Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 3998–4002, June 1985
Biochemistry

Calmodulin genes in trypanosomes are tandemly repeated and produce multiple mRNAs with a common 5′ leader sequence

(Ca^{2+} binding protein/DNA sequence conservation/oligonucleotide-specific primer extension/mRNA formation)

CHRISTIAN TSCHUDI*, ALEXANDER S. YOUNG*, LARRY RUBEN†, CURTIS L. PATTON†, AND FRANK F. RICHARDS*

*Department of Internal Medicine and †Department of Epidemiology and Public Health, Yale University, School of Medicine, New Haven, CT 06510

Communicated by Dorothy M. Horstmann, February 19, 1985

ABSTRACT **In *Trypanosoma brucei gambiense*, the Ca^{2+} binding protein calmodulin is encoded by three identical tandemly repeated genes. The transcripts of these genes consist of several RNA species similar in size. A 35-nucleotide spliced leader sequence is present at the 5′ end of each mRNA but is not encoded by DNA contiguous to these genes. We have identified two different sites for the fusion of the leader to the mRNA. These results strongly support the idea that a novel, possibly discontinuous, transcription mechanism is used by these parasites.**

African trypanosomes can change their antigenic identity by successively expressing variant surface glycoproteins (VSGs). This variation allows the trypanosome to escape the host immune response, resulting in relapsing parasitemia. The study of nuclear gene expression in these parasitic protozoa has been limited almost exclusively to the mechanism of antigenic variation (reviewed in ref. 1). Little is known about the expression of structural genes other than VSG genes. In order to extend the understanding of gene regulation in trypanosomes, we began to analyze the calmodulin genes. Calmodulin, a calcium-binding protein, is present in all animal and plant cells in which it has been looked for (reviewed in ref. 2). Its amino acid sequence of 148 residues shows a remarkable degree of conservation during evolution. Furthermore, many calcium-dependent processes are known to be regulated by calmodulin. The calmodulin from African trypanosomes has been isolated and shown to be physically and functionally related to other calmodulins (3, 4). In this communication, we report the structure and expression of the calmodulin genes in *Trypanosoma brucei gambiense*. The calmodulin locus contains three identical coding regions which are arranged in a tandem array. We also present results demonstrating that a 35-nucleotide leader sequence is present at the 5′ end of calmodulin mRNA and discuss the implication of these findings for both the control of Ca^{2+} metabolism and gene expression in trypanosomes.

MATERIALS AND METHODS

Trypanosomes. *T. b. gambiense* cloned variant antigen types of the Texas trypanozoon antigen type (TXTat) serodeme were used (5). Bloodstream trypanosomes were grown in irradiated rats and purified by chromatography on DEAE-cellulose (6).

Isolation of Nucleic Acids. High molecular weight trypanosome DNA and total cytoplasmic RNA was isolated as described (5). Poly(A)$^+$ RNA was prepared by three passages through an oligo(dT)-cellulose column.

Construction and Screening of Genomic Clones. High molecular weight DNA from *T. b. gambiense* variant 1 was partially digested with the restriction enzyme *Sau*3A. Fragments ranging in size from 18–24 kilobases (kb) were isolated on a 5–24% NaCl gradient and used for cloning into the bacteriophage vector EMBL4 (7). This vector was cut with *Bam*HI and ligated to the size selected *Sau*3A fragments. The ligation products were packaged *in vitro* (8) and plated on a lawn of Q359 cells to yield about 2×10^5 recombinants per μg of vector DNA.

The library was screened by the method of Benton and Davis (9) using the nick-translated electric eel calmodulin cDNA pCM109 as a probe. Baked nitrocellulose filters were preincubated for 2–6 hr at 37°C in 50% deionized formamide containing 5× NaCl/Na Cit (1× NaCl/Na Cit = 0.15 M NaCl/0.015 M Na citrate, pH 7), 0.1% bovine serum albumin, 0.1% polyvinylpyrrolidone, 0.1% Ficoll, and 0.2% $NaDodSO_4$. Hybridization with labeled DNA was done in the same solution for 48 hr at 37°C. Filters were then washed three times at room temperature for 30 min in 2× NaCl/Na Cit containing 0.2% $NaDodSO_4$.

Primer Extension. The synthetic oligonucleotide (5′ T-C-A-A-G-T-G-G-A-T-G-T-T-A-C 3′) was 5′-end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase and was annealed for 2–4 hr at 65°C with 2 μg of poly(A)$^+$ RNA in 10 μl of 10 mM Pipes, pH 6.4/0.4 M KCl. At the end of the hybridization, the reaction mixture was diluted to 0.1 M KCl and adjusted to 50 mM Tris·HCl (pH 8.0), 50 mM $MgCl_2$, 10 mM dithiothreitol and 0.5 mM dATP, dGTP, dCTP, and TTP. Avian myeloblastosis virus reverse transcriptase (5–10 units) was added, and the incubation was continued at 43°C for 1 hr. The sample was extracted with phenol, precipitated with ethanol, and analyzed by electrophoresis on 8% polyacrylamide/7 M urea gels. For DNA sequence analysis, the extension products were cut out, extracted in 0.3 M NH_4OAc, and ethanol-precipitated with carrier tRNA.

DNA Sequence Analysis. End-labeled fragments were sequenced by the chemical degradation method of Maxam and Gilbert (10). Phage M13 hybrid single-stranded DNA was used as template in the dideoxy chain-termination method of Sanger *et al.* (11). The radioactive label in the sequencing reactions was α-^{35}S-labeled deoxyadenosine 5′-[α-thio]triphosphate (New England Nuclear), and the reaction products were fractionated on gradient gels (12).

RESULTS

Isolation and Characterization of Calmodulin Genomic Clones. To isolate trypanosome calmodulin genes, a genomic phage library of *T. b. gambiense* DNA was screened by using the electric eel calmodulin cDNA pCM109 as a hybridization probe. pCM109 contains sequences encoding amino acids 93–148 in addition to 124 nucleotides of the 3′ untrans-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: VSG, variable surface glycoprotein; SL, spliced leader; TXTat, Texas trypanozoon antigen type; kb, kilobases; bp, base pair(s).

lated region (13). Using the low-stringency hybridization conditions as described, we screened 11,000 recombinants, or about five genome equivalents. Two weakly hybridizing clones were isolated and called λCM51 and λCM55.

Restriction enzyme analysis showed that these two recombinants contain overlapping DNA regions covering a total of 35 kb (Fig. 1). The pattern of hybridization of pCM109 to Southern blots of λCM51 and λCM55 revealed about 3.5 kb of calmodulin-like sequences in the two phage DNAs. In addition, these experiments indicated that the cross-hybridizing sequence is contained within a 0.85-kb repeat unit.

Protein Coding Regions for Trypanosome Calmodulin Are Tandemly Repeated. To ensure that the sequences isolated were indeed coding for trypanosome calmodulin, we sequenced the DNA fragments that contained the regions of homology with pCM109. The sequencing strategies are outlined in Fig. 1 and the results are displayed in Fig. 2. Several important conclusions follow from these data. The sequence analysis yielded the complete protein coding region for trypanosome calmodulin. An open reading frame starts from an initiation codon AUG at position 1, extends into the calmodulin-specific sequence and terminates at position 448 with an amber codon. Encoded is a 148-amino acid protein that bears extensive sequence homology with calmodulin isolated from a variety of eukaryotic organisms. In particular, when compared with the amino acid sequence of the bovine brain calmodulin (15), the trypanosome protein shows 89% homology with a total of 17 substitutions (indicated by stars in Fig. 2). These results agree with the amino acid sequence recently obtained for trypanosome calmodulin (4).

The physical map of the isolated genomic clones indicated that trypanosome calmodulin genes are tandemly reiterated in the genome. The nucleotide sequence shown in Fig. 2 confirmed the structures deduced from restriction enzyme analysis. The calmodulin locus was found to contain three identical protein coding regions (designated genes *A*, *B*, and *C*), uninterrupted by intervening sequences.

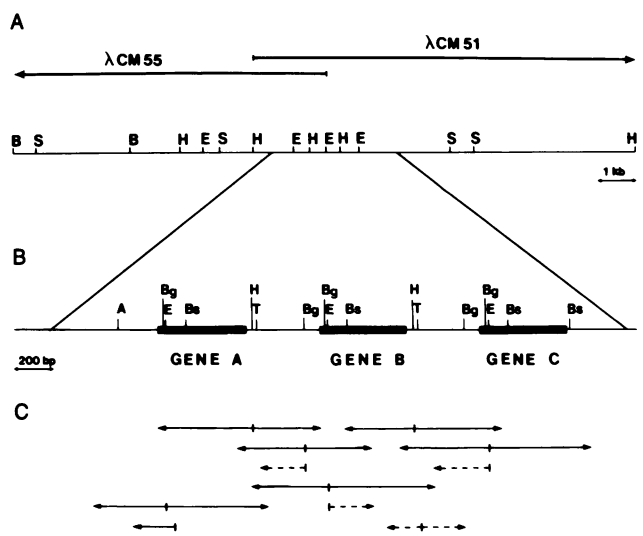

FIG. 1. Physical map of the calmodulin locus. (*A*) A partial restriction map of the two overlapping bacteriophage recombinants (λCM51 and λCM55) was obtained by analyzing single and double digests along with Southern hybridization (14) using nick-translated pCM109. (*B*) The region subjected to DNA sequence analysis is shown. Only the main restriction sites are indicated. The positions of the three coding regions are denoted by solid boxes. The direction of transcription is from left to right. (*C*) Sequencing strategy for the calmodulin genes. The arrows depict the direction and extent of sequencing from the restriction enzyme sites. Solid and broken arrows indicate regions sequenced by the chain-termination (11) and the chemical degradation (10) methods, respectively. A, *Apa* I; B, *Bam*HI; Bg, *Bgl* II; Bs, *Bst*NI; E, *Eco*RI; H, *Hin*dIII; S, *Sal* I; T, *Taq* I.

Examination of the sequences flanking the coding regions revealed that the homology upstream from the AUG of all three genes is limited to 57 bp (this sequence is underlined in Fig. 2). Examining further in the 5′ direction, we note that the homology between the 5′ untranslated region of gene *A* and the other two genes drops off precipitously to random levels. Therefore, the 5′ boundary of the repeated structure is located at position −57. We can define two complete copies of a calmodulin repeat unit of 843 bp (nucleotides −57 to 786 and 787 to 1629, Fig. 2). Comparison of these sequences reveals a remarkable degree of homology; there are only three base changes, all in the intergenic region. The third unit starts at position 1630, and the sequence homologous to the calmodulin repeat is 512 bp long including 448 bp of protein coding region. The 3′ flanking region of gene *C* is different in nucleotide sequence when compared to that of gene *A* or *B*. A clear line of demarcation can be drawn immediately after the termination codon (nucleotide 2142).

The results presented in this section demonstrate that the calmodulin genes in trypanosomes are organized in three tandemly repeated units. Using the trypanosome calmodulin gene as a radioactive probe, we have confirmed this structural arrangement by comparing the restriction map of the recombinant clones with genomic fragments detected on Southern blots (results not shown).

Transcripts from the Calmodulin Genes Are Synthesized with a Common 5′ Leader Sequence. The transcripts derived from the calmodulin region were analyzed by blot hybridization. The 0.84-kb *Eco*RI fragment of phage λCM51, which contains one repeat unit of the calmodulin locus, was subcloned into pUC8, and the resulting plasmid was designated p512. This fragment was used to probe a blot of TXTat 1 poly(A)$^+$ RNA, and the result is shown in Fig. 3. By this analysis, we detected a transcript of approximately 800 nucleotides. Because of the repetition of the calmodulin genes, we cannot exclude the possibility that several transcripts comigrate with an apparent size of 800 nucleotides.

The 5′ end(s) of the mRNA(s) was determined more precisely by primer extension analysis. A ^{32}P-labeled synthetic oligonucleotide complementary to nucleotides −21 to −7 (Fig. 2) was hybridized with poly(A)$^+$ RNA and used as a primer for extension with reverse transcriptase. This primer is part of the 57 nucleotides that are conserved in the 5′ untranslated region of all three genes (Fig. 2). The size of the extension products was determined by electrophoresis on a polyacrylamide/urea gel. For use as size markers, dideoxy-sequencing reactions were carried out with the same primer hybridized to M13 mp8 subclones of the corresponding region. The predominant extension products had lengths of 78–103 nucleotides, corresponding to a 5′ untranslated region of 84–109 nucleotides (Fig. 4). Several minor bands were occasionally observed. Because the relative amount of these bands varied from experiment to experiment, they are probably premature termination products of reverse transcription. None of the five major bands was obtained when poly(A)$^+$ RNA was omitted from the reaction (lane 2 in Fig. 4).

The five major cDNA fragments were sequenced by the method of Maxam and Gilbert (10). These experiments showed that the extension products contain sequences corresponding to three different mRNAs. Two of these sequences, designated cDNA 1 and cDNA 2, can be aligned with the 5′ untranslated region of the calmodulin locus (see below). The third cDNA, on the other hand, contains sequences only partially related to the 5′ untranslated region. The origin of this cDNA is unclear and has not yet been studied further.

Alignment of the nucleotide sequences of cDNA 1 and cDNA 2 with the genomic DNA sequence revealed that the

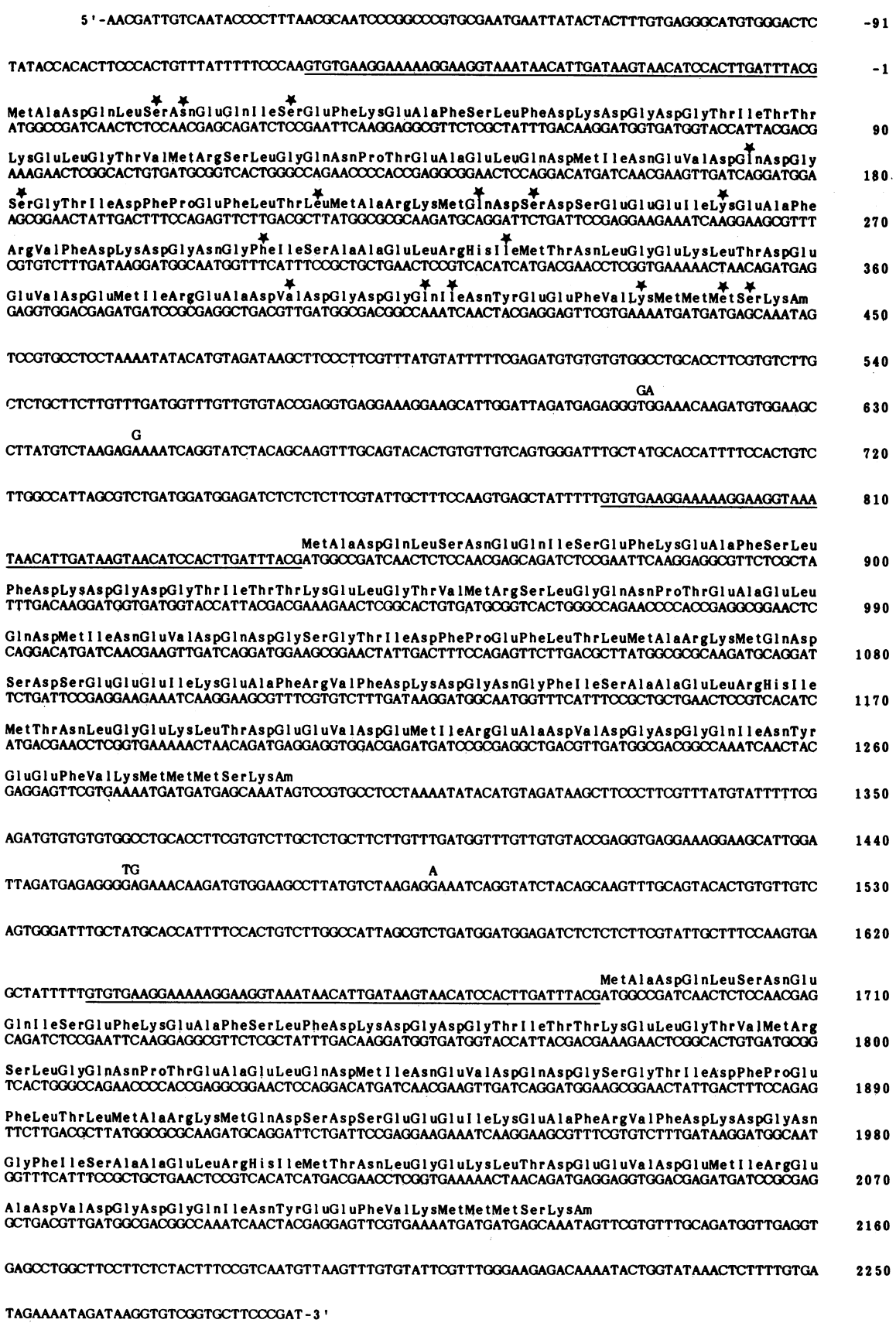

FIG. 2. Sequence of the calmodulin locus. The mRNA coding strand of the nucleotide sequence is shown. Numbering starts at the A of the initiation codon of gene *A*. Negative numbers refer to the 5' untranslated region of gene *A*. The conserved sequence (57 nucleotides) found upstream from the ATG of all three genes is underlined. The three nucleotide changes in the intergenic region are indicated above the lines (see text). The deduced amino acid sequences are shown above the corresponding nucleotides. The stars mark 17 substitutions as compared to the amino acid sequence of bovine brain calmodulin (15). Amino acid sequence analysis of trypanosome calmodulin revealed a posttranslational modification that does not appear in these data (unpublished data). The lysine residue at nucleotide position 345 is methylated in the trypanosome calmodulin but is not modified in the bovine brain protein.

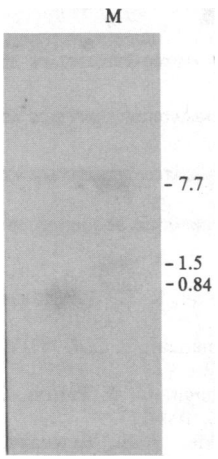

FIG. 3. Blot-hybridization analysis of calmodulin mRNA expressed in trypanosomes. Bloodstream-form poly(A)$^+$ mRNA from TXTat 1 was electrophoresed on an agarose/formaldehyde gel, transferred to nitrocellulose, and hybridized to ^{32}P-labeled p512 DNA. λCM51 digested with *Hin*dIII (lane M) was used as molecular weight marker (shown in kb).

two mRNAs contain an identical 35-nucleotide-long leader sequence that is not encoded in the calmodulin locus or in adjacent genomic DNA (Fig. 5). We hybridized an oligonucleotide complementary to 16 nucleotides of this leader to λCM51 and λCM55 and did not detect any homology (results not shown). This same leader sequence is found at the 5′ terminus of all mRNAs coding for the VSGs (16, 17).

Two Different Sites Are Used for the Fusion of the Leader Sequence. Fig. 5 shows the nucleotide sequences of cDNA 1, cDNA 2, and the genomic DNA of gene *B* upstream from the AUG. The sequence obtained from cDNA 1 is identical to the 5′ noncoding region up to position −49. The genomic sequence at the divergence point (indicated by a slash) with the mRNA is 5′ T-T-G-T-G-T-G-A-A-G/G 3′. Comparison of the cDNA 2 sequence with the genomic sequence of gene *B* showed that 69 nucleotides upstream from the AUG are identical. The fusion of the leader occurs within the sequence 5′ G-C-T-T-T-C-C-A-A-G/T 3′. The two fusion points are not extensively homologous to each other. The attachment of the leader to the two sites conforms to the G-T....A-G rule for nucleotides immediately flanking exon borders. However, further flanking sequences are not in agreement with favored nucleotide frequencies for the 3′ splice site of eukaryotic genes (18). It is also worth mentioning that in gene *A* an extra A-G is present in the attachment site for cDNA 1, since it has been noted that this dinucleotide never occurs in the −15 through −5 region of an acceptor.

DISCUSSION

We have examined the calmodulin genes in *T. b. gambiense* and found that they occur in the form of three tandemly repeated segments. Each unit is composed of a translated region and an intergenic, or spacer region, which separates one gene from the next. The significance of this specific arrangement is not clear at this point. Analysis of calmodulin genomic and cDNA clones from chicken (19) and *Xenopus* (20) revealed that at least two different, nonallelic genes are present in these genomes. The unusual organization in trypanosomes may be related to the control of calmodulin gene expression (see below). The observation that the two intergenic regions and the 3′ flanking region of gene *C* hybridize to mRNA isolated from bloodstream trypanosomes (unpublished data) suggests that at least two calmodulin genes are expressed. Whether trypanosomes at different stages of the life cycle also contain three tandemly repeated genes which are transcribed in a similar way remains to be determined.

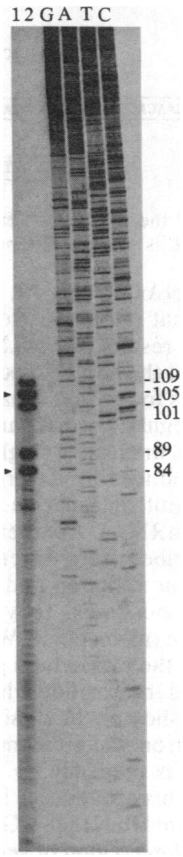

FIG. 4. Primer extension analysis of the 5′ end of calmodulin mRNA. A synthetic oligonucleotide complementary to position −21 to −7 of the 5′ noncoding region was used to generate extension products on mRNA templates as described. The reaction mixture for lane 1 contained TXTat 1 poly(A)$^+$ RNA and reverse transcriptase and for lane 2 contained reverse transcriptase but no RNA. A parallel dideoxynucleotide sequencing reaction was carried out with the same primer and a M13 mp8 subclone of the 5′ noncoding region of gene *B* (see Fig. 2). The nucleotide sequence obtained by this method is complementary to the one shown in Fig. 2 and starts at position 809. The sizes (in nucleotides) of the five extension products are given with respect to the A of the ATG initiation codon. The arrowheads indicate the extension products that contain the cDNA 1 (lower arrowhead) and cDNA 2 (upper arrowhead) sequence.

Our sequence analysis showed a remarkable degree of homology between the three repeat units (Fig. 2). We were surprised to detect only three base changes located in the intergenic region. This extraordinary sequence conservation could mean that a strong selective pressure operates in trypanosomes to maintain sequence homogeneity within tandemly repeated sequences. Alternatively, an amplification of the calmodulin genes could have occurred very recently and divergent sequences did not yet accumulate. In view of the fact that VSG genes undergo rearrangements in African trypanosomes, it is possible that non-VSG genes are also subject to rearrangements. It would also be interesting to see whether this tandem organization of calmodulin genes is present in other trypanosomes.

Our second main result is that a 35-nucleotide long leader sequence is present at the 5′ end of two calmodulin messenger RNAs. This sequence, called the spliced leader (SL), is not detectably linked to the calmodulin locus but is encoded in a tandemly repeated DNA segment (21, 22). De Lange *et al.* (21) suggested that the cluster of SL repeat units functions as a multiple promoter for VSG gene transcription. Instead, recent evidence indicated that the leader sequence is not unique to VSG mRNA but is attached to the 5′ end of

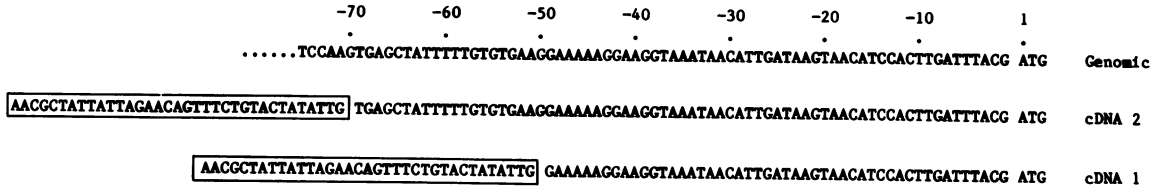

FIG. 5. Sequences of the two cDNA fragments and the corresponding genomic DNA. The genomic sequence upstream from the ATG initiation codon of gene *B* is shown. The nucleotide sequence of the spliced leader is boxed.

many trypanosome RNAs (23–25, 29, 30). Our experiments clearly demonstrate that the expression of structural genes other than VSG genes results in a mRNA containing the SL sequence. The fact that the two calmodulin mRNAs differ in the site where the SL is found in the mature transcript can be explained by several regulatory mechanisms. One possibility is that transcription occurs from a single gene and that either the two RNAs are produced by differential processing of a single primary transcript or separate promoters direct the synthesis of different mRNAs. Another possibility is that the two RNAs are transcribed from different genes. While the signals for transcription initiation and RNA processing are not yet clear in trypanosomes, they might lie in the sequences that flank each fusion point. We have compared the genomic sequences of the calmodulin genes *A* and *B* and of four VSG genes around the position where the SL is found in the RNA (results not shown). In all six sequences, the two nucleotides upstream from the attachment site are A-G. The invariant dinucleotide is preceded by a C+T-rich region in all genes except one. This sequence is located between 1 and 11 nucleotides upstream from the A-G. The location of this sequence in the 5′ flanking region of gene *A* is different when compared with that of gene *B* or *C* and correlates well with the position of the fusion points. This interpretation also suggests that gene *A* produces mRNA 1 and that gene *B* and/or gene *C* transcribes mRNA 2. This conserved motif might be important in modulating the level of transcription as suggested for yeast genes (26), in splicing, or in both.

Regulation of gene expression through the production of mRNAs with identical coding but different 5′ noncoding sequences has been described for the mouse α-amylase gene *Amy-1*^{a} (27) and the yeast invertase gene *SUC2* (28). The calmodulin mRNAs contain identical coding regions and the same 35-nucleotide 5′ terminal sequence. However, cDNA 2 differs from cDNA 1 by 20 nucleotides of 5′ untranslated sequence. Whether or not calmodulin-dependent Ca^{2+} metabolism in trypanosomes is regulated in the manner described above is unknown at present. Nevertheless, Ca^{2+} metabolism is known to be intimately involved with motility in these parasites, and there may well be different motility requirements in procyclic and metacyclic trypanosomes of the tsetse vector and in the bloodstream trypanosome forms of the mammalian host. We recently have shown that bloodstream trypanosomes contain a larger concentration of calmodulin (3). It is not unlikely, therefore, that calmodulin genes may be expressed differentially at different stages of the African trypanosome life cycle.

We thank Anthony Means and his colleagues for providing the eel calmodulin cDNA clone and we are very grateful to Peter Walter for the synthesis of the oligonucleotide. We thank Elisabetta Ullu and David Solnick for critical comments on the manuscript. This work was supported by U.S. Public Health Service Grants AI-08614, AI-15742, and AI-07136 and by the MacArthur Foundation. C.T. is a recipient of a fellowship from the James Hudson Brown–Alexander B. Coxe Foundation. The authors are investigators of the MacArthur Foundation Consortium on the Biology of Parasitic Diseases.

1. Parsons, M., Nelson, R. G. & Agabian, N. (1984) *Immunol. Today* **5**, 43–50.
2. Klee, C. B. & Vanaman, T. C. (1982) *Adv. Protein Chem.* **35**, 213–321.
3. Ruben, L., Egwuagu, C. & Patton, C. L. (1983) *Biochim. Biophys. Acta* **758**, 104–113.
4. Ruben, L., Strickler, J. E., Egwuagu, C. & Patton, C. L. (1984) in *Molecular Biology of Host-Parasite Interactions*, eds. Agabian, N. & Eisen, H. (Liss, New York), pp. 267–278.
5. Merritt, S. C., Tschudi, C., Konigsberg, W. H. & Richards, F. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1536–1540.
6. Lanham, S. M. (1968) *Nature (London)* **218**, 1273–1274.
7. Frischauf, A.-M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170**, 827–842.
8. Scherer, G., Telford, J., Baldari, C. & Pirrotta, V. (1981) *Dev. Biol.* **86**, 438–447.
9. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
10. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
11. Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
12. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
13. Munjaal, R. P., Chandra, T., Woo, S. L. C., Dedman, J. R. & Means, A. R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2330–2334.
14. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
15. Kasai, H., Kato, Y., Isobe, T., Kawasaki, H. & Okuyama, T. (1980) *Biomed. Res.* **1**, 248–264.
16. Boothroyd, J. C. & Cross, G. A. M. (1982) *Gene* **20**, 281–289.
17. Van der Ploeg, L. H. T., Liu, A. Y. C., Michels, P. A. M., De Lange, T., Borst, P., Majumder, H. K., Weber, H., Veeneman, G. H. & Van Boom, J. (1982) *Nucleic Acids Res.* **10**, 3591–3604.
18. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
19. Stein, J. P., Munjaal, R. P., Lagace, L., Lai, E. C., O'Malley, B. W. & Means, A. R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6485–6489.
20. Chien, Y.-H. & Dawid, I. B. (1984) *Mol. Cell. Biol.* **4**, 507–513.
21. 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.
22. Nelson, R. G., Parsons, M., Barr, P. J., Stuart, K., Selkirk, M. & Agabian, N. (1983) *Cell* **34**, 901–909.
23. Nelson, R. G., Parsons, M., Selkirk, M., Newport, G., Barr, P. J. & Agabian, N. (1984) *Nature (London)* **308**, 665–667.
24. Parsons, M., Nelson, R. G., Stuart, K. & Agabian, N. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 684–688.
25. Parsons, M., Nelson, R. G., Watkins, K. P. & Agabian, N. (1984) *Cell* **38**, 309–316.
26. Dobson, M. J., Tuite, M. F., Roberts, N. A., Kingsman, A. J., Kingsman, S. M., Perkins, R. E., Conroy, S. C., Dunbar, B. & Fothergill, L. A. (1982) *Nucleic Acids Res.* **10**, 2625–2637.
27. Schibler, U., Hagenbüchle, O., Wellauer, P. K. & Pittet, A. C. (1983) *Cell* **33**, 501–508.
28. Carlson, M. & Botstein, D. (1982) *Cell* **28**, 145–154.
29. De Lange, T., Michels, P. A. M., Veerman, H. J. G., Cornelissen, A. W. C. A. & Borst, P. (1984) *Nucleic Acids Res.* **12**, 3777–3790.
30. De Lange, T., Berkvens, T. M., Veerman, H. J. G., Frasch, A. C. C., Barry, J. D. & Borst, P. (1984) *Nucleic Acids Res.* **12**, 4431–4443.