Volume 16 Number 18 1988 Nucleic Acids Research

Structure and sequence of the gene for the largest subunit of trypanosomal RNA polymerase III

Josef Köck, Raymond Evers and Albert W.C.A.Cornelissen*

Max-Planck-Institut für Biologie, Molecular Parasitology Unit, Spemannstrasse 34, 7400 Tübingen, FRG

Received July 15, 1988; Revised and Accepted August 22, 1988 Accession no. X12494

ABSTRACT

As a first step in the analysis of the transcription process in the African trypanosome, *Trypanosoma brucei*, we have started to characterise the trypanosomal RNA polymerases. We have previously described the gene encoding the largest subunit of RNA polymerase II and found that two almost identical RNA polymerase II genes are encoded within the genome of *T.brucei*. Here we present the identification, cloning and sequence analysis of the gene encoding the largest subunit of RNA polymerase III. This gene contains a single open reading frame encoding a polypeptide with a Mr of 170 kD. In total, eight highly conserved regions with significant homology to those previously reported in other eukaryotic RNA polymerase largest subunits were identified. Some of these domains contain functional sites, which are conserved among all eukaryotic largest subunit genes analysed thus far. Since these domains make up a large part of each polypeptide, independent of the RNA polymerase class, these data strongly support the hypothesis that these domains provide a major part of the transcription machinery of the RNA polymerase complex. The additional domains which are uniquely present in the largest subunit of RNA polymerase I and II, respectively, two large hydrophylic insertions and a C-terminal extension, might be a determining factor in specific transcription of the gene classes.

INTRODUCTION

Eukaryotic transcription is regulated by three DNA-dependent RNA polymerases, classified as I, II and III. Each of these enzymes is responsible for synthesising different classes of RNA and differ in their sensitivity to the potent RNA polymerase inhibitor, alpha-amanitin (for review see 1,2). In recent years considerable progress has been made in understanding the transcription process in eukaryotes at the molecular level. Although other proteins play a major role in the initiation of this complex process (3,4), the RNA polymerases can be considered the constant entity of the transcription process.

The transcription process of kinetoplastid flagellates, such as *Trypanosoma brucei,* is rather unique among eukaryotes and characterised by a number of biochemical peculiarities, such as discontinuous transcription and *trans*-splicing (for review see 5-7). One of the first steps in the characterisation of the transcription process in this group of protozoan parasites has been the analysis of the DNA-dependent RNA polymerases. Initial characterisation of kinetoplastid RNA polymerases has resulted in conflicting biochemical observations. *In vitro* transcription assays with isolated nuclei of *T.brucei* in the presence of various concentrations of alpha-amanitin showed that transcription of the genes for rRNA, tubulin and 5S RNA is, repectively, insensitive, sensitive and intermediately sensitive to alpha-amanitin (8,9), as would be expected for transcription of these genes (1,2). This suggested the presence of RNA polymerase I, II and III in *T.brucei*. However, separation of individual enzyme activities by DEAE-sephadex chromatography, the standard technique used to separate eukaryotic RNA polymerases, resulted in at most two RNA polymerase activity peaks in *T.brucei*, one sensitive and one insensitive to alpha-amanitin (10-13).

Another, unexpected, outcome of the nuclear run-on experiments mentioned above was that transcription of surface antigen genes is not inhibited by alpha-amanitin and showed a profile identical to that of the rRNA genes (8,9). This phenomenon has been observed in each trypanosome variant analysed thus far (8; 14-19). The nature of the RNA polymerase transcribing the surface antigen genes is unknown. There is, however, evidence in favour of RNA polymerase I, which normally only transcribes the rRNA genes (18,19). Alternatively, an additional RNA polymerase may exist for the transcription of this specific class of genes (20).

Taken together, these observations suggest that transcription in trypanosomes does not follow the classical eukaryotic pattern, in which three RNA polymerases, I, II and III, are responsible for synthesising the different RNA classes (1,2). In order to determine how many RNA polymerases are present in *T.brucei* and which RNA polymerase transcribes VSG genes, we started to analyse the trypanosomal RNA polymerases. We have

isolated the genes encoding the largest subunits of RNA polymerases by heterologous hybridisation with yeast and *Drosophila* probes. Here, we report the isolation and complete sequence of the trypanosomal gene encoding the largest subunit of RNA polymerase III. In this paper, we also report a comparative analysis of the primary sequence of trypanosomal and other eukaryotic genes. Unique sequence motifs that enable the three classes of RNA polymerases to be readily distinguished have been identified. Moreover, our analysis not only allows us to refine the previously described homology regions, but also enables us to identify important functional domains. The most important conclusion is that the conserved domains of the largest subunit of eukaryotic RNA polymerases probably provide a large part of the transcription machinery of the enzyme complex, while the RNA polymerase I and II-specific domains might be involved in determining the specificity of the RNA polymerase complex.

EXPERIMENTAL PROCEDURES

Materials

Restriction enzymes and modifying enzymes were purchased from Pharmacia-LKB. ^{32}P-labeled nucleotides were purchased from Amersham. Sequence analysis was performed with the program described by Queen and Kan (21; Microgenie™, Beckman Instruments). Recombinant clones carrying the largest subunit genes of yeast RNA polymerase I (pA4; 22), III (pC4; 23) and *Drosophila* RNA polymerase II (DmRpII215; 24) were kindly provided by, respectively, Dr. A. Sentenac and Dr. A. Greenleaf.

Trypanosomes

The cloned variant antigen type Mitat 1.1c (14) and an uncloned population 1.1cR (R for relapse) from *T.brucei* were used. Trypanosomes were grown in rats to a density of 10^9 parasites per ml of blood and purified from blood elements by anion exchange chromatography (25). Procyclic culture form trypanosomes were grown in the semi-defined medium described by Brun and Schönenberger (26).

Nucleic acid isolation

Total RNA was prepared from isolated trypanosomes by LiCl-Urea precipitation (27). DNA was removed from the preparations by incubation with RNase-free DNase. Poly(A)$^+$ RNA was purified by oligodeoxythymidilate-cellulose column chromatography (28). Trypanosomal DNA was isolated by standard procedures (21). Plasmid DNA was isolated by the alkaline lysis procedure (30).

DNA and RNA manipulations

Restriction endonuclease digestions were done under the conditions specified by the manufacturer. Electrophoresis and transfer of DNA to nitrocellulose filters were performed as

described (31,32). RNA (northern) blots were performed according to the formaldehyde procedure (33). Southerns and northerns were prehybridised as described (32), and hybridised with nick-translated (70) ^{32}P-labelled probes as described by Jeffreys and Flavell (34), with the addition of 10% dextran sulphate to increase the efficiency of hybridisation (35). All post-hybridisation washes were to a final stringency of 0.1xSSC, 0.1% SDS at 65°C (1xSSC = 0.15M NaCl, 0.015M sodium citrate, pH 7.0) or as indicated in the text. Dot blot analysis was essentially performed as described (36). 200ng DNA was bound to nitrocellulose and hybridised with 1mg total RNA in 10ml as described above, washed posthybridisationally to 0.3xSSC, and subsequently hybridised with a kinased 22mer mini-exon probe ($^{3'}$TCTTGTCAAAGACATGATATAA$^{5'}$) at 30°C as described (37). Non-specifically bound activity was removed by washing in 3xSSC at 35°C.

Isolation of pTrp28

The recombinant plasmid pTrp28 was isolated from a library containing HindIII-restricted genomic DNA of *T.brucei*. This library was constructed in the vector pBR322 using *E.coli* DH1 as the host (38). Three plasmids were isolated using the 2.8 kb EcoRI fragment of pC4 as a hybridisation probe under low stringency conditions (3xSSC at 65° C).

Pulsed field gradient gel electrophoresis

Trypanosomes were prepared for PFGE according to standard procedures (39), separation of the chromosomes was performed according to the modifications and specifications of Johnson and Borst (40). The separation shown in Figure 6 was obtained with a 0.5% agarose gel and a run of 24 hr at 16°C, and the electric field (190 V) was switched every 260 sec.

DNA sequencing

The 6.0 kb EcoRI/PvuII fragment of pTrp28 (Figure 2) was isolated from the genomic clone and subcloned into pGEM-3. Restriction fragments of the resulting plasmid, pTrp6, were subcloned into pEMBL 8/9. Cloning and preparation of template DNA was performed using standard protocols (33,41). DNA sequencing was performed using the dideoxy method (42) with modifications described by Biggin et al. (43); both strands were sequenced.

RESULTS

Cloning of the largest subunit gene of RNA polymerase III

The coding region of the trypanosomal largest subunit gene of RNA polymerase III was identified by hybridising Hind III digested genomic DNA of *T.brucei* with a yeast RNA polymerase III probe. At reduced stringency conditions (3xSSC at 65°C) a 28 kb fragment was detected which did not hybridise with probes encoding the largest subunit gene of yeast RNA polymerase I and *Drosophila* RNA polymerase II. This fragment, pTrp28, was isolated from a Hind III genomic library of *T.brucei* in pBR322 (Figure 1, lanes A-C). Southern analysis using different

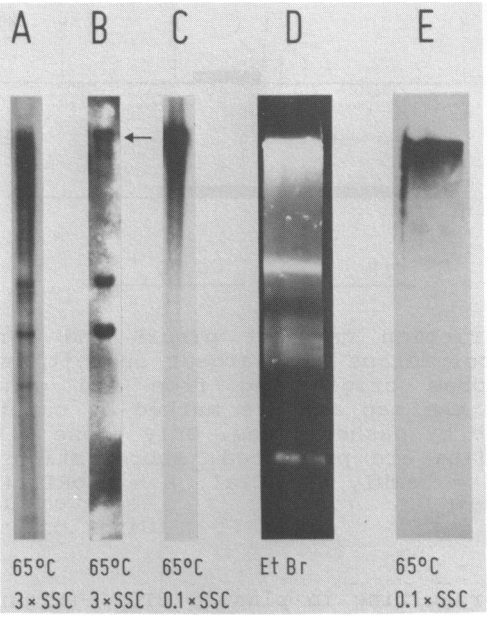

Figure 1. Identification and isolation of the largest subunit gene of RNA polymerase III of *T.brucei*. Four ug of DNA from variant 1.1c was digested with HindIII, size fractionated on a 0.7% agarose gel, and transferred to nitrocellulose. The blots were probed with nick-translated probes encoding the largest subunits of *Drosophila* RNA polymerase II (9.5 kb XbaI fragment from clone DmRpII215, (24)) (lane A) and yeast RNA polymerase III (2.8 kb EcoRI fragment from clone pC4, (23)) (lane B). The arrow indicates the single 28 kb fragment that cross-hybridises to the yeast RNA polymerase III probe at low stringency conditions. An identical blot was hybridised with the insert of the isolated plasmid pTrp28, containing the 28 kb HindIII genomic fragment, at high stringency conditions (lane C). The chromosomal location of the trypanosomal RNA polymerase III gene was determined using PFGE (lanes D and E). Lane D shows the ethidium bromide-stained gel; the chromosomes were blotted onto nitrocellulose filters and the blot was hybridised with the gene-internal probe B (lane E).

restriction enzymes, of pTrp28 and genomic DNA, showed that the clone had not undergone rearrangements; heterologous hybridisation experiments showed that the cross-hybridisation signal of the yeast RNA polymerase III gene is confined to the 6.0 kb EcoRI/PvuII fragment (data not shown; Figure 2).

Sequence analysis

The 6.0 kb EcoRI/PvuII fragment of pTrp28 was subcloned into the

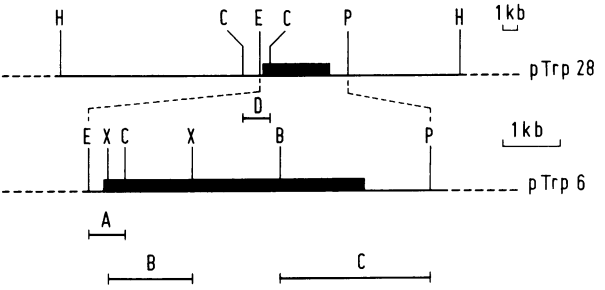

Figure 2. Restriction maps of pTrp28 and pTrp6. The coding region of RNA polymerase III largest subunit is indicated by a black box. Probes originating from the genomic clone are indicated below the map and are marked by capitals A-C. Vector DNA is indicated by dashed lines. Only those sites necessary to interpret the data are presented. Abbreviations of restriction enzyme sites: B = BamHI, C = ClaI, E = EcoRI, H = HindIII, P = PvuII and X = XmnI.

pGEM-3 vector, resulting in plasmid pTrp6 (Figure 2), which was used for further analysis. A continuous 5245 bp stretch of DNA downstream of the EcoRI site in pTrp6 was sequenced by subcloning restriction fragments into pEMBL 8/9, and sequencing both strands by the dideoxy method of Sanger et al. (42). This sequence revealed a single open reading frame extending from nucleotide 271 (amino acid 1) to nucleotide 4861 (amino acid 1530) (Figure 3).

A comparison of the deduced amino acid sequence with all other eukaryotic RNA polymerase largest subunit sequences reveals an extensive conservation of residues. Eight homology blocks named A to H are evident; these have been found to be a general characteristic of all largest subunits of eukaryotic RNA polymerases (Figure 4, see also below). The presence of these homology blocks indicates that pTrp6 carries the coding region of an RNA polymerase largest subunit gene. Three lines of evidence identify the Trp6 coding region as the gene most closely resembling the largest subunit of RNA polymerase III. First, the highest degree of sequence conservation is found with yeast RNA polymerase III, as shown by direct comparison using dot matrix analyses (Figure 5, panel C versus A, B and D). Second, and more importantly, the gene lacks the carboxy-

terminal extension which is uniquely present in the largest subunit of RNA polymerase II, as has been observed in yeast (44,45), mouse (46), hamster (47,48), *Drosophila* (48,49) and in a modified form also in *T.brucei* (20). Third, the gene lacks the two hydrophylic domains of about a hundred residues uniquely present in the N- and C-terminal region of yeast RNA polymerase I (22). Moreover, the lowest degree of conservation is found between yeast RNA polymerase I and the pTrp6 coding region: only the highly conserved motifs within the general homology blocks are present in the yeast RNA polymerase I (Figure 4 and 5).

These results strongly indicate that the identified open reading frame encodes the largest subunit of RNA polymerase III in *T.brucei*. The predicted molecular weight of this subunit, based on the sequence data, is 170.290 daltons which is within the size range of other eukaryotic RNA polymerase III largest subunits (reviewed in 1,50).

Conserved motifs in the largest subunit of RNA polymerase III of *T.brucei*

Our comparative analysis shows that the trypanosomal RNA polymerase III shares eight homology blocks with genes encoding the largest subunit of other eukaryotic RNA polymerases (Figure 4; 2,44; Jokerst, Weeks, Zehring and Greenleaf, personal communication).

The main characteristic of the amino-terminal block A is the presence of regularly spaced cysteine and histidine residues; the intermittent amino acid residues are less well conserved. Homology block B is characterised by a high proline content in the first half of the domain, followed by paired acidic residues in the second half of this domain. Block B of yeast RNA polymerase I contains an insertion of 18 amino acid residues between the two halves of this domain.

The subsequent five homology blocks, with the exception of block E, all cover a relatively large fragment of the protein. Although the trypanosomal RNA polymerase III domains C-G show a high degree of conservation with the homologous domains in yeast (I-III), *Drosophila* (II), mouse (II) and trypanosomes (II), the highest degree of conservation is observed in a direct comparison with yeast RNA polymerase III. Each of these blocks is characterised by the presence of higly conserved sequences.

```
                                                                                                                    120
GAATTCTTGTCCAGTCTGCTTATTTTTCCTCCTTTATTTCTCCCCCTTGTGTGGGGTATAGTTACACAACCCTTCCTCTTTTTGTCTTCTTATGCCTGCACGTCGGCACAACCATTGTTG
                                                                                                                    240
AGTATTGTCTCGTTGCTGTACTCCTCTATGGCTTACTTGCCTGTGATATCAAGTACCTTCATAGCGGTTGTCTCCCCTCTGTCGAAGGAAGGAGAAAAAAGAGGGGTAACATTACACATT
                                                                                                                    360
ACCGGGAATATTCACACCGTTTCTTAACCGATGCTAAAAGGAAGTAGCAGTACATCTTTCCTGCTTCCGCAGCAGTTTGTTGAACCGCTTCCTCACGCACCTGTAGAGATTAGTGCACTA
                                  M  L  K  G  S  S  S  T  S  F  L  L  P  Q  Q  F  V  E  P  L  P  H  A  P  V  E  I  S  A  L
                                                                                                                    480
CACTACGGCCTTCTTTCTCGTAACGATGTGCATCGTCTTTCCGTACTTCCATGTCGCCGTGTTGTTGGTGATGTTAAAGAGTACGGAGTCAATGACGCCCGTCTTGGCGTATGCGACCGT
 H  Y  G  L  L  S  R  N  D  V  H  R  L  S  V  L  P  C  R  R  V  V  G  D  V  K  E  Y  G  V  N  D  A  R  L  G  V  C  D  R
                                                                                                                    600
CTTTCCATCTGCGAAACTTGCGGACTCAACAGCATCGAGTGTGTTGGACACCCGGGTCACATTGACCTCGAGGCACCAGTTTTTCATCTTGGTTTCTTCACCACAGTGTTGCGAATTTGC
 L  S  I  C  E  T  C  G  L  N  S  I  E  C  V  G  H  P  G  H  I  D  L  E  A  P  V  F  H  L  G  F  F  T  T  V  L  R  I  C
                                                                                                                    720
CGTACTATATGCAAACGCTGCTCGCATGTGTTACTAGATGATACTGAAATCGATTATTATAAGAGGCGTTTGTCGTCTTCATCCCTCGAACCGCTGCAGCGCACGATGCTCATTAAAACG
 R  T  I  C  K  R  C  S  H  V  L  L  D  D  T  E  I  D  Y  Y  K  R  R  L  S  S  S  S  L  E  P  L  Q  R  T  M  L  I  K  T
                                                                                                                    840
ATACAAACGGATGCATATAAGACACGTGTGTGCCTTAAGTGTGGCGGTCTGAACGGCGTAGTACGGCGTGTGCGACCAATGAGGCTTGTTCACGAAAAATATCACGTGGAGCCTCGCCGT
 I  Q  T  D  A  Y  K  T  R  V  C  L  K  C  G  G  L  N  G  V  V  R  R  V  R  P  M  R  L  V  H  E  K  Y  H  V  E  P  R  R
                                                                                                                    960
GGTGAGGGGCCTAGGGAGAACCCTGGTGGTTTCTTTGACGCTGAGCTGCGAACTGCATGTGCTTACAACAAGGTCGTCGGCGAGTGTAGAGAGTTCGTCCACGACTTCCTTGACCCCGTA
 G  E  G  P  R  E  N  P  G  G  F  F  D  A  E  L  R  T  A  C  A  Y  N  K  V  V  G  E  C  R  E  F  V  H  D  F  L  D  P  V
                                                                                                                   1080
CGTGTCCGGCAGCTATTTCTTGCGGTTCCACCTGGGGAAGTCATACTGTTAGGCCTAGCTCCTGGAGTTAGTCCTACGGATCTACTCATGACAACACTTCTTGTGCCTCCCGTGCCCGTG
 R  V  R  Q  L  F  L  A  V  P  P  G  E  V  I  L  L  G  L  A  P  G  V  S  P  T  D  L  L  M  T  T  L  L  V  P  P  V  P  V
                                                                                                                   1200
CGGCCACGAGGGTGTGCGGGGACAACGACCGTTCGCGATGATGATCTCACAGCTCAGTACAACGACATCCTGGTCAGCACAGATACCATGCAGGACGGATCTCTTGATGCCACTAGGTAT
 R  P  R  G  C  A  G  T  T  T  V  R  D  D  D  L  T  A  Q  Y  N  D  I  L  V  S  T  D  T  M  Q  D  G  S  L  D  A  T  R  Y
                                                                                                                   1320
ACGGAGACGTGGGAGATGTTGCAGATGAGGGCAGCGAGGCTACTTGATTCGAGTCTTCCGGGGTTTCCACCAAATGTCCGTACCAGTGATTTGAAATCATATGCTCAAAGACTAAAGAGC
 T  E  T  W  E  M  L  Q  M  R  A  A  R  L  L  D  S  S  L  P  G  F  P  P  N  V  R  T  S  D  L  K  S  Y  A  Q  R  L  K  S
                                                                                                                   1440
AAACATGGCAGGTTCCGCTGTAACCTCAGTGGGAAACGTGTAGACTACTCGGGTCGCTCCGTCATTTCACCCGACCCAAATCTTGATGTGGATGAGCTCGCAGTGCCCCTTCATGTTGCC
 K  H  G  R  F  R  C  N  L  S  G  K  R  V  D  Y  S  G  R  S  V  I  S  P  D  P  N  L  D  V  D  E  L  A  V  P  L  H  V  A
                                                                                                                   1560
CGTGTGCTGACCTACCCTCAAAGAGTTTTCAAGGCCAACCATGAGCTAATGCGTCGACTTGTGCGCAACGGCCCGCACGTTCACCCTGGCGCCACTACCGTTTACCTGGCGCAGGAAGGC
 R  V  L  T  Y  P  Q  R  V  F  K  A  N  H  E  L  M  R  R  L  V  R  N  G  P  H  V  H  P  G  A  T  T  V  Y  L  A  Q  E  G
                                                                                                                   1680
TCAAAGAAGTCCCTCAAGAACGAGCGGGACCGACACCGCCTGGCAGCACGCCTTGCAGTTGGAGACATTGTTGAGCGACATGTGATGAATGGTGACTTGGTTTTATTTAATCGCCAACCT
 S  K  K  S  L  K  N  E  R  D  R  H  R  L  A  A  R  L  A  V  G  D  I  V  E  R  H  V  M  N  G  D  L  V  L  F  N  R  Q  P
                                                                                                                   1800
TCGCTTCATCGTGTTTCAATGATGGCCCATCGTGCTCGAGTGCTGCCTTTCCGCACGTTTCGTTTCAATGAGTGTTGTTGCGCACCTTATAATGCTGATTTCGATGGTGACGAAATGAAT
 S  L  H  R  V  S  M  M  A  H  R  A  R  V  L  P  F  R  T  F  R  F  N  E  C  C  C  A  P  Y  N  A  D  F  D  G  D  E  M  N
                                                                                                                   1920
GTGCATTTTGTTCAAACCGAAAAGGCTCGGGCAGAAGCTCTTCAACTTATGTCCACCGCGAGGAACATCATTTCAGCGAAAAATGGTGAGCCAATTATTGCGTGTACACAAGATTTTTTG
 V  H  F  V  Q  T  E  K  A  R  A  E  A  L  Q  L  M  S  T  A  R  N  I  I  S  A  K  N  G  E  P  I  I  A  C  T  Q  D  F  L
                                                                                                                   2040
GCCGCTGCATACCTTGTAACATCTCGCGATGTTTTTTTTGACCGTGGCGAGTTCTCGCAGATGGTGTCGCATTGGCTCGGCCCAGTAACGCAATTTAGACTGCCCATTCCCGCTATATTG
 A  A  A  Y  L  V  T  S  R  D  V  F  F  D  R  G  E  F  S  Q  M  V  S  H  W  L  G  P  V  T  Q  F  R  L  P  I  P  A  I  L
                                                                                                                   2160
AAACCCGTAGAGCTATGGACGGGAAAACAGTTGTTTGAACTCATTGTACGCCCCTCACCTGAGGTGGATGTACTGTTGAGTTTTGAGGCTCCAACAAAGTTCTACACGAGGAAGGGTAAG
 K  P  V  E  L  W  T  G  K  Q  L  F  E  L  I  V  R  P  S  P  E  V  D  V  L  L  S  F  E  A  P  T  K  F  Y  T  R  K  G  K
                                                                                                                   2280
CACGACTGTGCAGAGGAGGGCTATGTGGCATTCCTTGATTCCTGTTTTATATCAGGCCGCCTCGACAAGAAGTTATTAGGTGGAGGAGCAAAGGATGGACTTTTGCACGGCTGCACACC
 H  D  C  A  E  E  G  Y  V  A  F  L  D  S  C  F  I  S  G  R  L  D  K  K  L  L  G  G  G  A  K  D  G  L  F  A  R  L  H  T
                                                                                                                   2400
ATTGCTGGGGGTGGGTACACAGCGCGTGTTATGTCTCGCATTGCACAGTTCACTTCACGGTACCTAACAAATTATGGGTTTAGCCTGGGCCTCGGCGATGTTGCGCCAACCCCAGAGTTG
 I  A  G  G  G  Y  T  A  R  V  M  S  R  I  A  Q  F  T  S  R  Y  L  T  N  Y  G  F  S  L  G  L  G  D  V  A  P  T  P  E  L
                                                                                                                   2520
AACAAACAGAAAGCAGCAGTTCTAGCGCGCTCTGTAGAGGTATGTGATGGTTTAATCAAATCAGCGAAGACGGGGCGGATGATACCCCTTCCGGGTCTTACTGTCAAGCAATCACTGGAG
 N  K  Q  K  A  A  V  L  A  R  S  V  E  V  C  D  G  L  I  K  S  A  K  T  G  R  M  I  P  L  P  G  L  T  V  K  Q  S  L  E
                                                                                                                   2640
GCGCGTCTCAATACGGAGTTGTCGAAGGTGCGTGACGAGTGCGGTACCGCAGCCGTGCAAACACTTTCCATCCACAACAATACGCCACTGATCATGGTGCAATCGGGTAGTAAAGGTAGT
 A  R  L  N  T  E  L  S  K  V  R  D  E  C  G  T  A  A  V  Q  T  L  S  I  H  N  N  T  P  L  I  M  V  Q  S  G  S  K  G  S
                                                                                                                   2760
GCCCTGAACATTGCTCAAATGATGGCCTGCGTGGGCCAGCAGACTGTAAGTGGGAAGCGTATTTTGGACGCCTTTCAGGACCGCTCGCTCCCACATTTCCACCGTTTTGAAGAGGCACCT
 A  L  N  I  A  Q  M  M  A  C  V  G  Q  Q  T  V  S  G  K  R  I  L  D  A  F  Q  D  R  S  L  P  H  F  H  R  F  E  E  A  P
                                                                                                                   2880
GCAGCCCGTGGGTTTGTGGCTAACTCGTTTTACTCTGGCCTCTCCCCGACAGAGTTTTTCTTTCACACAATGGCAGGACGTGAGGGTCCCGTCGATACGGCAGTCAAGACGGCAGAAACG
 A  A  R  G  F  V  A  N  S  F  Y  S  G  L  S  P  T  E  F  F  F  H  T  M  A  G  R  E  G  I  V  D  T  A  V  K  T  A  E  T
```

```
                                                                                                                    3000
GGGTATATATATCGGCGCCTGATGAAGGCAATGGAAAACTTGAGCGTACGGTACGACGGCACGGTGCGTAACACGAAAGGAGACGTAATACAGCTAAGGTTTGGTGAAGATGGCCTTGAC
 G  Y  I  Y  R  R  L  M  K  A  M  E  N  L  S  V  R  Y  D  G  T  V  R  N  T  K  G  D  V  I  Q  L  R  F  G  E  D  G  L  D
                                                                                                                    3120
CCTCAGCTCATGGAGGGTAATAGCGGCACACCTCTGAACTTGGAGCAGGAATGGCTTAGTGTTCGAGCGGCGTATGCACGTTGGGTCGTGGGACTGCTTGCTGGAAGTAAGACGGCAAGT
 P  Q  L  M  E  G  N  S  G  T  P  L  N  L  E  Q  E  W  L  S  V  R  A  A  Y  A  R  W  V  V  G  L  L  A  G  S  K  T  A  S
                                                                                                                    3240
GACGGAAATGCAATACGAGATAATGAAAATTACTTTAACGAGTTTATTTCGATGCTTCCAACAGAGGGACCTTCGTTTGTAGAAGCGTGCTTGAATGGGGATCAAGAGGCGCTAAAAGTT
 D  G  N  A  I  R  D  N  E  N  Y  F  N  E  F  I  S  M  L  P  T  E  G  P  S  F  V  E  A  C  L  N  G  D  Q  E  A  L  K  V
                                                                                                                    3360
TGTGAGGAACAGGAGTCTCGAGAAGATGCTTTGCATAACAGCAACGGGAAAACTAATGATAGGGAGAGCAGGCCACGTACAGGAAGACTGCGTCGGGCAGTGTTAATTTCCCATCTGGTT
 C  E  E  Q  E  S  R  E  D  A  L  H  N  S  N  G  K  T  N  D  R  E  S  R  P  R  T  G  R  L  R  R  A  V  L  I  S  H  L  V
                                                                                                                    3480
AAAGTTTGCAGTAGAAAGTTCAAAGATGATATCCAAGACTTCTTCGTGAAAAAGGTGAGGGAACAGCAACGGATCCGCAACCTGCTCAACCTTCCCAACACGAGTCGTGAGCGAACGGAG
 K  V  C  S  R  K  F  K  D  D  I  Q  D  F  F  V  K  K  V  R  E  Q  Q  R  I  R  N  L  L  N  L  P  N  T  S  R  E  R  T  E
                                                                                                                    3600
GGGGGAGGTGACAATAGTGGGCCTATTGCAAATAAGAGAACCAAAAAGAGAGCACCGTCACTGAAGGTCAAAGACAGCAAAGAAGGGGGAAGGGTTTCCGAGTTGCGGGACCTTGAGATG
 G  G  G  D  N  S  G  P  I  A  N  K  R  T  K  K  R  A  P  S  L  K  V  K  D  S  K  E  G  G  R  V  S  E  L  R  D  L  E  M
                                                                                                                    3720
CTTCAGACCGAACTACTTCCGCTGACACGCGGAATGGTAACACGATTCATTGCTCAATGCGCAAGCAAGTATCTGCGTAAGGCATGTGAGCCTGGTACCCCGTGCGGCGCCATTGCTGCC
 L  Q  T  E  L  L  P  L  T  R  G  M  V  T  R  F  I  A  Q  C  A  S  K  Y  L  R  K  A  C  E  P  G  T  P  C  G  A  I  A  A
                                                                                                                    3840
CAATCTGTTGGTGAGCCAAGCACCCAGATGACTTTGCGTACCTTTCACTTTGCTGGTGTGGCGAGCATGAGCATTACGCAGGGTGTTCCTCGACTAGTGGAGGTCATCAACGCCAATCGC
 Q  S  V  G  E  P  S  T  Q  M  T  L  R  T  F  H  F  A  G  V  A  S  M  S  I  T  Q  G  V  P  R  L  V  E  V  I  N  A  N  R
                                                                                                                    3960
AATATTGCTACACCCGTTGTCACAGCTCCAGTACTACTTATGGAAGGCGAGGAAAACCATTGCGAAATATTCAGGAAGCGGGCACGGTTCGTGAAGGCCCAAATTGAGCGCGTGCTGTTA
 N  I  A  T  P  V  V  T  A  P  V  L  L  M  E  G  E  E  N  H  C  E  I  F  R  K  R  A  R  F  V  K  A  Q  I  E  R  V  L  L
                                                                                                                    4080
CGTGAGGTTGTGTCCGAAATTGTGGAGGTGTGCAGTGACACTGAATTTTACCTCCGTGTGCACCTTAATATGTCCGTTATTACAAAACTGCATCTACCCATCAACGCCATCACCGTTCGT
 R  E  V  V  S  E  I  V  E  V  C  S  D  T  E  F  Y  L  R  V  H  L  N  M  S  V  I  T  K  L  H  L  P  I  N  A  I  T  V  R
                                                                                                                    4200
CAGCGCATTCTTGCGGCCGCTGGGCATACCATGTCTCCCCTTCGGATGCTAAACGAGGATTGTATAGAGGTATTCAGCCTCGACACACTCGCTGTATATCCCCACTTTCAGGATGCCCGT
 Q  R  I  L  A  A  A  G  H  T  M  S  P  L  R  M  L  N  E  D  C  I  E  V  F  S  L  D  T  L  A  V  Y  P  H  F  Q  D  A  R
                                                                                                                    4320
TGGGTTCACTTTAGTCTCCGACGCATCCTCGGGTTACTCCCCGACGTTGTAGTTGGGGGCATCGGAGGCATTAACCGTGCGATGATATCCTCAAACGGCACTGAAGTGTTGGCAGAAGGC
 W  V  H  F  S  L  R  R  I  L  G  L  L  P  D  V  V  V  G  G  I  G  G  I  N  R  A  M  I  S  S  N  G  T  E  V  L  A  E  G
                                                                                                                    4440
GCGGAGCTGCGTGCCGTTATGAATCTTTGGGGTGTCGACTCTACTCGGGTGGTCTGTAATCACGTGGCCGTGGTGGAGCGTGTGCTGGGTATTGAAGCTGCGCGCCGTGTGATCGTGGAC
 A  E  L  R  A  V  M  N  L  W  G  V  D  S  T  R  V  V  C  N  H  V  A  V  V  E  R  V  L  G  I  E  A  A  R  R  V  I  V  D
                                                                                                                    4560
GAGATACAAAACATACTCAAGGCATACAGTCTAAGCATTGACGTTCGGCACGTGTACCTCCTTGCGGATCTCATGACTCAACGAGGTGTGGTGCTTGGAATTACGCGATACGGCATTCAA
 E  I  Q  N  I  L  K  A  Y  S  L  S  I  D  V  R  H  V  Y  L  L  A  D  L  M  T  Q  R  G  V  V  L  G  I  T  R  Y  G  I  Q
                                                                                                                    4680
AAAATGAACTTCAACGTTCTTACCATGGCTTCATTTGAGCGCACCACGGACCATTTGTACAATGCCGCAGCCACTCAACGAGTTGACCGGGATCTCAGTGTGTCAGACAGTATTATCGTA
 K  M  N  F  N  V  L  T  M  A  S  F  E  R  T  T  D  H  L  Y  N  A  A  A  T  Q  R  V  D  R  D  L  S  V  S  D  S  I  I  V
                                                                                                                    4800
GGGAAACCTGTCCCATTAGGGACAACATCCTTCGATCTCTTGTTAGACGGTTCCATAAGCAACGATATTCTCCCTCCTCAGCGTTGCGTGAAACGCGGGATGGGCCCCAACTTCCACACC
 G  K  P  V  P  L  G  T  T  S  F  D  L  L  L  D  G  S  I  S  N  D  I  L  P  P  Q  R  C  V  K  R  G  M  G  P  N  F  H  T
                                                                                                                    4920
GCAAAACGACATCATCTGGTACCTCTTGCCGCTGAGGGTGTTTTTCGTTTGGATCTTTTTTAAACGTTTTCAAGGAACGAAATGGCAATAACGGTGATGATGACGGAAACAAAATAAGTA
 A  K  R  H  H  L  V  P  L  A  A  E  G  V  F  R  L  D  L  F  *
                                                                                                                    5040
AAAGGTTGAAGTTATGTGTGGGCAGGTGATTGTGGTCGCGCGCATATGGGCGAGTACTTTGCCTAATGTTAGTATCTTCTTGAGGGATGAGAGGGGAGAATGAAGCTGTGTATTTCTTTA
                                                                                                                    5160
CACACACAAGCCAGTGGAGTTGTGTGACTGATGGTGGTGCTTGTAAGGTGCCAGTTGCATGGGCTGGCGTTGGAAACGTGTTGAATATTACGTGTAACGGTGCTTAATTGCCATGCTGAA

GCATGTGGCGTATCCAAAGGAAGATAGTTGGGGAAGTGACGGGAGTGTGGGGAAAATGTTTTACTTTATTTTCTGCTCCTGGTAC
```

Figure 3. Nucleotide sequence of the trypanosomal RNA polymerase III largest subunit gene. The predicted amino acid sequence is displayed beneath the nucleotide sequence. The determined nucleotide sequence of 5245 bp starts at the EcoRI restriction site indicated in Figure 2.

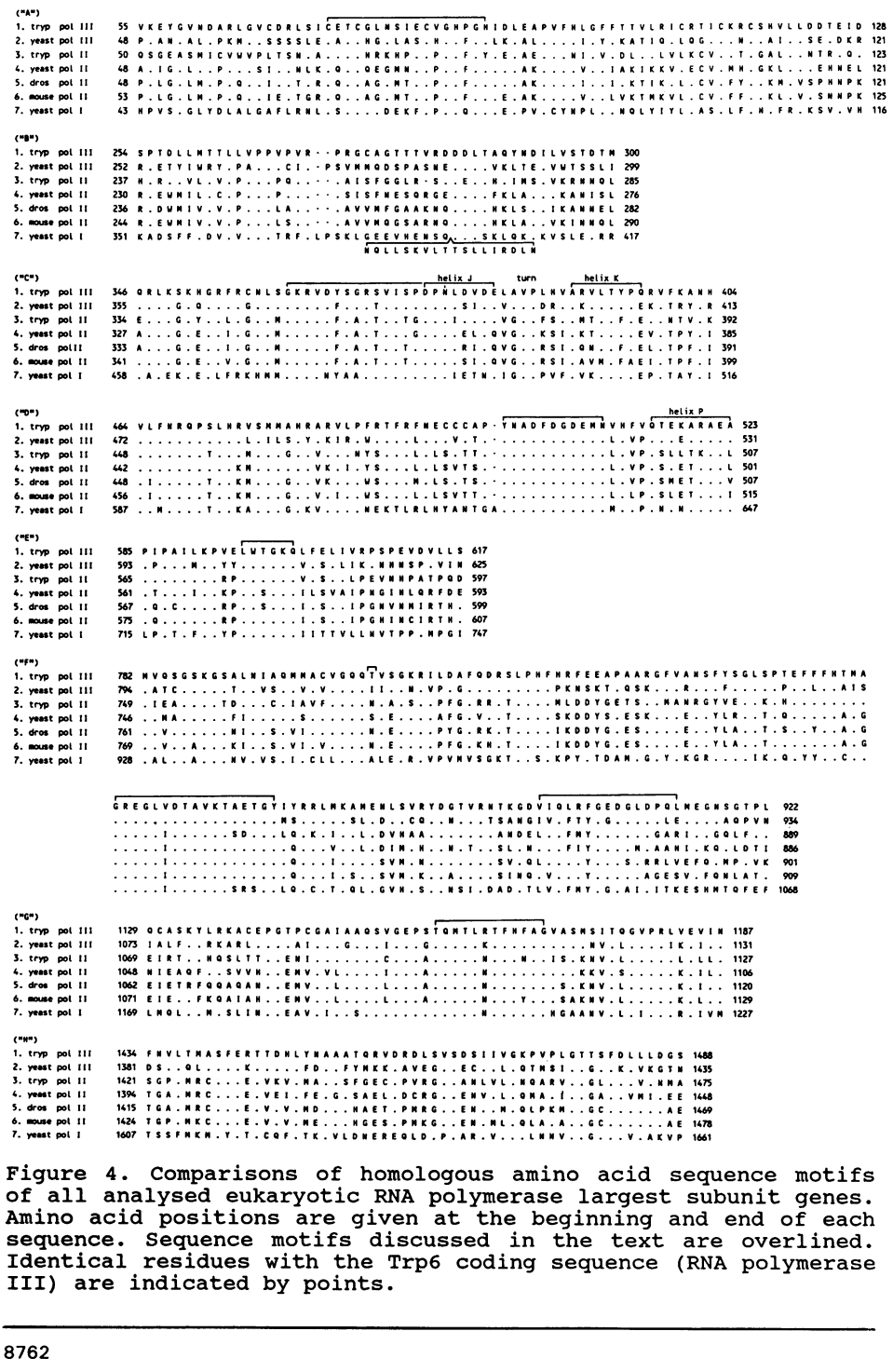

Figure 4. Comparisons of homologous amino acid sequence motifs of all analysed eukaryotic RNA polymerase largest subunit genes. Amino acid positions are given at the beginning and end of each sequence. Sequence motifs discussed in the text are overlined. Identical residues with the Trp6 coding sequence (RNA polymerase III) are indicated by points.

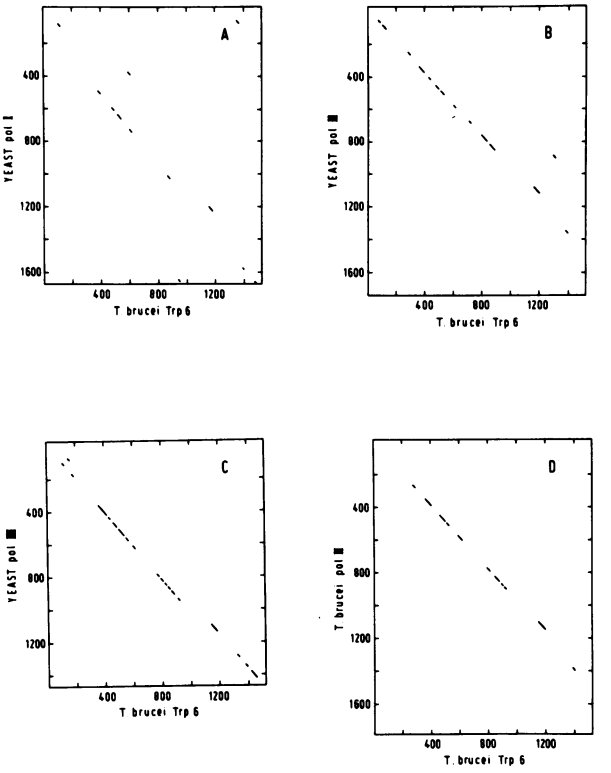

Figure 5. Dot matrix comparison of the largest subunits of eukaryotic RNA polymerases. The Trp6 coding region was compared with the amino acid sequences of the largest subunits of yeast RNA polymerase I (panel A; (22)), II (panel B; (44)) and III (panel C; (44)) and of trypanosomal RNA polymerase II (panel D; (20)). Dots were generated in the matrix whenever four residues in a stretch of six amino acids were identical between the compared sequences. A similar degree of conservation (panels B and D) was observed when the trypanosomal RNA polymerase III gene was compared with the largest subunit genes of RNA polymerase II of mouse (55) and *Drosophila* (kindly provided by Dr. A. Greenleaf).

Block C is centred around the consensus sequence GKRVdfsaRtVIspDPN. Block D is characterised by the motif YNADFDGDEMN, this motif is the longest conserved region which is universally present in all largest subunits (2). Homology block E covers only about 30 amino acid residues and is centred around the motif LWtGKQ. The longest conserved sequence motifs in block F and G shared between the trypanosomal RNA polymerase III and

the other largest subunits are, respectively, GREGLiD'AVKTaetGY and TQMTLnTFHfAG. The carboxy-terminal block H shows the lowest degree of conservation; only short amino acid stretches can be found that are common to all polypeptides.

Copy number and chromosomal location

Quantitative hybridisation experiments have demonstrated that all analysed RNA polymerase subunit genes are present as a single copy within eukaryotic genomes (reviewed in 2). One striking exception is the RNA polymerase II largest subunit gene from African trypanosomes, here two almost identical genes have been identified (20). To determine whether the gene encoding RNA polymerase III is present in one or more copies, we prepared a Southern blot from genomic DNA of *T.brucei* digested with eleven different restriction endonucleases; the blot was subsequently hybridised with probe B, spanning the 5' coding region (Figure 2). All but one of the endonucleases resulted in single hybridising fragments. Only MluI generated two hybridising fragments. Since trypanosomes are diploid organisms (51-53) and none of the restriction enzymes used cut within the probe (B), we can conclude that the RNA polymerase III gene is a single copy gene present in the two allelic forms, which can be discriminated by the MluI restriction site (data not shown).

To assign the chromosomal localisation of the trypanosomal RNA polymerase III largest subunit gene and to determine whether this gene is linked to the RNA polymerase II genes, we prepared blots of *T.brucei* karyotypes. Trypanosomal chromosomes were fractionated by pulsed field gradient gel electrophoresis (PFGE, (31,40)) and blotted to nitrocellulose. These blots were hybridised with nick-translated probes specific for the largest subunit of RNA polymerase III. The polymerase III-specific probe hybridised to material remaining in the slot (Figure 1, lanes D and E). We have previously shown that the RNA polymerase II-specific probes hybridise to the large resolved chromosomes (2 Mb region; (20)).

In summary our results strongly suggest that the trypanosomal gene encoding the largest subunit of RNA polymerase III is a single copy gene which is not physically linked to the two RNA polymerase II largest subunit genes.

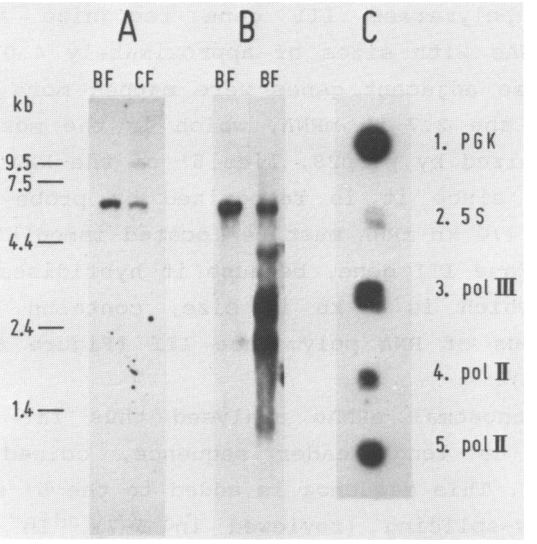

Figure 6. Northern blot analysis of pTrp28. About 6 ug poly $(A)^+$ RNA from bloodstream form (BF) and culture form (CF) trypanosomes were separated in a 1% formaldehyde agarose gel. RNA was transferred to nitrocellulose, and blots were hybridised with the gene-internal probe B (panel A; panel B lane 1) and with the 28 kb insert of pTrp 28 (panel B, lane 2).
Trypanosomal RNA polymerase mRNAs contain a mini-exon (panel C). Fragments of the clones indicated were used for a sandwich dot-blot hybridisation (36). Dotted fragments, approximately 200 ng DNA/fragment, were used to select their homologous mRNA, and the hybrid was then hybridised to a labeled mini-exon probe. DNAs spotted were: **1**, 4 kb HindIII fragment from clone pTgPGK-2 (68); **2**, 0.7 kb PstI fragment from clone p5S-2 (Kooter and Borst, unpublished and (69)); **3**, probe A (Figure 2); **4**, 0.5 kb HincII fragment from clone pTrp4.8 and **5**, 0.6 kb HincII fragment from clone pTrp5.9 representing the 5'ends of the two RNA polymerase II genes (Evers et al., unpublished and (20)).

Northern analysis of pTrp28

As is obvious from the map provided in Figure 2, we cloned sequences in addition to the coding region of the largest subunit gene of RNA polymerase III, approximately 14 kb at the 5' end and 9 kb at the 3'end. To characterise these areas in more detail, we isolated poly$(A)^+$ RNA from bloodstream form and procyclic trypanosomes, prepared northern blots and hybridised two identical strips with either the complete insert of pTrp28 or the gene-internal probe B. The gene-internal probe hybridises to a 5.5 kb mRNA (Figure 6, panel A); the DNA sequences adjacent

to the RNA polymerase III gene recognize at least four additional mRNAs with sizes of approximately 4.0, 2.7, 2.0 and 1.0 kb. The two adjacent genes were mapped more precisely. The gene encoding the 2.7 kb mRNA, which is the most dominant RNA species recognized by pTrp28, lies 5' of the RNA polymerase III coding region since it is recognized by probe D. The coding region of the 4.0 kb mRNA must be located immediately downstream of the polymerase III gene, because it hybridises with probe C. This probe, which is 3 kb in size, contains 1.5 kb of the carboxy-terminus of RNA polymerase III (Figure 6, panel B and data not shown).

All trypanosomal mRNAs analysed thus far start with an identical 39 nt long leader sequence, coined mini-exon or spliced leader. This sequence is added to the 5' end of the main exon by *trans*-splicing (reviewed in 5-7). In order to test whether the mRNAs of the RNA polymerases carry the mini-exon sequence, we performed sandwich dot-blot hybridisations (36). In this experiment total RNA was first hybridised with polymerase gene-specific probes and the resulting hybrid was probed with a kinased mini-exon probe. Gene-specific fragments of phosphoglycerate kinase (PGK) and 5S RNA served as, respectively, the positive and negative controls. No hybridisation could be detected to the 5S RNA genes. The PGK and the polymerase-specific fragments resulted in positive signals indicating that both RNA polymerase II and III mRNAs carry a mini-exon (Figure 6, panel C).

DISCUSSION

RNA polymerases in *T.brucei*

Separation of different RNA polymerase fractions by conventional chromatography (10-13) has resulted in the identification of an alpha-amanitin sensitive and an alpha-amanitin resistant RNA polymerase activity in *T.brucei*. We have previously reported the isolation of the genes encoding the largest subunit of RNA polymerase I and II (20). The experimental findings presented in this paper show the identification of the largest subunit of RNA polymerase III. These experiments thus indicate that *T.brucei* possesses the normal eukaryotic complement of RNA polymerases I,

II and III, and contrast with the previously reported chromatographical data. Our findings are also consistent with the reports of Kooter and Borst (8) and Laird et al. (9), which indicated the presence of RNA polymerase I, II and III, and possibly a fourth RNA polymerase activity (see also 2), based on template specificity and sensitivity to alpha-amanitin in *in vitro* run-on assays.

Our northern analysis shows that the relative abundance of the RNA polymerase III mRNA is equal in both procyclic- and bloodstream trypanosomes (Figure 6, panel A and unpublished data). This implies that the expression of the RNA polymerase III gene is not developmentally regulated within the life cycle of *T.brucei*, as we observed previously for both trypanosomal RNA polymerase II genes (20). Tittawella (13), however, provided some evidence for a developmental regulation of the RNA polymerase enzyme levels. If this observation turns out to be correct, it would imply that stage-specific regulation of RNA polymerase levels occurs at the translational level or by enzyme (in)activation (cf. 54).

Subunit structure and functional significance

Various groups have reported the structure and sequence of genes encoding the largest subunits of the different eukaryotic RNA polymerases: I (yeast, (22)), II (yeast, (44); mouse, (55); *Drosophila*, (24) and Jokerst et al., personal communication) and III (yeast, (44)). A comparative analysis of the sequences encoding the largest subunits of RNA polymerase I, II and III from evolutionarily distinct organisms shows two characteristic features among all these polypeptides. First, eight homology regions are conserved. Second, class-specific additional domains are present in RNA polymerase I and II.

The molecular weight of the largest subunits of RNA polymerase I and II are significantly higher than that of RNA polymerase III (1,50). As is evident from the sequence analysis, the higher molecular weight of RNA polymerase II is essentially due to the presence of a C-terminal extension. The C-terminal extension is characterised by a very unusual repeat structure, consisting of 26-52 tandem repeats of the heptapeptide YSPTSPS (44-46, 48, 55, 56, Jokerst et al., personal communication).

Since the repeat structure was found to be present in yeast, *Drosophila*, mouse and hamster it has been postulated to be a general feature of eukaryotes. In contrast with this, we have shown that the trypanosomal RNA polymerase II subunit lacks this repeat structure (20), although a C-terminal extension that is rich in serine and tyrosine residues is present. The larger molecular weight of RNA polymerase I originates from two additional domains between homology blocks A and B and between blocks G and H, each consisting of approximately 100 mainly hydrophylic residues (22). Analogous domains are absent in the RNA polymerase II and III largest subunits.

The highly conserved domains contain sequence motifs which probably determine general functions of the largest subunits. For example, the arrangement of cysteine and histidine residues in the N-terminal block A is identical to the "Zinc finger" motif of the C_2H_2 class, Cys-X_2-Cys-X_9-His-X_2-His (57). This "Zinc finger" is characteristic for trancription factors and other DNA-binding proteins (57-59). It is worth noting here that, in contrast to the latter proteins, the "Zinc finger" is not repeated in the RNA polymerase genes. It is rather striking that the trypanosomal largest subunit genes encoding RNA polymerase II lacks the conserved C_2H_2 motif, since the second histidine residue is substituted by tyrosine. However, a "Zinc finger" of the C_X class (57) lies immediately 3' to the mutated "Zinc finger" and most likely substitutes its function. "Zinc finger" motifs have also been observed in the second largest subunit gene of yeast and *Drosophila* RNA polymerase II (2, 60, 61). Since RNA polymerases are known to contain tightly bound Zn atoms, which is a prerequisite for the formation of a functional enzyme (reviewed in 62), this motif is very likely to be the metal chelating domain.

Blocks C and D contain sequences that possess homology to the helix-turn-helix motif in the DNA binding domain of *E.coli* DNA polymerase I and T7 DNA polymerase (44). Two other conserved domains which might be involved in DNA template binding were identified: a 28 amino acid stretch containing a high percentage of alternating basic and hydrophobic residues preceeding the helix-turn-helix motif in block C, and the 3' end of domain F,

in which a stretch of 15 amino acid residues shows 60% homology with the 3' end of a DNA binding protein of SV40 (63).

In block D another striking motif was identified, the motif YNADFDGDEMN (2). This motif, with unknown function, is the longest absolutely conserved domain universally present in all largest subunits. The same sequence is found in archaebacterial RNA polymerases (Zillig, personal communication) and, when restricted to the heptapeptide NADFDGD, also in *E.coli* (64), vaccinia virus (65) and chloroplast (66) RNA polymerases. The strict conservation of this motif among all RNA polymerase subunits, and the absence of this motif in any of the sequences contained in the NBRF protein sequence data bank (release 13), indicate that the conservation of this motif in RNA polymerases is the result of stringent functional constraint.

Sequence analysis of blocks E-H did not reveal any known functional properties. Block F of RNA polymerase II, however, might contain the binding site for the RNA polymerase inhibitor alpha-amanitin, since the mutation conferring resistance to this toxin has been mapped within this domain. In mouse a single point mutation, giving rise to an asparagine-to-aspartic acid substitution, resulted in a mutant gene resistant to alpha-amanitin (67).

Another striking feature we observed was the even spacing of the regions separating the homology blocks, except for a marked difference in length of the amino acid stretches separating blocks F and G. Moreover, the stretches seperating F and G are characterised by the lowest overall homology between the individual subunits. This is exemplified most drastically between mouse and *Drosophila* RNA polymerase II where this represents the only major difference between these otherwise very homologous subunits.

In summary, we have identified and cloned the largest subunit gene of RNA polymerase III of *T.brucei*. The gene is characterised by eight homology regions, which are shared with other eukaryotic largest subunit genes. The rather strict conservation of these domains, some of them with known functional motifs, suggest that they reflect functional entities adding general properties, such as DNA template binding and

catalytic functions, to the enzyme core. Interactions of the largest subunits with the smaller subunits of the RNA polymerase complex are, most likely, mediated by more complex structural domains, which are not identified by simple sequence comparisons. However, our sequence analysis has shown that the absence or presence of hydrophylic insertions or the C-terminal extension determines the RNA polymerase class among eukaryotes. These domains might, therefore, be responsible for directing gene-specific transcription, either directly or indirectly trough interactions with the class-specific smaller subunits of the RNA polymerase complex.

ACKNOWLEDGEMENTS
We wish to thank Dr. Arno Greenleaf (Duke University, USA) and Dr. André Sentenac (C.E.N.S., Paris, France) for providing, respectively, the *Drosophila* and yeast probes. We thank Dr. Carol Gibbs for valuable comments on the manuscript, Mr. Klaus Lamberty for artwork and Ms. Carmen Müller for photography and Ms. Brigitte Wall for preparing the manuscript. The work was supported by Bundesministerium für Forschung und Technologie grant 0318885 and an research award of the Erna and Victor Hasselblad Foundation to Albert W.C.A. Cornelissen.

*To whom correspondence should be addressed

References

1. Sentenac, A. (1985) *CRC Crit. Rev. Biochem.* 18, 31-90.
2. Cornelissen, A.W.C.A., Evers, R and Köck. J. (1988) *Oxford Surveys Euk. Gen.* 5, in press.
3. Dynan, W.S. and Tjian, R. (1985) *Nature* 316, 774-778.
4. Struhl, K. (1987) *Cell* 49, 295-297.
5. Boothroyd, J.C. (1985) *Ann. Rev. Microbiol.* 39, 475-502.
6. Borst, P. (1986) *Ann. Rev. Biochem.* 55, 701-732.
7. Van der Ploeg, L.H.T. (1986) *Cell* 47, 479-480.
8. Kooter, J.M., and Borst, P. (1984) *Nucl. Acids Res.* 12, 9457-9472.
9. Laird, P.W., Kooter, J.M., Loosbroek, N., and Borst, P. (1985) *Nucl. Acids Res.* 13, 4253-4266.
10. Kitchin, P.A., Ryley, J.F. and Gutteridge, W.E. (1984) *Comp. Biochem.Physiol.* 77B, 223-231.
11. Earnshaw, D.L., Beebee, T.J.C. and Gutteridge, W.E. (1985) *Biochem. Biophys. Res. Comm.* 131, 844-848.
12. Earnshaw, D.L., Beebee, T.J.C. and Gutteridge, W.E. (1987) *Biochem. J.* 241, 649-655.
13. Tittawella, I. (1988) *FEBS Lett.* 227, 122-126.
14. 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.
15. 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.
16. Murphy, N.B., Guyaux, M., Pays, E. and Steinert, M. (1987) In: *Molecular Strategies of Parasitic Invasion* (Agabian, N., Goodman, H. and Nogueira, N. eds.), Alan R. Liss Inc., New York, 449-469.
17. Shea, C., Gwo-Shu Lee, M., and Van der Ploeg, L.H.T. (1987) *Cell* 50, 603-612.
18. Alexandre, S., Guyaux, M., Murphy, N.B., Coquelet, H., Pays, A., Steinert, M. and Pays, E. (1988) *Mol. Cell. Biol.* 8, 2367-2378.
19. Shea, C. and Van der Ploeg, L.H.T. (1988) *Mol. Cell. Biol.* 8, 854-859.
20. Evers, R., Hammer, A., Köck, J., Jess, W., Borst, P., Mémet, S. and Cornelissen, A.W.C.A. (1988) *submitted.*
21. Queen, C. and Kan L.J. (1984) *Nucl. Acids Res.* 12, 581-599.
22. Mémet, S., Gouy, M., Marck, C., Sentenac, A. and Buhler, J.-M. (1988) *J. Biol. Chem.* 263, 2830-2839.
23. Riva, M., Mémet, S., Micouin, J.-Y., Huet, J., Treich, I., Dassa, J., Young, R., Buhler, J.-M., Sentenac, A. and Fromageot, P. (1986) *Proc. Natl. Acad.Sci. USA.* 83, 1554-1558.
24. Biggs, J., Searles, L. L. and Greenleaf, A. L. (1985) *Cell* 42, 611-621.
25. Lanham, S.M. and Godfrey, D.G. (1970) *Exp. Parasitol.* 28, 521-534.
26. Brun, R. and Schönberger, M. (1979) *Acta Trop.* 36, 289-292.
27. Auffray, C., and Rougeon, F. (1980) *Eur. J. Biochem.* 107, 303-314.
28. Hoeijmakers, J.H.J., Borst, P., Van den Burg, J., Weissmann, C., and Cross, G.A.M. (1980) *Gene* 8, 391-417.
29. Van der Ploeg, L.H.T., Liu, A.Y.C., Michels, P.A.M., De Lange, T., Borst, P. Majumder, H.K., Weber, H. and van Boom, J.H. (1982) *Nucl. Acids Res.* 10, 3591-3604.
30. Birnboim, H.C. and Doly, J. (1979) *Nucl. Acids Res.* 7, 1515-1523.
31. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.
32. Bernards, A., Van der Ploeg, L.H.T., Frasch, A.C.C., Borst, P., Boothroyd, J.C., Coleman, S., and Cross, G.A.M. (1981) *Cell* 27, 497-505.
33. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular cloning.* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).
34. Jeffreys, A.J., and Flavell, R.A. (1977) *Cell* 12, 1097-1108.
35. Wahl, G.M., Stern, M. and Stark, P.G.R. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3683-3687.
36. Guyaux, M., Cornelissen, A.W.C.A., Pays, E., Steinert, M. and Borst, P. (1985) *EMBO J.* 4, 995-998.
37. De Lange, T., Liu, A.Y.C., Van der Ploeg, L.H.T., Borst, P., Tromp, M.C., Van Boom, J.H. (1983) *Cell* 34, 891-900.
38. Kooter , J.M., Winter, A.J., d'Oliveira, C., Wagter, R. and Borst, P. (1988) *Gene*, in press.
39. Van der Ploeg, L.H.T., Schwartz, D.C., Cantor, C.R. and Borst, P. (1984) *Cell* 37, 77-84.
40. Johnson, P.J. and Borst, P. (1986) *Gene* 43, 213-220.
41. Messing, J. (1983) *Meth. Enzymol.* 101, 20-89.
42. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl.*

Acad. Sci. USA. 74, 5463-5467.
43. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) *Proc. Natl. Acad. Sci. USA* 80, 3963-3965.
44. Allison, L. A., Moyle, M., Shales, M. and Ingles, C. J. (1985) *Cell* 42, 599-610.
45. Nonet, M., Sweetser, D. and Young, R.A. (1987) *Cell* 50, 909-915.
46. Corden, J. L., Cadena, D. L., Ahearn, J. M. and Dahmus, M. E. (1985) *Proc. Natl. Acad. Sci. USA* 82, 7934-7938
47. Ingles, C. J., Moyle, M., Allison, L. A., Wong, J. K.-C., Archambault, J. and Friesen, J. D. (1987) *UCLA Symp. Mol. Cell. Biol.* 52, 383-393.
48. Allison, L.A., Wong, J.K.-C., Fitzpatrick, V.D., Moyle, M. and Ingles, C.J. (1988) *Mol. Cell. Biol.* 8, 321-329.
49. Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S., and Greenleaf, A.L.(1988) *Proc. Natl. Acad. Sci. USA* 85,3698-3702.
50. Paule, M.R. (1981) *Trends Biochem. Sci.* 5, 128-131.
51. Tait, A. (1980) *Nature* 287, 536-538.
52. Borst, P., Van der Ploeg, M., Van Hoek, J.F.M., Tas, J. and James, J. (1982) *Molec. Biochem. Parasitol.* 6, 13-23.
53. Gibson, W.C., Osinga, K.A., Michels, P.A.M. and Borst, P. (1985) *Molec. Biochem. Parasitol.* 16, 231-242.
54. Bateman, E. and Paule, M.R. (1986) *Cell* 47, 445-450.
55. Ahearn J. M., Bartolomei M. S., West, M. L., Cisek, L. J. and Corden, J. L. (1987) *J. Biol. Chem.* 262, 10695-10705.
56. Bartolomei, M. S., Halden, N. F., Cullen, C. R. and Corden, J. L. (1988) *Mol. Cell. Biol.* 8, 330-339.
57. Evans, R.M. and Hollenberg, S.M. (1988) *Cell* 52, 1-3.
58. Berg, J.M. (1986) *Science* 232, 485-487.
59. Klug, A. and Rhodes, D. (1987) *Trends Biochem. Sci.* 12, 464-469.
60. Sweetser, D. Nonet, M. and Young, R.A. (1987) *Proc. Natl. Acad. Sci. USA* 84, 1192-1196.
61. Falkenburg, D., Dworniczak, B., Faust, D.M. and Bautz, E.K.F. (1987) *J. Mol. Biol.* 195, 929-937.
62. Wu, F.Y.-H. and Wu, C.-W. (1981) *Adv. Organic Chem.* 3, 143-166.
63. Jay, G., Nomura, S., Anderson, C.W. and Khoury, G. (1981) *Nature* 291, 346-349.
64. Ovchinnikov, Y.A., Monastryrskaya, G.S., Gubanov, V.V., Guryev, S.O., Salomatina, I.S., Shuvaea, T.M., Lipkin, V.M. and Sverdlov, E.D. (1982) *Nucl. Acids Res.* 10, 4035-4044.
65. Broyles, S.S. and Moss, B. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3141-3145.
66. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aoto, S.-i., Inokuchi, H. and Ozeki, H. (1986) *Nature* 322, 572-574.
67. Bartolomei, M.S. and Corden, J.L. (1987) *Mol. Cell. Biol.* 7, 586-594.
68. Osinga, K.A., Swinkels, B.W., Gibson, W.C., Borst, P., Veeneman, G.H., van Boom, J.H., Michels, P.A.M. and Opperdoes, F.R. (1985) *EMBO J.* 4, 3811-3817.
69. Lenardo,M.J., Dorfman, D.M., Reddy, L.V. and Donelson, J.E. (1985) *Gene* 35, 131-141.
70. Rigby, P.W., Dieckmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.