Volume 12 Number 11 1984 Nucleic Acids Research

# Comparison of the genes coding for the common 5' terminal sequence of messenger RNAs in three trypanosome species

Titia De Lange, Theo M.Berkvens, Harry J.G.Veerman, A.Carlos C.Frasch*, J.Dave Barry** and Piet Borst

Division of Molecular Biology, Antoni van Leeuwenhoekhuis, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Received 12 April 1984; Accepted 8 May 1984

ABSTRACT

Messenger RNAs of Trypanosoma brucei share a common 5' terminal sequence of 35 nucleotides, encoded by a mini-exon located in 1.35-kb tandemly linked repeats. We show here that sequences, almost identical to the mini-exon of T.brucei, are present in mRNAs from members of two other kinetoplastid subgenera: Trypanosoma vivax and Trypanosoma cruzi. As in T.brucei, these mini-exons are encoded by small tandemly linked repeat elements. We have determined and compared the nucleotide sequences of the mini-exon repeats from T.brucei, T.vivax and T.cruzi. This analysis shows that the mini-exon, its immediate flanking sequences and a T-rich stretch downstream are conserved, but little else Our data establish the generality of the novel transcription system, that was first found in T.brucei and that yields mRNAs with common, repeat-encoded, 5' termini.

## INTRODUCTION

Among unicellular flagellates of the genus Trypanosoma are many medically and economically important parasites. Some of these, e.g. the South-American T. cruzi, multiply in host cells; others, e.g. the African species T. brucei and T. vivax, live in the mammalian bloodstream and change the antigenic nature of their surface coat to evade elimination by the immune system (see references 1-3 for reviews).

We have shown that the variable protein in the surface coat of T.brucei, the Variant Surface Glycoprotein or VSG (4), is encoded by a split gene (5). The first 35 nucleotides of all VSG mRNAs are identical and are encoded by a mini-exon (5-10); the remainder of the VSG mRNAs is encoded by one of the numerous VSG main-exons (11). Recently, we have found that the common mini-exon sequence of VSG mRNAs is also present at the 5' end of many (if not all) non-VSG mRNAs in T.brucei (12). There are about 200 mini-exon genes per nucleus (9). Each gene is part of a 1.35-kb repeat element and these repeats are head-to-tail linked in a few

© IRL Press Limited, Oxford, England.



Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020

long tandem arrays (9,10). The clustering of mini-exon genes and the abundance of mRNAs with a mini-exon encoded 5' end have led to the proposal that transcription of protein-coding genes in T.brucei must be highly unusual and may be a discontinuous process (3,12).

We have previously shown that the mini-exon repeat element is highly conserved in the subgenus Trypanozoon and that the DNAs of T.vivax (subgenus Duttonella) and T.cruzi (subgenus Schizotrypanum) have homology to the mini-exon of T.brucei (9). Here we report a detailed analysis of the mini-exon repeat elements of these three species and show that many mRNAs of T.vivax and T.cruzi have the mini-exon sequence at their 5' ends.

## MATERIALS AND METHODS

Trypanosomes. The T. brucei stock used in this study (427) has been described by Cross (4). Growth and isolation of these trypanosomes is described by Fairlamb et al. (13). The rodent-adapted T. vivax clone used was isolated from the Y486 stock by Barry and Gathuo (14). Growth of T. cruzi strain Tulahuèn has been described (15). Crithidia fasciculata was grown in Boné medium (16).

Isolation and blotting analysis of DNA and RNA. DNA and RNA were isolated as described by Bernards et al.(17) and Van der Ploeg et al. (5) respectively. Poly(A)$^+$ RNA was isolated as described by Hoeijmakers et al.(18). DNA was analysed by restriction analysis and blotting as described by Bernards et al.(17). RNA blots were prepared using glyoxylated RNA according to the procedure of Thomas (19). Hybridization of nitrocellulose bound nucleic acids to a chemically synthesized probe of 22 nucleotides specific for the mini-exon (see text and Figure 5) was done as described previously (9). Post-hybridization washes were in 3 x SSC at 30°C.

Molecular cloning and sequence analysis of mini-exon repeat elements.

From T.brucei. A clone bank of T.brucei (MITat 1.2a) DNA fragments generated by partial digestion with MboI and ligated in the BamHI site of pAT 153 (20) were introduced in Escherichia coli DH1 (21). A recombinant plasmid containing approximately ten mini-exon repeat units was isolated by screening (22) with a



Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020

probe for the mini-exon repeat element (a 1.35-kb XmnI fragment described previously (12)). This mini-exon repeat array was used to determine the nucleotide sequence by the chemical degradation procedure (23) following the sequence strategy shown in Figure 2. From T.vivax. T.vivax DNA fragments generated by partial digestion with XmnI were ligated in the HindII site of phage M13mp8 and introduced into E.coli JM101 (24). Plaques were screened with a 22-mer probe for the mini-exon of T.brucei (see Figure 5). One clone was obtained and sequenced in both orientations (in M13mp8 and M13mp9 (24)) using the dideoxy-chainstopper method (25) and by the chemical degradation method (23) on the replicative form (for sequence strategy see Fig.2).

From T.cruzi. T.cruzi DNA fragments generated by digestion with PstI were ligated in the PstI site of phage M13mp8 , introduced into E.coli JM101 (24) and screened as described for T.vivax. One clone containing a 609 bp insert was obtained and analysed as described for the T.vivax repeat element.

Analysis of the 5' ends of mRNAs. The 5' ends of T.brucei and T.vivax mRNAs were sequenced by reverse transcriptase catalysed cDNA synthesis with the synthetic 22-mer complementary to part of the mini-exon sequence (9) (see Figure 5) as primer using the reaction conditions described by Michels et al. (26).

## RESULTS

### The genomic organization of mini-exons in T.brucei, T.vivax and T.cruzi is similar.

A chemically synthesized probe of 22 nucleotides complementary to part of the common 5' end sequence of mRNAs in T.brucei hybridizes to nuclear DNA from T.vivax and T.cruzi (9). We have characterized the hybridizing elements in T.vivax and T.cruzi DNA by restriction analysis and Southern blotting as shown in Figure 1; the resulting maps are presented in Figure 2. T.vivax contains a repeat element of 683 bp with one HindIII site and two XmnI sites only 27 bp apart (Fig. 2). The majority of the repeats must be tandemly linked because partial digestion with XmnI yields unit length multimers (Fig. 1). Likewise T.cruzi contains a major hybridizing repeat element of 609 bp that is tandemly linked as judged from partial digestion with PstI (Fig. 1 and Fig. 2). Thus, as in T.brucei, the majority of the mini-exons in



Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020

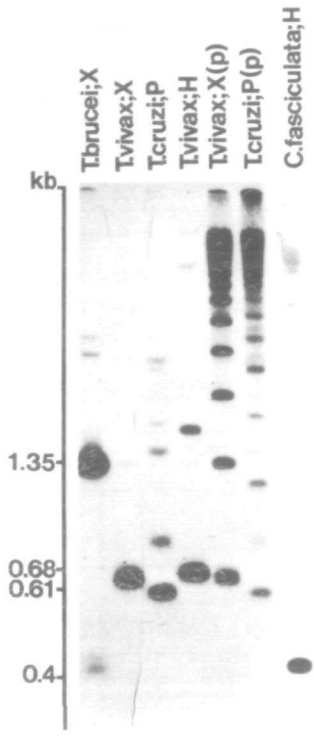



Figure 1. Blot analysis of DNA from three trypanosome species and C. fasciculata showing the presence of mini-exon repeat elements. About 2 µg DNA from each trypanosomatid was digested with the indicated restriction enzymes, size-fractionated on a 1% agarose gel and transferred to a nitrocellulose filter. The filter-bound DNA was hybridized to a probe of 22 nucleotides complementary to position +13 to +34 of the mini-exon sequence of mRNAs in T.brucei (9) (see Figure 5). Abbreviations: H, HindIII; P, PstI; P(p), partially digested with PstI; X, XmnI; X(p), partially digested with XmnI.

T.vivax and T.cruzi is present on small repeat elements that are organised in long tandem arrays.

We extended this analysis to the DNA from the insect trypanosome C.fasciculata. Figure 1 shows that this DNA also contains a putatively repeated element that hybridizes to the mini-exon probe. Digestion with HindIII (Fig. 1) and TaqI (not shown) yields a fragments of 400 bp, whereas digestion with XmnI, PstI or PvuII yields a hybridizing band at the top of the gel (>25 kb) (not shown). The simplest interpretation is that C.fasciculata contains tandemly linked mini-exon repeat elements of about 400 bp, containing one HindIII and one TaqI site, but no sites for XmnI, PstI or PvuII.

Many mRNAs in T.vivax and T.cruzi have a mini-exon at their 5' end.

The bulk of the mRNAs in T.brucei have a common 5' terminus that is encoded by the 35-bp mini-exon (12) . We verified this for T.vivax and T.cruzi by hybridization of the 22-mer mini-exon



Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020

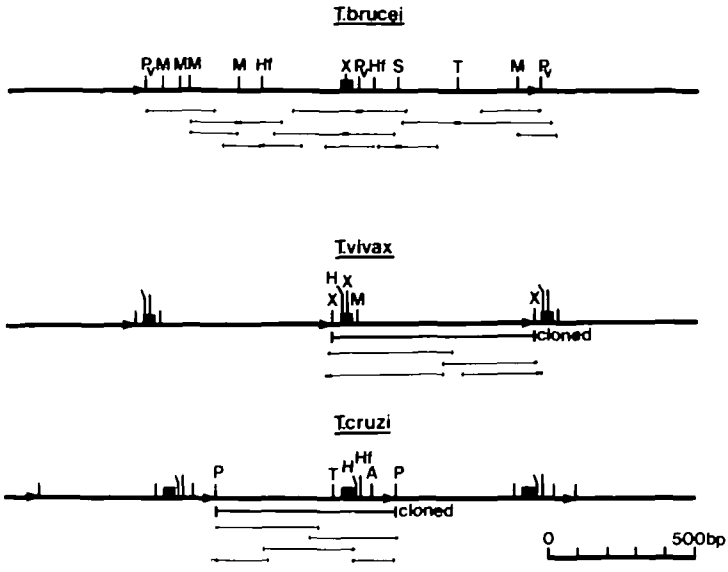



Figure 2. Physical maps of the mini-exon repeat elements in three trypanosome species. The physical maps of the tandemly linked mini-exon repeats were derived from blotting analysis of nuclear and cloned DNA as shown in Figure 1. Mini-exons are drawn as black boxes; the direction of transcription is from left to right. For simplicity, the restriction sites in the T.brucei repeats are identified in one unit only. The bars below the T.vivax and T.cruzi maps indicate the restriction fragments that have been cloned; from T.brucei a tandem array of approximately ten units was cloned. Below each map the strategy used for sequence analysis is shown. Arrows with open circles indicate analysis by the dideoxy-chainstopper method; arrows with closed circles indicate analysis by the chemical degradation method; and arrows with both open and closed circles indicate analysis by the chemical degradation method using restriction sites outside the cloned mini-exon repeat, i.e. in the polylinker of the cloning vehicle. Abbreviations: A, AvaII; H, HindIII; Hf, HinfI; M, MspI; P, PstI; Pv, PvuII; S, Sau3A; T, TaqI; X, XmnI.

probe to oligo(dT)-selected, size-fractionated RNA. Figure 3 shows that both species contain poly(A)$^+$ RNAs with mini-exon sequences. The size distribution of these RNAs is similar to that seen in T.brucei (Fig.3). The prominent band in T.brucei represents VSG mRNA; the band at 1400 nt in T.vivax RNA might also represent VSG mRNA.

Figure 4 shows that mRNAs from T.vivax, like T.brucei, have the mini-exon sequence predominantly at their 5' ends. The major product of 22-mer primed cDNA synthesis is 35 nucleotides long and very few run-through products are visible. The deduced se-



Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020

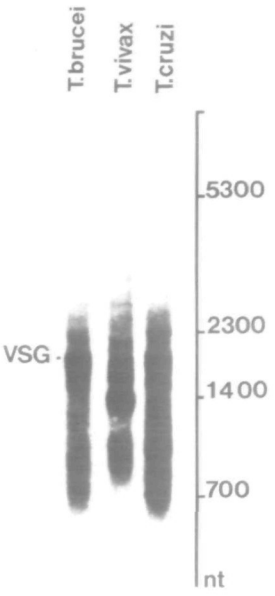



Figure 3. Many mRNAs of T.brucei, T.vivax and T.cruzi contain mini-exon sequences. Approximately 2 µg glyoxylated poly(A)$^+$ RNA from T.brucei (MITat 1.8b, see ref. 27), T.vivax and T.cruzi was size-fractionated on a 1.2% agarose gel, transferred to nitrocellulose and hybridized to the 22-mer probe for the mini-exon of T.brucei (see Figure 5). Molecular weights were calibrated by co-migration of PvuII-digested, glyoxylated DNA of a cosmid clone (cPR 1, see ref. 9) that contains several mini-exon repeat units. The VSG mRNA of T.brucei is indicated.

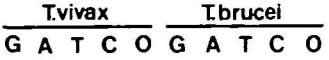
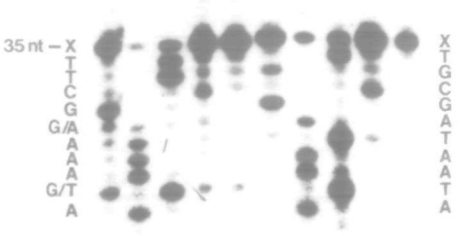
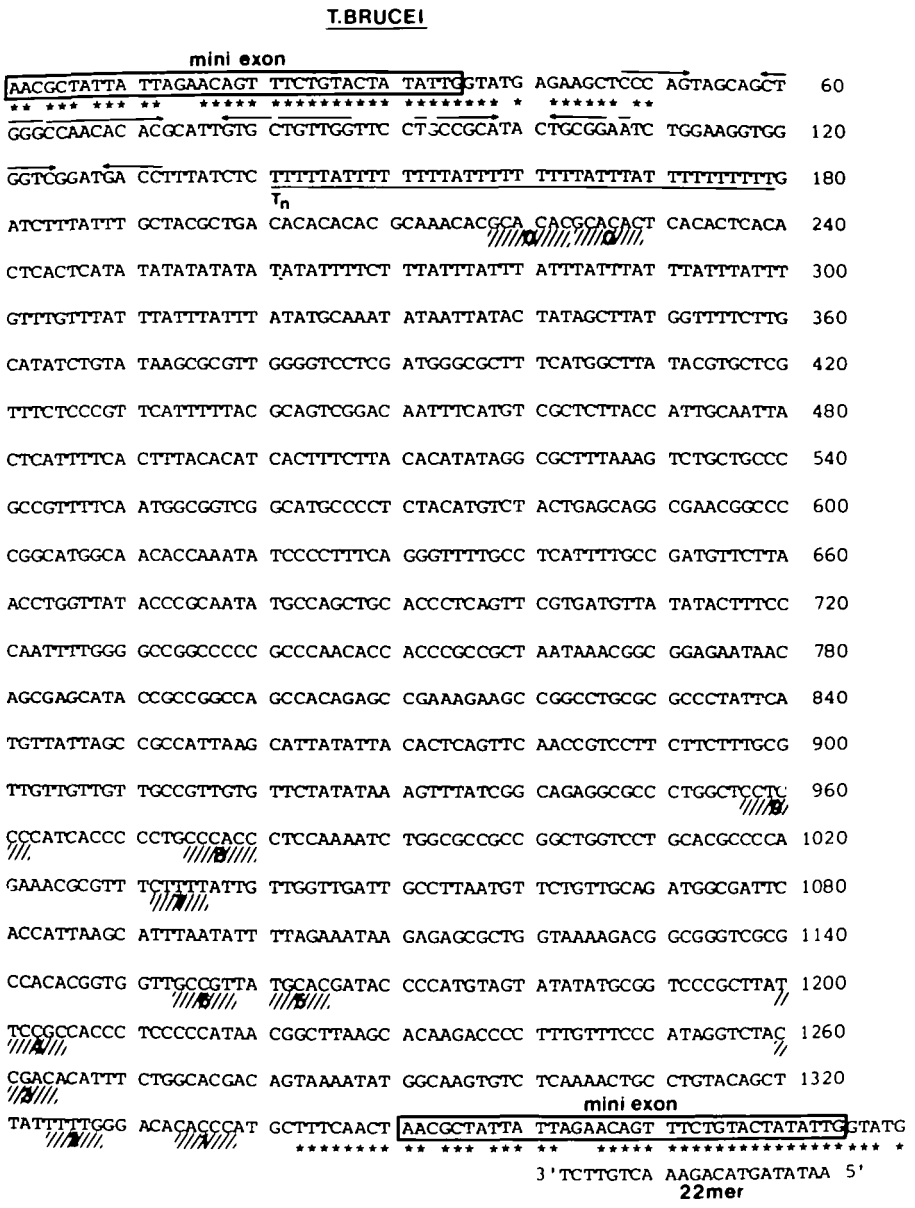


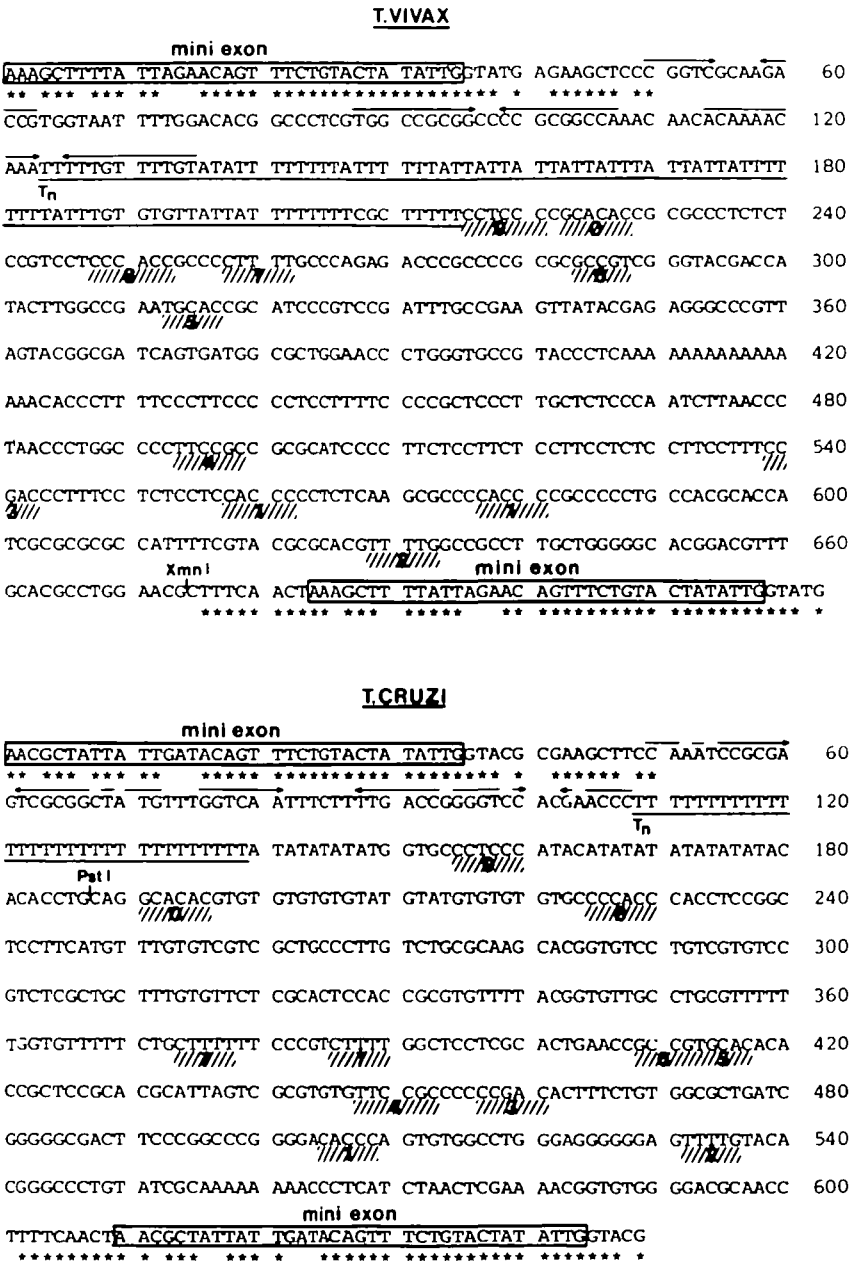

Figure 4. The mini-exon sequence of trypanosome mRNAs is predominantly located at 5' ends. Approximately 1 µg poly(A)$^+$ RNA from T.vivax and T.brucei was copied into DNA by reverse transcriptase using the 22-mer mini-exon probe (9)(see Figure 5) as primer . Either one (G, A, T or C) or no (O) dideoxynucleotide and $\alpha$-$^{32}$P-dATP were added to the reaction mixture.




Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020




Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020




Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020

quence of the 5' ends of mRNAs from T.vivax agrees with the sequence of a genomic mini-exon (see below). However, at two positions the RNA sequence is ambiguous (see Fig 4); therefore, the mini-exons of T.vivax may be heterogeneous. Priming with the 22-mer on T.cruzi RNA also gave a cDNA product of 35 nucleotides (not shown). The yield was very low, however, probably because the primer has a 3-bp mismatch at the 3' end with the T.cruzi mini-exon (see below).

Comparison of the nucleotide sequence of three mini-exon repeat elements

We have cloned and sequenced the mini-exon repeat elements from T.brucei, T.vivax and T.cruzi (see Materials and Methods for experimental details); a printout of the sequences is given in Figure 5. The sequence of the T.brucei element was determined on a cloned array of ± ten repeats by the chemical degradation procedure (23). As a consequence of this strategy, approximately ten repeat elements were analysed simultaneously. As this approach did not result in ambiguities in the sequence, the majority of the mini-exon repeats in this array must have the sequence in Figure 5. However, minor sequence heterogeneity between mini-exon repeats from different clusters does exist because in a second repeat array we found three differences within 300 bp (not shown).

Both the T.vivax and the T.cruzi repeat sequence were determined on a single cloned repeat. As a consequence we have not sequenced across the restriction sites at the ends of the

Figure 5. Comparison of the nucleotide sequences of the mini-exon repeats of T.brucei, T.vivax and T.cruzi. To emphasize the tandem linkage of these elements the (non-coding strand) sequences are printed starting and ending with a mini-exon (boxed). Conserved elements are indicated in three ways: asterikses show the conserved nucleotides in and immediately adjacent to the mini-exons; the conserved T-rich stretch downstream of the mini-exons is underlined; and the (numbered) hatched boxes indicate sequences of five or more nucleotides that occur in approximately the same order upstream (1-9) or downstream (0) of the three mini-exons. The latter elements were found by matrix screening using a programme provided by B. de Vries (University of Amsterdam, Amsterdam, The Netherlands). Arrows denote hyphenated dyad symmetries immediately downstream of the mini-exons. Below the T.brucei sequence the 22-mer probe for the mini-exon is shown. The sequencing strategy is described in the Materials and Mehods section and in Figure 2. The XmnI and PstI sites shown in the T.vivax and the T.cruzi sequence, respectively, mark the ends of the cloned repeat elements and are the positions where a few nucleotides could be missing (see text).



Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020

cloned fragments and cannot exclude that we have missed a small restriction fragment. However, if a fragment is missed, it cannot be larger than 15 bp as we do not detect a partial digestion product indicative for such a fragment (Fig. 1). This ambiguity is indicated in Figure 5.

A general characteristic of the repeat elements is the presence of long stretches of simple repeat sequences: e.g the alternating purine/pyrimidine sequence in the T.brucei (position 200) and the T.cruzi (position 138) repeats of which parts may have the potential to adopt the Z conformation (for review see ref. 28); the $(ATTT)_n$ sequence in the T.brucei repeat at position 270; and the $C_nT_m$ blocks in the T.vivax (position 425) and in the T.cruzi (around position 360) repeats. However, as none of these elements are present in all three repeats their biological significance is doubtful.

Between the three repeat elements there is very little sequence conservation (see Figure 5 for conserved blocks). The highest level of conservation is found in and around the mini-exon: 29 of the 35 bp of the mini-exon are identical in the three repeats and immediately upstream of the mini-exon a perfectly conserved octanucleotide is present. Further upstream of the mini-exons there are nine conserved blocks of at least five nucleotides. These motifs and the octanucleotide preceding the initiation site could play a role in transcription. As we do not know which polymerase transcribes this gene, we cannot predict the position or nature of promoter sequences. It is even possible that the high level of conservation within the mini-exon reflects a promoter function for this block.

Another prominent conserved region is found immediately 3' of the mini-exon (see Fig. 5). This area is part of the putative splice donor site of the mini-exon (5'TTG/GTAPyG3'), that resembles the consensus sequence (29) for splice donor sites (5'CAG/GTPuAG3') and it is possible that this function imposes the observed sequence constraint. Further downstream yet another conserved feature is present. Approximately 110-140 bp from the end of the mini-exons the repeat elements contain long stretches of T-residues in the strand shown in Figure 5. The T-stretches are preceded by an area that contains three hyphenated dyad symmetries, which are not conserved in size or nucleotide se-



Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020

quence (see Fig. 5). It is possible that either or both motifs play a role in termination of transcription; the analysis of nascent transcripts and steady state RNA from T.brucei indicates that mini-exon transcripts end in the T-stretch (J.M. Kooter, pers. comm.). This transcript is not detected in the experiment in Figure 3 and other RNA blots (see ref. 9 and 12) because of its low abundance and the inefficient retention of small RNAs by nitrocellulose.

## DISCUSSION

In a previous paper we demonstrated that many mRNAs in T.brucei share a common sequence at their 5' ends (12). This mini-exon is located within a 1.35-kb repeat element present in ± 200 copies per nucleus (9). Here we report that mRNAs of two other trypanosomes, T.vivax and T.cruzi, have a similar mini-exon sequence and that this sequence is encoded by a small repeat element. As in T.brucei, mini-exon repeats in T.vivax and T.cruzi are tandemly linked in long arrays. Hence the number of mini-exon repeat arrays is low and probably insufficient to supply every gene with its own mini-exon. We have proposed alternative explanations for this observation (12). First, genes could be clustered downstream of a mini-exon repeat array. Secondly transcription could be a discontinuous process either involving a jumping polymerase or bimolecular splicing. Our present analysis indicates that this novel transcription system is not confined to trypanosomes that show antigenic variation, but is a more general phenomenon in the genus Trypanosoma. This conclusion is further substantiated by the observation that DNA from Trypanosoma cyclops and Trypanosoma rangeli hybridizes to the 22-mer probe for the mini-exon of T.brucei (L.H.T. van der Ploeg, pers. comm.) and it may hold for all trypanosomatids as DNA from C.fasciculata and Herpetomonas muscarum also has homology to the mini-exon (this paper and L.H.T. Van der Ploeg, pers. comm.).

We have determined and compared the nucleotide sequence of the mini-exon repeat elements from T.brucei, T.vivax and T.cruzi. Although the overall level of conservation of the mini-exon repeats is very low, sequences in and around the mini-exon are rather conserved. Critical testing of the function of these



Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020

sequences awaits the development of systems in which mutagenized templates can be probed for their biological activity.

ACKNOWLEDGEMENTS

We thank Mrs. F. Fase-Fowler for the isolation of DNA from C.fasciculata, L.H.T. Van der Ploeg and J.M. Kooter for sharing unpublished results, B. De Vries (University of Amsterdam, Amsterdam) for providing computer facilities and Dr. D. Kabat (Oregon Health Sciences University, Oregon, USA) for the initial blotting analysis of T.vivax and T.cruzi DNA. This work was supported in part by a grant from the Foundation for Fundamental Biological Research (BION), which is subsidized by The Netherlands Organization for the Advancement of Pure Research (ZWO). This research was also partly supported by a grant (J.D.B.) from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases.

* Catedra de Bioquimica, Facultad de Medicina, Universidad de Buenos Aires, Paraguat 2155, 5th piso, (1121) Capital Federal, Buenos Aires, Argentina.
**Departement of Genetics, University of Glasgow, Church St., Glasgow G11 5JS United Kingdom.

Abbreviations: bp, base pair(s); cDNA, complementary DNA; kb, kilo-base-pairs; MITat, Molteno Institute Trypanosome antigen type; mRNA messenger RNA; nt, nucleotide(s); VSG, Variant Surface Glycoprotein.

REFERENCES

1. Vickerman, K. (1978) Nature 273, 613-617.
2. Englund, P.T., Hajduk, S.L., Marini, J.C. (1982) Ann. Rev. Biochem. 51, 695-726.
3. Borst, P., Bernards, A., Van der Ploeg, L.H.T., Michels, P.A.M., Liu, A.Y.C., De Lange, T. and Kooter, J.M. (1983) Eur. J. Biochem. 137, 383-389.
4. Cross, G.A.M. (1975) Parasitology 71, 393-417.
5. Van der Ploeg, L.H.T., Liu, A.Y.C., Michels, P.A.M., De Lange, T., Borst, P., Majumder, H.K., Weber, J., Veeneman, G.H. and Van Boom, J. (1982) Nucl. Acids Res. 10, 3591-3604.
6. Boothroyd, J.C. and Cross, G.A.M. (1982) Gene 20, 281-289.
7. Liu, A.Y.C., Van der Ploeg, L.H.T., Rijsewijk, F.A.M. and Borst, P. (1983) J. Mol. Biol. 167, 57-75.
8. Bernards, A., De Lange, T., Michels, P.A.M., Liu, A.Y.C., Huisman, M.J. and Borst, P. (1984) Cell 36, 163-170.
9. De Lange , T., Liu, A.Y.C., Van der Ploeg, L.H.T., Borst, P., Tromp, M.C., and Van Boom, J.H. (1983) Cell 34, 891-900.
10. Nelson, R.G., Parsons, M., Barr, P.J., Stuart, K., Selkirk, M. and Agabian, N. (1983) Cell 34, 901-909.
11. Van der Ploeg, L.H.T., Valerio, D., De Lange, T., Bernards, A., Borst, P.



Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020

and Grosveld, F.G. (1982) Nucl. Acids Res. 10, 5905-5923.
12. De Lange, T., Michels, P.A.M., Veerman, H.J.G., Cornelissen, A.W.C.A. and Borst, P. (1984) Nucl. Acids Res., in press.
13. Fairlamb, A.H., Weislogel, P.O., Hoeijmakers, J.H.J. and Borst, P. (1978) J. Cell. Biol. 76, 293-309.
14. Barry, J.D. and Gathuo, H. (1984) Parasitology, in press
15. Warren, L. (1960) J.Parasitol. 46, 529-539.
16. Boné, G.J. and Steinert, M. (1956) Nature 178, 308-309.
17. 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.
18. Hoeijmakers, J.H.J., Borst, P., Van den Burg, J., Weissmann, C. and Cross, G.A.M. (1980) Gene 8, 391-417.
19. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. U.S. 77, 5201-5205.
20. Twigg, A.J. and Sherratt, D. (1980) Nature 283, 216-220.
21. Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.
22. Grosveld, F.G., Dahl, H.H.M., De Boer, E. and Flavell, R.A. (1981) Gene 13, 227-237.
23. Maxam, A. and Gilbert, W. (1980) in: Methods in Enzymology (Grossman, L. and Moldave, K., Eds) Vol. 65, Academic Press, New York, pp 499-560.
24. Messing, J. (1981) Third Cleveland Symposium on Macromolecules: Recombinant DNA, A. Walton Ed., Elsevier, Amsterdam, pp 143-153.
25. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S. 74, 5463-5467.
26. 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., Tromp, M.C. and Van Boom, J.H. (1983) J. Mol. Biol. 166, 537-556.
27. Michels, P.A.M., Van der Ploeg, L.H.T., Liu, A.Y.C. and Borst, P. (1984) EMBO J., in press.
28. Zimmerman, S.B. (1982) Ann.Rev. Biochem. 51, 395-427.
29. Mount, S.M. (1982) Nucl. Acids Res. 10, 459-472.

Downloaded from https://academic.oup.com/nar/article-abstract/12/11/4431/1109763 by guest on 25 July 2020