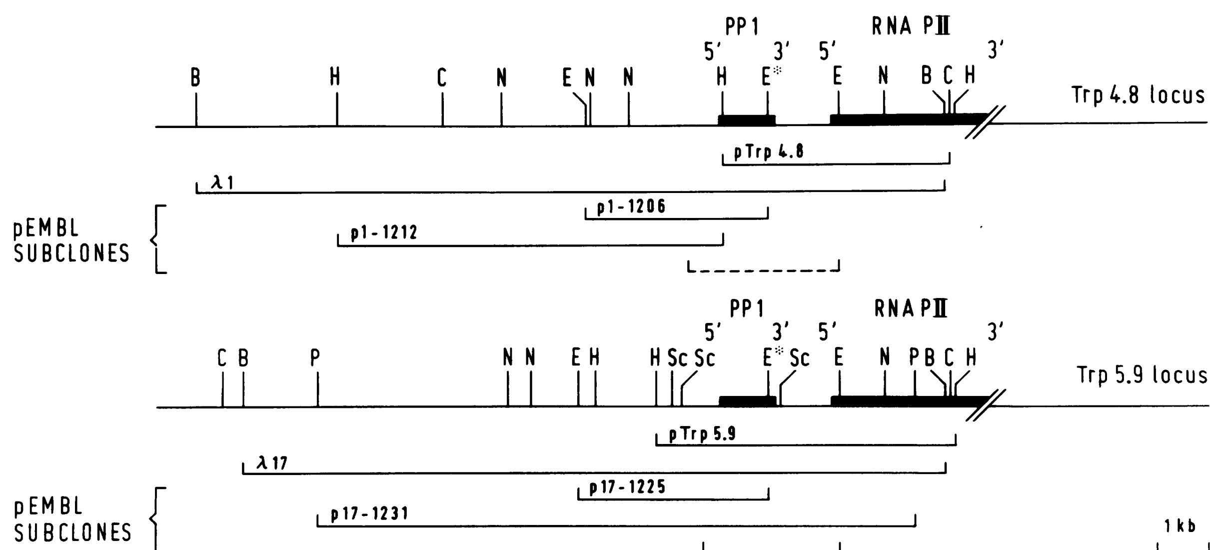

Figure 1. Physical map of the Trp4.8 and Trp5.9 loci. The restriction sites mapped within both loci were derived from mapping data of the isolated lamda clones 1 and 17, and by Southern blotting analyses of genomic DNA (Figure 2A, and 13). The pEMBL subclones derived from the lambda clones 1 and 17 are depicted below the maps. The black boxes represent the coding regions of the protein phosphatase gene (PP1, see text) and RNAP II largest subunit gene (13). The broken line indicates the DNA segment of which the sequence is shown in Figure 3. Abbreviations of restriction enzyme sites: B = Bgl II, C = Cla I, E = Eco RI, H = Hind III, N = Nru I, P = Pvu II, Sc = Sca I.

Cloning of the gene located upstream of the RNAP II genes

The northern blotting experiments indicated that probe 1, which is localized immediately upstream of the RNAP II gene in the Trp4.8 locus, recognizes an mRNA of approximately 2 kb. The homologous probe of the Trp5.9 locus recognizes a similar-sized mRNA. As outlined in the Introduction, we wanted to analyze a gene, localized close to the RNAP II genes, within the duplicated segment. This would allow us to determine the homology of these genes and to be able to compare this with the extent of homology between both RNAP II genes. To identify the gene encoding the 2 kb mRNA, we subcloned and sequenced the regions indicated in Figure 1. The sequencing strategy and procedures are outlined in Materials and Methods.

The sequence analysis of the gene localized in the Trp4.8 locus revealed an open reading frame (ORF) of 1038 nt. The nucleotide sequence of the ORF, the deduced amino acid sequence, and the flanking nucleotide sequences are depicted in Figure 3. To analyse the same region in the Trp5.9 locus, we sequenced the 2.7 kb region immediately upstream of the RNAP II gene.

The deduced amino acid sequence of the gene encoded on the Trp4.8 locus predicts a 39 kD protein. A computer search for homologous sequences indicated that the gene had a striking homology to the catalytic subunit of protein phosphatases (PPs). PPs exist in two classes, named PP1 and PP2, which are in turn subdivided in isoforms (18). The *T.brucei* PP-gene showed an overall absolute homology of 40% with the gene encoding the rat PP2a gene and 56% with the rat $PP1_{alpha}$ gene (Figure 4). Since most amino acid substitutions in the *T.brucei* gene were found to be clustered in the N- and C-termini, we removed fifty 'trypanosomal-specific' N- and C-terminal amino acids in order to improve a subsequent computer comparison. This intervention resulted in an increased absolute homology (64%) with $PP1_{alpha}$. Hence we conclude that the sequence corresponds to the eukaryotic PP1 genes rather than PP2, and therefore named this gene $PP1_{4.8}$. Since the identity of the *T.brucei* PP1 gene was similar to all known PP1 subclasses and also lacked isoform-specific domains (P.T.W. Cohen, personal communication), we are unable to classify the *T.brucei* gene to any of the known isoforms.

The sequence of the gene localized on the Trp5.9 locus is highly homologous to $PP1_{4.8}$ (Figure 2) and therefore also encodes a putative PP1 gene ($PP1_{5.9}$). That this gene was also transcribed was shown by gene specific probes located just upstream from the genes (not shown). Most differences between both *T.brucei* genes were found to be clustered. There are 37 nucleotide substitutions in the N-terminus and 9 nucleotide substitutions in the C-terminus, resulting in 10 and 4 amino acid replacements, respectively. Interestingly, the homology between both loci is lost immediately up- and downstream of the $PP1_{4.8}$ and $PP1_{5.9}$ genes.

The PP1 and RNAP II genes are transcribed as a polycistronic precursor RNA

Most trypanosomal housekeeping genes analysed to date have been found in tandem arrays and are transcribed as single polycistronic transcription units (see for review 19,20). To determine whether the RNAP II gene and the PP1 gene of both loci belong to a single transcription unit or whether they belong

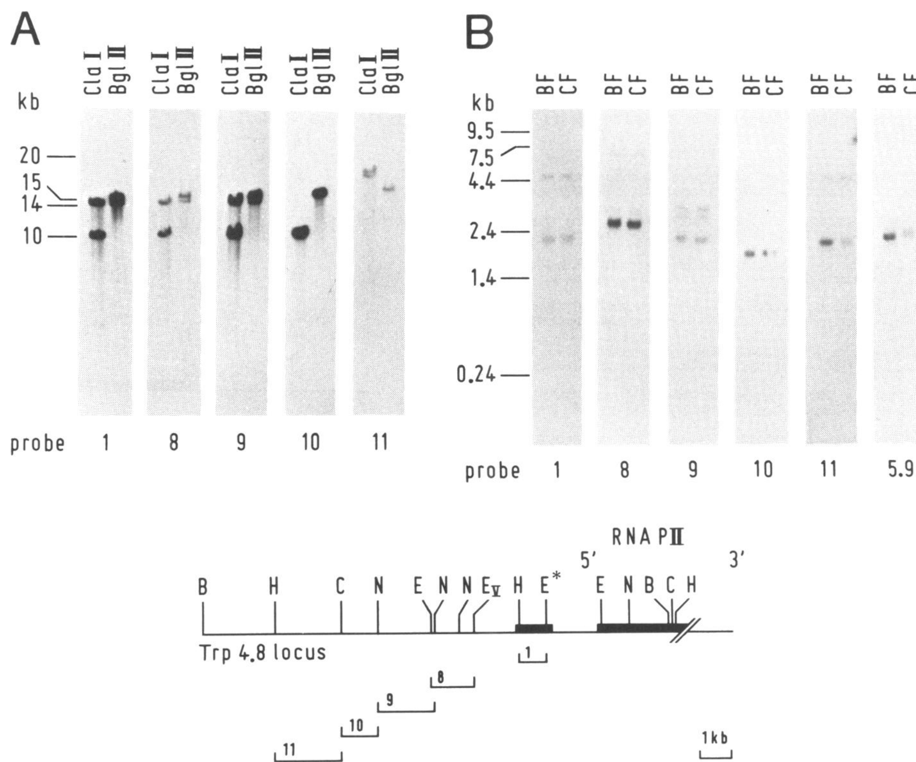

Figure 2. Mapping of the breakpoint between the Trp4.8 and Trp5.9 loci. Four ug of *T.brucei* DNA was digested with BglII or ClaI, size fractionated on a 0,5% agarose gel and transferred to Biodyne B membranes. Identical blots were hybridised to nick-translated fragments derived from the Trp4.8 locus (Panel A). The fragments used are indicated below the map. Abbreviations of restriction enzyme sites: Ev = Eco Rv, others as in Figure 1. Five ug poly $(A)^+$ RNA of bloodstream and procyclic trypanosomes was size-fractionated in a 1% formaldehyde agarose gel and transferred to Biodyne B membranes (Panel B). Blots were hybridised to the same probes as in panel A.

Figure 3. Nucleotide and deduced amino acid sequence of the $PP1_{4.8}$ and $PP1_{5.9}$ genes. The upper line shows the deduced amino acid sequence of the $PP1_{4.8}$ gene. Amino acid changes of $PP1_{5.9}$ are given behind the slash above the codon in question. The lower line shows the nucleotide sequence of the Trp4.8 locus. The sequence of the Trp5.9 locus is depicted in the bottom line, only the differences between the $PP1_{4.8}$ and $PP1_{5.9}$ genes are indicated. The 5′ and 3′ flanking sequences are depicted in two lines, the top line gives the sequence of the Trp4.8 locus, the bottom line that of the Trp5.9 locus. The putative splice acceptor sites are underlined, the putative polyadenylation signals are underlined twice. These sequences will appear in sequence databases under the accession numbers X52745 and X52746 for the Trp4.8 and Trp5.9 locus, respectively.

Figure 4. Dot matrix analysis showing the homology between $PP1_{4.8}$ gene and rat protein phosphatase genes. The deduced amino acid sequence of the $PP1_{4.8}$ gene was run against the amino acid sequences of the $PP1_{alpha}$ and PP2a enzymes of rat liver (52, 53). Dots were produced whenever 6 or more out of 8 residues were identical.

Figure 5. Mapping of the $PP1_{4.8}$ transcript. S-1 nuclease protection analyses were performed with the fragments indicated below each panel. The hybridisation temperature was 50°C. Lane 1, input fragment 4 (3′ end-labeled), lane 2, input fragment 4 (5′ end-labeled), lane 3, input fragment 6 (5′ end-labeled). The fragments in lanes 4, 6 and 8 were hybridised to tRNA. The fragments in lanes 5, 7 and 9 were hybridised to total bloodstream form RNA.

to two separate units, we mapped the start and stop of transcription of the appropriate DNA segments by S1 analysis.

The $PP1_{4.8}$ mRNA was mapped using end-labelled probes. The 5′ end of the $PP1_{4.8}$ gene was mapped with the 700 bp Sca I/Hind III fragment (fragment 5) and a 210 bp AluI fragment (fragment 6). Fragment 5 resulted in one prominent protected

fragment of 500 nt (data not shown). The outcome of this experiment was confirmed in an additional experiment with fragment 6. Fragment 6, which must contain the 5′ start site, was labelled at both 5′ ends. A single fragment of 145 nt was protected (Figure 5). This position maps at an AG-splice acceptor site which is preceded by the trypanosome specific oligo-T stretch (Figure 3; cf.(21)). Since the mature mRNA of the $PP1_{4.8}$ is approximately 2 kb in size and the mapping of the 5′ end maps 145 nt upstream of the start codon, the 3′ end of the $PP1_{4.8}$ gene has to be located within the 3′ 770 bp Nar I/Eco RI fragment (fragment 4). Fragment 4 was 3′ end-labeled at the Nar I site and the $PP1_{4.8}$ RNA protects approximately 250 nt of this fragment (Figure 5). This position maps at a putative, weakly defined, polyadenylation signal (cf. 22) With the use of the same fragment, we were able to confirm the previously mapped 5′ end of the RNAP II mRNA (15), by labelling fragment 4 at the 5′ end of the Eco RI site. This resulted in two protected fragments, approximately 240 nt and 380 nt in size (cf. (15)). The smaller fragment can be explained by the loss of homology between the sequence of the Trp4.8 and Trp5.9 loci in this area (13,15).

In conclusion, these S1 protection experiments suggest that the $PP1_{4.8}$ RNA contains, in addition to the ORF of 1008 nt, a 5′ non-translated sequence of 470nt and a 670 nt non-translated tail sequence.

The long 3′ tail leaves only 130 bp between the 3′ end of the $PP1_{4.8}$ mRNA and the 5′ end of the RNAP II mRNA. If both genes belong to a single transcription unit, we should be able to protect the 130 nt spacer region in S1 mapping experiments. For this experiment, to rule out the possibility of re-annealing of the fragment, we used single-stranded pEMBL clones (see Materials and Methods) encoding the 1.3 kb Eco RI fragment of both the Trp4.8 and Trp5.9 loci (Figure 1). These DNAs were hybridized to steady state total RNA. The coding strand served as a control for both the specificity of hybridization and the direction of transcription. If PP1 and RNAP II belong to the same polycistronic transcription unit, the prediction is that the precursor RNA will generate three protected fragments; a fragment of 1300 nt, which spans the intergenic region between the RNAP II and PP1 genes, and fragments of 830 nt and 380 nt from the processed PP1 and RNAP II mRNAs, respectively. All these three fragments were detected for both the Trp4.8 and Trp5.9 loci (Figure 6), strongly suggesting that the PP1 and RNAP II genes are transcribed as a polycistronic transcription unit. Differences in intensities between bands are due to blotting artefacts.

We performed nuclear run-on assays in order to confirm the outcome of these results. In experiments in which a 120 nt gene-internal fragment of RNAP II resulted in a positive signal, we also obtained a positive signal with a 240 nt fragment of the intergenic region, which is consistent with the observations made by S1 analysis. However, in some of the run-on experiments we were unable to detect a signal to the RNAP II gene-internal probe as well as to the intergenic probe. This was most likely due to the necessity of using small restriction fragments and the low overall transcription level of the transcription unit (data not shown).

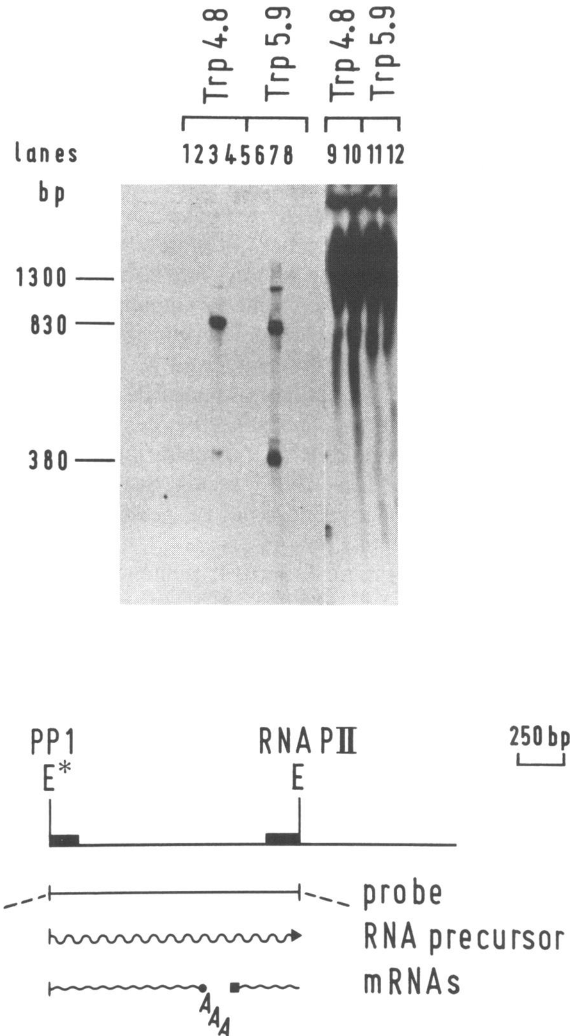

Figure 6. S1 analyses of the intergenic region between the PP1 and RNAP II genes. Single stranded pEMBL DNAs containing the 1300 bp Eco RI fragment spanning the intergenic region between the PP1 and RNAP II genes of the Trp4.8 and Trp5.9 loci (see Materials and Methods for procedure), were hybridised at 42°C to total RNA isolated from bloodstream trypanosomes and digested with S-1. The protected fragments were fractionated on a denaturing 4% PAA gel, transferred by semi-dry blotting to Biodyne B membranes and hybridised at 65°C to the 1300bp Eco RI probe from both the Trp4.8 and Trp5.9 locus. The fragments predicted for a polycistronic transcription unit are depicted below the figure. In lanes 1–4 the single standed probe of the Trp4.8 locus is hybridised to total RNA (1 and 3) or tRNA (2 and 4). Lanes 1,2 and 3,4 represent the sense and anti-sense probe, respectively. In lanes 5–8 the same probe of the Trp5.9 locus is hybridised to total RNA (5 and 7) or tRNA (6 and 8). Lanes 5,6 and 7,8 represent the sense and anti-sense probes, respectively. The input fragments are depicted in lanes 9–12, respectively.

DISCUSSION

The 5′ end of the duplicated segment

In this paper we describe the cloning and, Southern and northern blotting analyses of the 5′ end of the duplicated DNA segments that encode the RNAP II largest subunit genes (13). The data show that the 5′ boundary of the duplicated segment of the Trp4.8 and Trp5.9 loci is located approximately 7.5 kb upstream of both RNAP II genes. This is based on the cross-hybridization of DNA fragments derived from both loci to DNA downstream of the breakpoint and the presence of identical poly $(A)^+$ RNAs encoded on the duplicated segment (Figure 2). Upstream of the 5′ boundary the cross-hybridization is lost, even at reduced stringency conditions, and different poly $(A)^+$ RNAs were found. The homologous 5′ segments encode at least four poly $(A)^+$ RNAs, suggesting that these segments are tightly packed with genes. Since all transcripts were found in both bloodstream and procyclic trypanosomes, these genes are most likely general

housekeeping genes. However, whether the putative ORFs are intact and what the function of these genes is remains to be determined.

The duplicated segment encodes a protein phosphatase gene

The RNAP II genes of both the Trp4.8 and Trp5.9 loci are, based on our northern analyses, flanked by another gene at their 5′ end. We cloned and sequenced the DNA segments hybridizing to this mRNA. The amino acid sequences of these genes most closely resembles that of the catalytic subunit of PP1 genes (Figures 3 and 4). The PP1 enzymes are found throughout the eukaryotes, and are characterized by an extremely high homology among evolutionary distinct organisms (reviewed by (18,23,24)). We were unable to classify the trypanosomal PP1 genes to one of the known PP1 isoforms. The yeast PP gene that might be involved in the regulation of RNAP II transcription (25) belongs to the PP2 class of enzymes. Thus it seems reasonable to exclude the possibility that the trypanosomal PP genes identified here would have a similar function.

The S1 analyses (Figure 5 and 6), which indicate that the intergenic region of PP1 and RNAP II genes is transcribed, strongly suggest that the PP1 genes belong to the same transcription unit as the RNAP II genes. This situation seems to be standard in *T.brucei* since a similar organization has been observed for the tubulin genes (26,27), the phosphoglycerate kinase genes (28), the actin genes (29) and the VSG transcription unit (30–33). A similar organization has also been reported for *T.cruzi*, where putative precursors were detected from a tandemly arranged repetitive gene family with unknown function (34) and for the alpha- and beta tubulin mRNAs (35). Therefore, polycistronic transcription units might be a general feature of trypanosomes or all kinetoplastid parasites. It is interesting to note here that the transcription units of the PP1 and RNAP II genes are the first examples of an amanitin-sensitive transcription unit which is not composed of tandemly arranged genes.

Subsequent processing of the precursor RNA, leading to the formation of the mature mRNAs by the addition of a mini-exon, which is present at the 5′ end of all trypanosomal mRNAs (20,36,37), is also suggested by our S1 analyses. The transcription starts of both the PP1 and that of the RNAP II mRNAs map exactly at places where the DNA sequence contains a splice acceptor site that is preceded by a T-stretch (Figures 3 and 4), as has been observed in most instances (22). Moreover, in the case of the RNAP II mRNAs, it was shown by sandwich hybridization that the mature message contains the mini-exon sequence (11).

PPI, RNAP II and amanitin-resistant transcription

Nuclear run-on experiments showed that the trypanosomal transcription unit encoding the VSG gene is transcribed by an amanitin-resistant RNAP (4). It was subsequently demonstrated that the surface antigen genes of procyclic trypanosomes (6–8) and telomeric sequences (only in salivarian trypanosomes (9)) were also transcribed by an amanitin-resistant RNAP. This correlated with the presence of a second copy of the RNAP II largest subunit gene in all analysed salivarian trypanosomes, and the expression of both copies in both bloodstream and procyclic trypanosomes (13,15,17). In the case of VSG genes we were able to demonstrate that the RNAP involved in VSG gene transcription is probably modified by an auxiliary factor to generate amanitin resistance (17).

Our present analysis shows clearly that different replacement rates are found within the duplicated segment of both RNAP loci. Between both RNAP II genes 19 nucleotide substitutions are present, giving rise to three amino acid substitutions out of 1765 (13); between the PP1 genes there are 51 nucleotide substitutions (Figure 3), giving rise to 14 amino acid substitutions out of 347. If one assumes a replacement site divergence rate of 0,1% per million years (15), the genes would have diverged roughly one million and 50 million years ago. Since it is clear from the Southern, northern and S1 analyses that the RNAP II and PP1 genes have been duplicated in a single event, these data are not compatible and demonstrate the pitfalls of trying to date such an event (vs. (15)). The relatively high number of silent replacements is therefore most likely due to a functional constraint on the amino acid sequence of the RNAP II polypeptides.

The present analysis shows that the duplication must be older than previously anticipated (15) and could have coincided with the emergence of antigenic variation and amanitin-resistant RNAP II transcription in *T.brucei* and related salivarian trypanosomes.

MATERIALS AND METHODS

Materials

Restriction and modifying enzymes were purchased from Pharmacia-LKB, Boehringer Mannheim and Applied Genetechnology Systems. Radionucleotides were obtained from Amersham. Sequence analysis was performed with the program described by Queen and Korn ((38), Microgenie, Beckman Instruments).

Trypanosomes

The cloned variant antigen type MiTat 1.1c (39), an uncloned population 1.1cR (R for relapse), and 118a (40) from *T. brucei* 427 were used. Bloodstream trypanosomes were grown in rats to a density of 10^9 per ml blood and purified from blood cells by anion exchange chromatography. Procyclic trypanosomes were grown in the semi-defined medium described by Brun and Schönenberger (41).

DNA and RNA isolation

Trypanosomal total DNA and RNA were isolated by standard methods (42,43). Poly(A)$^+$ RNA was purified by oligo (dT) cellulose chromatography (44). Plasmid DNA was isolated by the alkaline lysis method (45).

Library preparation and screening

A BglII genomic library of size-selected fragments (10–20 kb) was generated by ligation of the isolated fragments in the Bam H1 site from lambda EMBL-3 (46). Recombinant phages were screened according to standard protocols (47) using the complete HindIII inserts of pTrp4.8 and pTrp5.9 (13).

DNA sequence analysis

The 3.7 kb Eco RI and 7.5 kb HindIII fragments from lambda 1, and the 3.8 kb Eco RI and 12 kb PvuII fragments from lambda 17 were subcloned in pEMBL 8/9 vectors (48), resulting in p1-1206, p1-1212, p17-1225 and p17-1231 (see Figure 1). The sequence from the PP-$1_{4.8}$ gene was determined by subcloning the 1.3 kb Eco RI and 0.8 kb Hind III/Eco RI fragments of pTrp4.8, and the 1,8kb NruI/HindIII fragment of p1206. These subclones were digested with the appropriate restriction enzymes, treated with Bal31 nuclease in order to obtain a series of progressively deleted fragments, and digested with a second

restriction enzyme. The fragments obtained in this way were subcloned into the appropriate pEMBL 8/9 vectors. Fragments spanning the HindIII and Eco RI site were obtained by cloning small fragments overlapping these sites or by using oligodeoxynucleotides as primers. The PP-$1_{5.9}$ gene was sequenced by subcloning the 1.7 kb ScaI and the 1.3 kb Eco RI fragments of pTrp5.9. Deletion clones were obtained as described above. All sequences were determined by the chain termination method (49). 7-Deaza dGTP was used as a substitute for dGTP in all reactions.

DNA and RNA manipulations

Restriction endonuclease digestion, electrophoresis of DNA or RNA and end labeling of DNA fragments were according to standard procedures (47). RNA and DNA were transferred to Biodyne B nylon membranes (PAL) as described by the manufacturer. DNA and RNA blots were hybridized with nick-translated ^{32}P-labelled probes (50) as previously described (13). All post hybridizational washes were to a final stringency of 0.1×SSC at 65°C (1×SSC = 0.15 M NaCl, 0.015 M sodium citrate, pH7.0).

S1 protection analyses

S-1 analyses were performed either by using end-labelled double stranded DNA fragments and direct analysis of the S-1 product(s) on polyacrylamide gels, or by using single stranded pEMBL DNAs. Single stranded templates of the 1300bp Eco RI fragment were prepared by the same method as the templates used for sequencing single stranded DNA (48). In the latter case the S-1 product(s) were separated on polyacrylamide gels, blotted to Biodyne-B membranes and subsequently hybridized with an appropriate probe. The S-1 analyses were carried out as previously described (51). Briefly, 10ng of end-labeled DNA was coprecipitated with 10μg total bloodstream form RNA and dissolved in 20μl S-1 buffer (52). After heating at 85°C for 15 min, the probes were incubated overnight at the temperatures indicated in the text. Following digestion with 350 U/μl S-1 nuclease (Boehringer-Mannheim) in a final volume of 300μl and incubation for 30 min at 30°C, the probes were precipitated and analysed on a 6% or 4% (w/v) denaturing acrylamide gel. Size markers were end-labelled fragments of pBR322/Msp I or pEMBL/Taq I.

ACKNOWLEDGEMENTS

We like to thank Andrea Hammer for technical assistance and Dr. Johannes Pohlner (Infektgenetik) for the synthesis of oligodeoxynucleotides. We also thank all our colleagues from the laboratory for helpful discussions throughout this work, and Drs. Piet Borst and Carol Gibbs for constructive criticism on the manuscript. The work was supported by the Bundesministerium für Forschung und Technologie grant 031885, and an award of the Erna and Victor Hasselblad Foundation to Albert W.C.A. Cornelissen.

REFERENCES

1. Vickerman, K. (1969) J. Cell Sci. **5**, 163–193
2. Vickerman, K. (1978) Nature **273**, 613–617
3. Cross, G.A.M. (1975) Parasitology **71**, 393–417
4. Kooter, J.M., and Borst, P. (1984) Nucl. Acids Res. **12**, 9457–9472
5. Laird, P.W., Kooter, J.M., Loosbroek, N., and Borst, P. (1985) Nucl. Acids Res. 13, 4253–4266
6. König, E., Delius, H., Carrington, M., Williams, R.O., and Roditi, I. (1989) Nucl. Acids Res. **18**, 8727–8740
7. Rudenko, G., Bishop, D., Gottesdiener, K., and Van der Ploeg, L.H.T. (1989) EMBO J. **8**, 4259–4263
8. Clayton, C.E., Fueri, J.P., Itzhaki, J.E., Bellofato, V., Sherman, D., Wisdom, G.S., Vijayasarathy, S., and Mowatt, M.R. (1990) Molec. Cell. Biol. **10**, in press
9. Rudenko, G., and Van der Ploeg, L.H.T. (1989) EMBO J. **8**, 2633–2638
10. Grondal, E.J.M., Evers, R., Kosubek, K., and Cornelissen, A.W.C.A. (1989) EMBO J. **8**, 3383–3389
11. Köck, J., Evers, R. and Cornelissen, A.W.C.A. (1988) Nucl. Acids Res. **16**, 8753–8772
12. Jess, W., Hammer, A., and Cornelissen, A.W.C.A. (1989) FEBS Lett. **249**, 123–128
13. Evers, R., Hammer, A., Köck, J., Jess, W., Borst, P., Mémet, S., and Cornelissen, A.W.C.A. (1989) Cell **56**, 585–597
14. Smith, J.L., Levin, J.R., and Agabian, N. (1989) J. Biol. Chem. **264**, 18091–18099
15. Smith, J.L., Levin, J.R., Ingles, C.J., and Agabian, N. (1989) Cell **56**, 815–827
16. Cornelissen, A.W.C.A., Evers, R., Grondal, E., Hammer, A., Jess, W., and Köck, J. (1989) Acta Leidensia **58**, 2–23
17. Grondal, E.J.M., Evers, R., Köck, J., Jess, W., and Cornelissen, A.W.C.A. (1990) UCLA symp. on Molec. Cell. Biol. New series, Vol. 131 (N. Agabian & A Cerami, Eds.), Alan R. Liss Inc. New York, pp. 279–296
18. Cohen, P., and Cohen, P.T.W. (1989) J. Biol. Chem. **264**, 21435–21438
19. Pays, E., and Steinert, M. (1988) Ann. Rev. Genet. **22**, 107–126
20. Boothroyd, J.C. (1989) Nucl. Acids and Molec. Biol. **3**, 216–230
21. Laird, P.W., Zomerdijk, J.C.B.M., De Korte, D., and Borst, P. (1987) EMBO J. **6**, 1055–1062
22. Laird, P.W. (1988) PhD thesis, University of Amsterdam
23. Cohen, P. (1988) Proc. R. Soc. London Ser. B **234**, 115–144
24. Cohen, P. (1989) Ann. Rev. Biochem. **58**, 453–508
25. Arndt, K.T., Styles, C.A., and Fink, G.R. (1989) Cell **56**, 527–537
26. Imboden, M.A., Laird, P.W., Affolter, M., and Seebeck, T. (1987) Nucl. Acids Res. **15**, 7357–7370
27. Muhich, M.L., and Boothroyd, J.C. (1988) Molec. Cell. Biol. **8**, 3837–3846
28. Gibson, W.C., Swinkels, B., and Borst, P. (1988) J. Molec. Biol. **201**, 315–325
29. Ben-Amar, M.F., Pays, A., Tebabi, P., Dero, B., Seebeck, T., Steinert, M., and Pays, E. (1988) Molec. Cell. Biol. **8**, 2166–2176
30. Kooter, J.M., van der Spek, H.J., Wagter, R., d'Oliveira, C.E., van der Hoeven, F., Johnson, P.J., and Borst, P. (1987) Cell **51**, 261–272
31. Johnson, P.J., Kooter, J.M., and Borst,P. (1987) Cell **51**, 273–281
32. Alexandre, S., Guyaux, M., Murphy, N.B., Coquelet, H., Pays, A., Steinert, M. and Pays, E. (1988) Molec. Cell. Biol. **8**, 2367–2378
33. Pays, E., Tebabi, P., Pays, A., Coquelet, H., Revelard, P., Salmon, D., and Steinert, M. (1989) Cell **57**, 835–845
34. Gonzales, A. Lerner, T.J., Huescas, M., Sosa-Pineda, B., Nogueira, N., and Lizardi, P. (1985) Nucl. Acids Res. **13**, 5789–5804
35. Soares, C.M.A., de Carvalho, E.F., Urmenyi, T.P., Carvalho, J.F.O., de Castro, F.T., and Rondinelli, E. (1989) FEBS Lett. **250**, 497–502
36. Borst, P. (1986). Ann. Rev. Biochem. **55**, 701–732
37. Van der Ploeg, L.H.T. (1987) Cell **51**, 159–161
38. Queen, C., and Korn, L.J. (1984) Nucl. Acids Res. **12**, 581–599
39. Cornelissen, A.W.C.A., Johnson, P.J., Kooter, J.M., Van der Ploeg, L.H.T., and Borst, P. (1985) Cell **41**. 825–832
40. Michels, P.A.M., Liu, A.Y.C., Bernards, A., Sloof, P., Van der Bijl, M.M.W., Schinkel, A.H., Menke, H.H., Borst, P., Veeneman, G.H., and Van Boom, J.H. (1983) J. Molec. Biol. **166**, 537–556
41. Brun, R., and Schönenberger, M. (1979) Acta Trop. **36**, 289–292
42. Van der Ploeg, L.H.T., Liu, A.Y.C., Michels, P.A.M., De Lange T., Borst, P., Majumder, H.K., Weber, H., Veeneman, G.H., and van Boom, J. (1982) Nucl. Acids Res. **10**, 3591–3604
43. Auffray, C., and Rougeon, F. (1980) Eur. J. Biochem. **107**, 303–314
44. Hoeijmakers, J.H.J., Borst, P., Van den Burg, J., Weissmann, C., and Cross, G.A.M. (1980) Gene **8**, 391–417
45. Birnboim, M.C., and Doly, J. (1979) Nucl. Acids Res. **7**, 1513–1523
46. Frischauf, A.-M., Lehrach, H., Poustka, A., and Murray, N. (1983) J. Molec. Biol. **170**, 827–842
47. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular cloning. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory)
48. Dente, L., Cesareni, G., and Cortese, R. (1983) Nucl. Acids Res. **11**, 1645–1655
49. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467