Volume 249, number 1, 123–128 FEB 07143 May 1989

Complete sequence of the gene encoding the largest subunit of RNA polymerase I of *Trypanosoma brucei*

Waldemar Jess, Andrea Hammer and Albert W.C.A. Cornelissen

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

Received 20 March 1989

We have set out to clone the trypanosomal gene encoding the largest subunit of RNA polymerase I. We screened a genomic library with a synthetic oligonucleotide probe encoding an eleven amino acid sequence motif, YNADFDGDEMN, which has been found in all eukaryotic RNA polymerase largest subunit genes analyzed so far. We isolated the Trp11 locus and determined the complete sequence of the gene encoded within this locus. The deduced amino acid sequence contains the highly conserved RNA polymerase domains as well as the previously identified RNA polymerase I-specific hydrophilic insertions. Therefore, the gene most closely resembles the largest subunit of RNA polymerase I.

RNA polymerase I; Largest subunit; Nucleotide sequence; Sequence homology; (*Trypanosoma brucei*)

1. INTRODUCTION

Eukaryotic RNA polymerases are multi-subunit enzymes responsible for transcription. Three classes of RNA polymerases have been described, which are named I(or A), II(or B) and III(or C). The nucleolar enzyme RNA polymerase I transcribes the ribosomal RNA genes, II is responsible for the transcription of all mRNA precursors and III transcribes the small RNA genes, encoding e.g. 5 S and tRNAs (reviewed in [1,2]).

Protein-coding genes are transcribed by RNA polymerase II which is, in most eukaryotes, characterized by its high sensitivity to the toxin α-amanitin [3]. This contrasts strongly with the transcription of the VSG and the physically linked ESAG genes of the African trypanosome, *Trypanosoma brucei* [4–7]. Transcription of these two gene classes is completely insensitive to α-amanitin [5]. Other trypanosomal protein-coding genes are, however, transcribed by an RNA polymerase II which is sensitive to α-amanitin [8,9].

To explain the unusual transcription of the VSG and ESAG genes, two alternative hypotheses have been proposed. Firstly, the eukaryotic RNA polymerase resistant to α-amanitin, RNA polymerase I, might transcribe these genes despite their protein-coding nature. This hypothesis is supported by the observation that the putative promoter region of two VSG transcription units shows homology with the consensus sequence of the eukaryotic RNA polymerase I promoter [10,11]. Secondly, the VSG transcription unit might be transcribed by a modified RNA polymerase II, which has gained resistance towards the toxin α-amanitin. The latter hypothesis is supported by the fact that *T.brucei* contains two slightly different copies of the Pol II gene [12]. In all other eukaryotic species analyzed so far, Pol II is encoded by a unique gene. Moreover, genetic analysis in higher eukaryotes has demonstrated that mutations conferring α-amanitin resistance to RNA polymerase II have all occurred in the largest subunit of RNA polymerase II [1,2]. Therefore, the presence of the second copy in trypanosomes might generate a modified RNA polymerase II, transcribing VSG and ESAG genes.

To obtain additional information on the RNA

Correspondence address: A.W.C.A. Cornelissen, Max-Planck-Institut für Biologie, Molecular Parasitology Unit, Spemannstrasse 34, 7400 Tübingen, FRG

Abbreviations: VSG, variant surface glycoprotein; ESAG, expression site-associated gene; Pol I, Pol II and Pol III indicate the largest subunit of the respective RNA polymerase

Published by Elsevier Science Publishers B. V. (Biomedical Division)
00145793/89/$3.50

polymerase transcribing the VSG transcription unit, we wanted to extend our analysis of trypanosomal RNA polymerases [12,16] to Pol I. We have previously reported the isolation of the locus, Trp11, most likely encoding Pol I [12]. In this study we report the isolation and the complete sequence of the gene encoding Pol I. Our aim is to use this gene as a tool to examine the transcription of VSG genes in further detail, allowing us to critically evaluate the two alternative hypotheses described above.

2. MATERIALS AND METHODS

2.1. *Constructing and screening of the EMBL3 library*

The genomic library of *T.brucei* in EMBL3 was prepared and screened with a synthetic oligonucleotide probe 5'TAC/T AAT GCT GAC/T TTC GAC/T GGT GAC GAA ATG AAT3' (33mer) as previously described [12]. Duplicate filters were hybridized to the 0.7 kb *Pvu*II/*Hin*dIII fragment of pTrp5.9 (encoding pol II sequences, [12]) at a final stringency of 1 × SSC at 50°C. At this stringency this probe also cross-hybridizes to the Trp6 locus [16], encoding Pol III sequences. By selecting those clones which only hybridized to the 33mer, we excluded clones encoding Pol II or Pol III sequences.

2.2. *DNA sequencing*

The isolated clone was mapped by Southern analysis. Genomic fragments of the 11 kb *Hin*dIII fragment were subcloned into pEMBL vectors [17]. Fragments of the Trp11 locus, containing the 3'half of the gene were isolated using fragment B (fig. 1) as a probe. Progressive unidirectional deletions of the inserted DNA were created using *Bal*31 exonuclease. DNA sequence analysis was performed using the dideoxy chain termination method [18] with modifications as described [19]. Both strands were sequenced using 7-deaza dGTP (Boehringer Mannheim) as subsitute for dGTP. Sequence analysis was performed with the program described by Queen and Korn ([20]; Microgenie™, Beckman Instruments).

3. RESULTS

3.1. *Identification and isolation of Trp11*

All eukaryotic RNA polymerases analysed to date are characterized by a higly conserved eleven amino acid motif, YNADFDGDEMN [2,21]. To identify all putative loci encoding trypanosomal polymerases, we hybridized *Hin*dIII digested genomic DNA of *T.brucei* with the 33mer, 5'TAC/T AAT GCT GAC/T TTC GAC/T GGT GAC GAA ATG AAT3', encoding this motif. This probe hybridized to four fragments, respectively, 4.8, 5.9, 11 and 28 kb in size. We have previously shown that the 4.8 and 5.9 kb *Hin*dIII fragments encode Pol II sequences [12], while the 28 kb *Hin*dIII fragment encodes Pol III sequences [16]. This suggests that the remaining locus, Trp11, might encode Pol I. This locus was isolated as a recombinant clone from a genomic library in EMBL 3 phage (see section 2.1 for details), and used for further analysis (fig. 1).

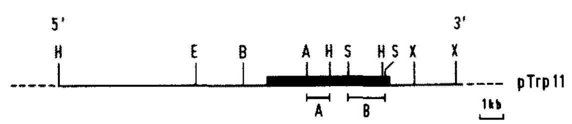
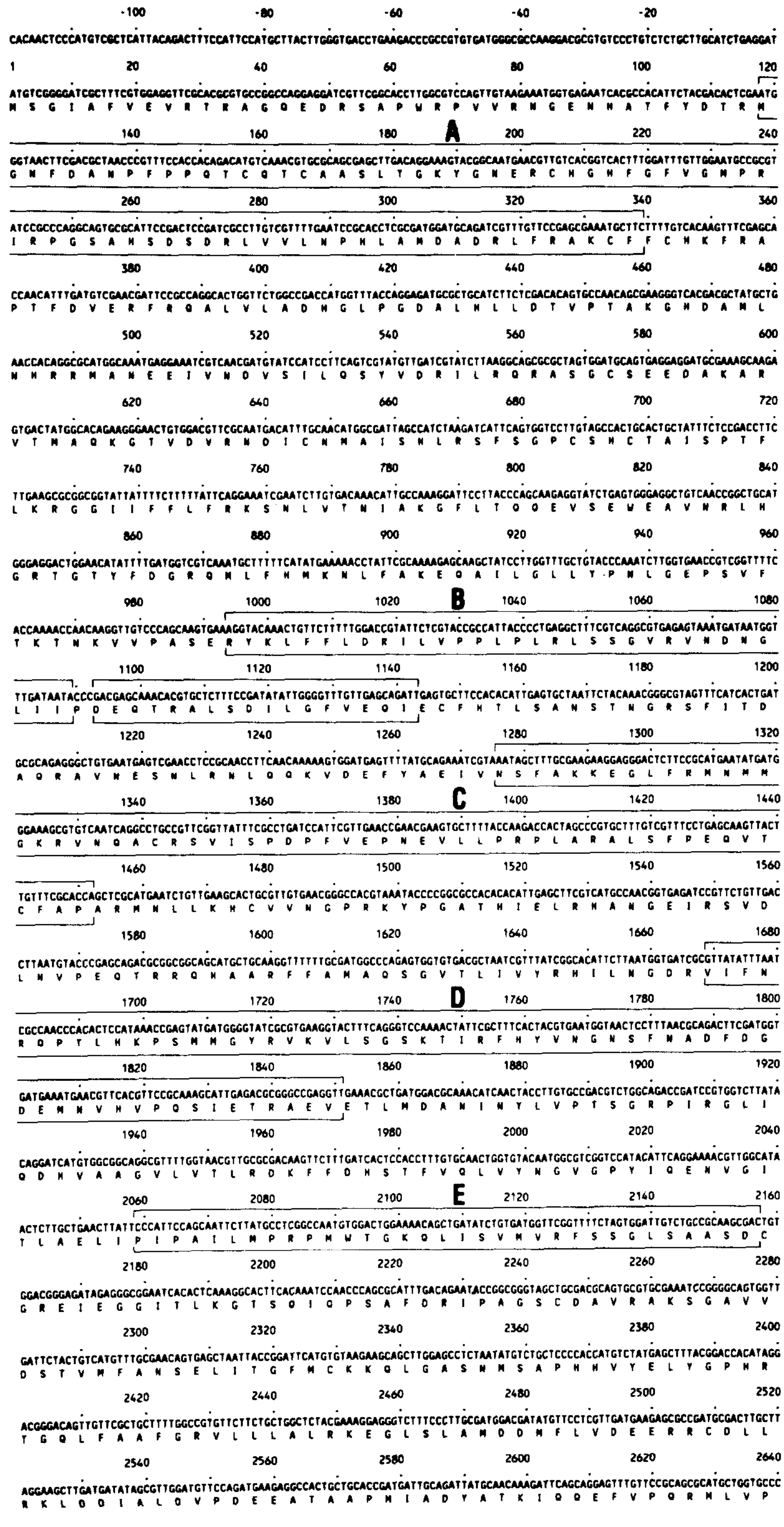


Fig. 1. Restriction map of the Trp11 locus of *T.brucei*. The coding region of the largest subunit of RNA polymerase I is indicated by a box. The probes used are indicated below the map. Abbreviations of restiction enzyme sites: A, *Ava*I; B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sal*I; X, *Xba*I.

3.2. *Sequence analysis of Trp11*

The nucleotide sequence and the deduced amino acid sequence of the gene encoded by the Trp11 locus are shown in fig. 2. The sequence revealed a single open reading frame of 5052 nt, which is in line with the size of the mRNA, 5 kb, detected in Northern blot experiments [22] with probe A (fig. 1). The predicted molecular mass of the trypanosomal subunit, based on the sequence data, is 185 kDa which is in the size range of other eukaryotic Pol Is [1,23]. The similarity of the deduced amino acid sequence and the presence of the eight homology regions A–H (fig. 2), which are characteristic for eukaryotic RNA polymerases [2,15,16,24], strongly suggest that Trp11 encodes a polymerase subunit. A dot-matrix analysis ([20]; data not shown) revealed that the sequence most closely resembled that of yeast Pol I.

A direct comparison of the trypanosomal Pol I amino acid sequence with that of the Pol I subunit of *Saccharomyces cerevisiae* [21] and *Schizosaccharomyces pombe* [25] reveals several notable features. The 18 amino acid insertion present in the B domain of both yeast subunits is absent in the trypanosomal subunit and is substituted for a single proline residue. Both yeast subunits contain two additional class-specific insertions [16] of approximately 100 mainly hydrophilic residues localized between domain A and B and between G and H [21,25]. Analogous insertions are present in the trypanosomal subunit. A striking feature was the marked difference in length of the amino acid stretches separating the G and H domains. This re-


A
B
C
D
E

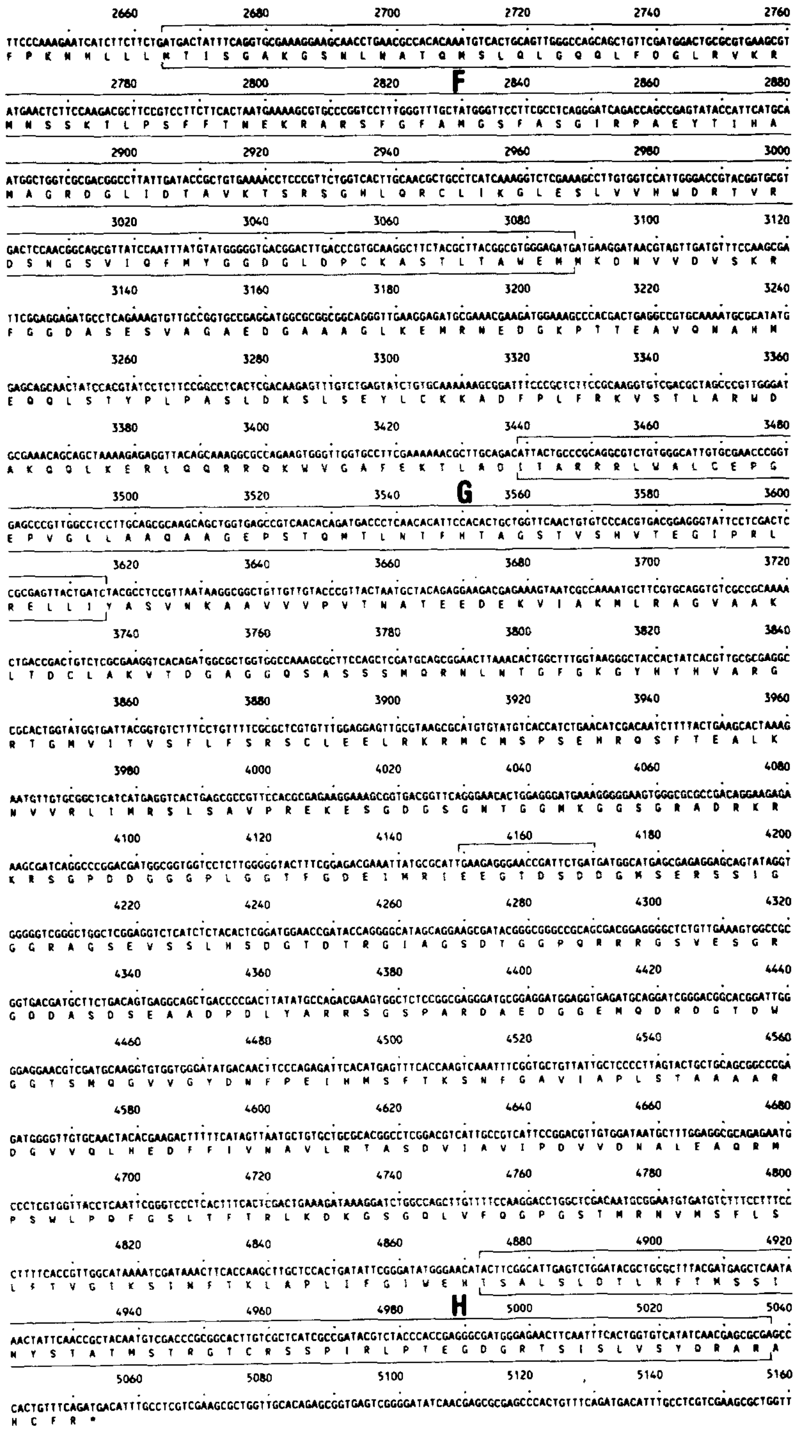


Fig. 2. Nucleotide sequence of the Trp11 gene of *T.brucei*, and the corresponding deduced amino acid sequence. The highly conserved homology regions A–H, which are characteristic for all genes encoding the largest subunit of the different eukaryotic RNA polymerases [2,15,16], are boxed. The sequence motif discussed in the text is overlined. The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession number X14399.

gion is approximately 40 amino acid residues longer in the trypanosomal subunit. These amino acids are, however, not clustered. Although this stretch shows a low overall homology between the trypanosomal and the yeast subunits, a single heptapeptide sequence EEG($^T/_I$)DSD was found at, respectively, 178 and 176 amino acid residues after the G domain of the trypanosomal and the *S.cerevisiae* subunits (fig. 2). This motif is not present in the RNA polymerase classes. However, the absence of this motif in *Sc.pombe* Pol I [25] reduces the significance of this motif. Our comparative analysis showed two other interesting differences of the trypanosomal pol I: (i) the absolutely conserved eleven amino acid motif YNADFDGDEMN [2,16,21] contains a single conservative change (*F*NADFDGDEMN); (ii) the H domain has the lowest degree of homology between identical domains of the individual subunits and is, moreover, enriched in serine and threonine residues.

4. DISCUSSION

4.1. *RNA polymerase I-specific domains*

Although our knowledge of the role of the individual subunits of the RNA polymerases is still fragmentary, it is clear that the largest subunit has an important functional role in the transcription process [1,2]. A direct comparison of the primary sequence of the trypanosomal (this paper) and yeast Pol Is [21,25] revealed the presence of two additional domains in the N- and C-terminal regions, which are absent in all analyzed Pol IIs [12–15] and Pol IIIs [13,16]. These Pol I-specific domains are highly hydrophilic, suggesting that they might be exposed to the surface and could have a functional role in the RNA polymerase I complex (cf. [21]). The significance of these domains may be in: (i) the direction of the enzyme complex to its nuclear compartment, the nucleolus, or (ii) the interaction with the nucleolar protein nucleolin ([26] and cited references), or (iii) the binding of class-I specific transcription initiation factors [27].

More detailed studies will be necessary to identify the actual role of these Pol I-specific domains in RNA polymerase I transcription.

4.2. *Trypanosomal RNA polymerases*

The VSG and ESAG genes of *T.brucei* are transcribed in an α-amanitin resistant fashion [5,8,9]. The exact mechanism of this process and the type of polymerase involved are still unclear (see [10–12]). In order to understand the basis of the α-amanitin resistant transcription, we have started to analyze the trypanosomal RNA polymerases. For this purpose, we have isolated clones encoding the putative coding regions of these proteins (this paper and [12,16]).

In this report, the structure of the gene encoding the Pol I subunit was determined by DNA sequence analysis. The Pol I structure of *T.brucei* followed the basic scheme typical for all polymerases analyzed so far (fig. 2; cf. [2,15,16]). The major deviations of the trypanosomal Pol I from the identical subunit of the yeasts *S.cerevisiae* and *Sc.pombe* are: the absence of an 18 amino acid insertion in domain B, the length of the amino acid stretch separating domain G and H, and the lack of significant homology of the trypanosomal domain H, which is enriched in serine and threonine residues.

Unfortunately, these obervations give no insight into the way these differences might affect trypanosomal RNA polymerase I transcription. As could be expected, they also provide no insight into whether or not RNA polymerase I transcribes the VSG transcription unit. However, the isolation of genomic clones encoding the trypanosomal Pol I–III subunits (this paper and [12,16]), will allow us to prepare specific antibodies by subcloning appropriate fragments into expression vectors and subsequent immunization of rabbits. These antibodies can be used as structural and functional probes in in vitro experiments and might help to identify the RNA polymerase trancribing the VSG genes.

Acknowledgements: We wish to thank our colleagues for valuable comments on the manuscript, Mr Klaus Lamberty for artwork and Ms Carmen Müller for photography. This work was supported by the Bundesministerium für Forschung und Technologie grant 0318885 and a research award of the Erna and Victor Hasselblad Foundation to Albert W.C.A. Cornelissen.

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, 91–131.
[3] Wieland, T. and Faulstich, H. (1978) CRC Crit. Rev. Biochem. 5, 185–260.
[4] Cully, D.F., Ip, H.S. and Cross, G.A.M. (1985) Cell 42, 173–182.

[5] 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.
[6] Johnson, P.J., Kooter, J.M. and Borst, P. (1987) Cell 51, 273–281.
[7] Gibbs, C.P. and Cross, G.A.M. (1988) Mol. Biochem. Parasitol. 28, 197–206.
[8] Kooter, J.M. and Borst, P. (1984) Nucleic Acids Res. 12, 9457–9472.
[9] Laird, P.W., Kooter, J.M. and Borst, P. (1985) Nucleic Acids Res. 13, 4253–4266.
[10] Shea, C., Lee, M.G.-S. and Van der Ploeg, L.H.T. (1987) Cell 50, 603–612.
[11] Alexandre, S., Guyaux, M., Murphy, N.B., Coquelet, H., Pays, A., Steinert, M. and Pays, E. (1988) Mol. Cell. Biol. 8, 2367–2378.
[12] Evers, R., Hammer, A., Köck, J., Jess, W., Borst, P., Mémet, S. and Cornelissen, A.W.C.A. (1989) Cell 56, in press.
[13] Allison, L.A., Moyle, M., Shales, M. and Ingles, C.J. (1985) Cell 42, 599–610.
[14] Ahearn, J.M., Bartolomei, M.S., West, M.L., Cisek, L.J. and Corden, J.L. (1987) J. Biol. Chem. 262, 10695–10705.
[15] Jokerst, R.S., Weeks, J.R., Zehring, W.A. and Greenleaf, A.L. (1989) Mol. Gen. Genet. 215, 266–275.
[16] Köck, J., Evers, R. and Cornelissen, A.W.C.A. (1988) Nucleic Acids Res. 16, 8753–8772.
[17] Dente, L., Cesareni, G. and Cortese, R. (1983) Nucleic Acids Res. 11, 1645–1655.
[18] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[19] Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963–3965.
[20] Queen, C. and Korn, L.J. (1984) Nucleic Acids Res. 12, 581–599.
[21] Mémet, S., Gouy, M., Marck, C., Sentenac, A. and Buhler, J.-M. (1988) J. Biol. Chem. 263, 2830–2839.
[22] Cornelissen, A.W.C.A., Evers, R., Grondal, E., Hammer, A., Jess, W. and Köck, J. (1989) Nova Acta Leopoldina, in press.
[23] Paule, M.R. (1981) Trends Biochem. Sci. 5, 128–131.
[24] Mémet, S., Saurin, W. and Sentenac, A. (1988) J. Biol. Chem. 263, 10048–10051.
[25] Hirano, T., Konoha, G., Toda, T. and Yanagida, M. (1989) J. Cell Biol. 108, 243–253.
[26] Borer, R.A., Lehner, C.F., Eppenberger, H.M. and Nigg, E.A. (1989) Cell 56, 379–390.
[27] Kownin, P., Batemann, E. and Paule, M.R. (1987) Cell 50, 693–699.