Gene, 146 (1994) 245–250

GENE 08017

Characterization of a short interspersed reiterated DNA sequence of *Trypanosoma cruzi* located at the 3′-end of a poly(A)$^+$ transcript

(Expression analysis; sequence analysis; cDNA isolation; genomic distribution; (TAA)*n* repeats)

Jose M. Requena[a], Francisco Martín[b], Manuel Soto[a], Manuel C. López[b] and Carlos Alonso[a]

[a]*Centro de Biología Molecular Severo Ochoa, CSIC-UAM, Madrid, Spain; and* [b]*Instituto de Parasitología y Biomedicina, CSIC, Granada, Spain. Tel. (34-958) 200-403*

Received by M. Salas: 6 September 1993; Revised/Accepted: 23 December 1993/23 February 1994; Received at publishers: 11 April 1994

SUMMARY

We have carried out the molecular characterization of a highly repeated DNA element, called *E12*, from *Trypanosoma cruzi*, which has been found to be interspersed along its genome. The *E12* element, repeated about 5.6×10^5 times, is found in most of the chromosomal bands of the parasite. Three subregions may be defined within the element on the basis of sequence similarities with other trypanosome genomic sequences. Northern blot analysis demonstrated that sequences of the *E12* element are present in several polyadenylated RNA species of *T. cruzi*. The isolation and characterization of a cDNA clone, pSPFM55, which showed hybridization with the *E12* probe, indicated that only one of the *E12* subregions, *E12A*, is found in the cDNA and that it is located at the 3′-end providing the site of polyadenylation addition. The location and high degree of nucleotide conservation of *E12A* suggest a possible functional role of this sequence in gene expression.

INTRODUCTION

The analysis at the molecular level of eukaryotic nuclear DNAs has revealed the presence of complex arrays of highly repeated sequences (Singer, 1982). The function of the major repetitive DNA families in these genomes has not been determined although it has been reported that they could be involved in events of unequal crossingover (Dover, 1989) and that they could be implicated in the posttranscriptional control of gene expression as it is the case of the *Alu*-like (Deininger and Daniels, 1986) elements from the rat identifier (ID) family repeats (Glaichenhaus and Cuzin, 1987) and the repeated sequences from the mouse B1 family found at the 3′-end of noncoding regions, conferring growth-dependent expression (Vidal et al., 1993). The existence of genetic repetitive material has been also recently interpreted in the context of the abilities of eukaryotes to develop evolutionary strategies which allow comparatively quick changes in the interactions of individual parts of the genome (Epplen, 1992).

At present, several families of nuclear repetitive sequences have been isolated from the genome of *T. cruzi*, the etiological agent of Chagas' disease, as a result of an intense search for DNA probes with diagnostic applications. The first reported repetitive sequence of this parasite was a 'satellite' DNA element which accomplishes for 9% of its nuclear genomic DNA (Sloof et al., 1983; Gonzalez et al., 1984). Another repetitive *T. cruzi* DNA

Correspondence to: Dr. C. Alonso, Centro de Biología Molecular Severo Ochoa, Universidad Autónoma de Madrid, 28049 Madrid, Spain. Tel. (34-91) 397-4863; Fax (34-91) 397-4799; e-mail: JMREQUENA@MVAX.CBM.UAM.ES

Abbreviations: bp, base pair(s); *E12*, *T. cruzi* repeated sequence; *E12A* and *E12B*, subfragments of *E12* sequence; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); OFAGE, orthogonal-field alternation gel electrophoresis; p, plasmid; SDS, sodium dodecyl sulfate; *SPFM55*, cDNA containing *E12*-related sequences; SSC, 150 mM NaCl/15 mM Na_3citrate pH 7.0; *T.*, *Trypanosoma*; *UTR*, untranslated region(s); VSG, variant surface glycoprotein.

SSDI 0378-1119(94)00228-K

sequence element, called *E13*, has been reported to be dispersed throughout the genome as a family of polymorphic sequences with a copy number of $(1–2)\times10^4$ (Requena et al., 1992). Repetitive sequences represented in lower abundance have been, moreover, isolated from the *T. cruzi* genome (Wincker et al., 1990; de Mendonça-Lima and Traub-Cseko, 1991).

The aim of the present study was to isolate a new DNA fragment from the *T. cruzi* genome, called *E12*, that is repeated about 5500 times and located on most of the chromosomes. We show that a particular subregion of the *E12* element, *E12A*, is transcribed and found at the 3'-end of a polyadenylated transcript.

EXPERIMENTAL AND DISCUSSION

(a) Isolation of a *T. cruzi* repeated sequence

In search for repeated DNA sequences from the *T. cruzi* genome we screened an *Eco*RI-digested DNA library with a *T. cruzi* total DNA probe as previously described (Requena et al., 1993). Clones showing a strong hybridization intensity were isolated, nick-translated (Maniatis et al., 1982) and used to probe *T. cruzi* endonuclease-restricted DNA on nylon membranes. A complex pattern of hybridization bands was observed with the pE12 clone, independent of the restriction enzyme used (Fig. 1A), as an indication of the high level of dispersion and restriction polymorphism of this DNA fragment throughout the *T. cruzi* genome. In fact, the *Eco*RI band corresponding to the cloned *E12* DNA fragment has a relatively low autoradiographic signal (see arrow in Fig. 1A). Bands of higher size showing strong hybridization with the *E12* element may be explained as formed by tandem arrays of either the entire *E12* element lacking *Eco*RI restriction sites or of particular *E12*-containing sequences.

To study the chromosomal distribution of the *E12* element, *T. cruzi* chromosomes were size separated by orthogonal-field alternation gel electrophoresis (OFAGE), transferred to nylon membranes and probed with the insert of the pE12 clone (Fig. 1B). Most of the chromosomal bands hybridize with the *E12* probe as a direct indication that the *E12* sequence is dispersed along the *T. cruzi* genome. Dot-blot hybridization assays (Fig. 1C) indicated that the *E12* sequences comprise about 2.5% of *T. cruzi* total DNA. Since the *E12* element is 1123-nt long (see section **b**), assuming a *T. cruzi* genome content of 2.5×10^8 bp (Lanar et al., 1981), we may deduce that the copy number of this element is 5.6×10^3.

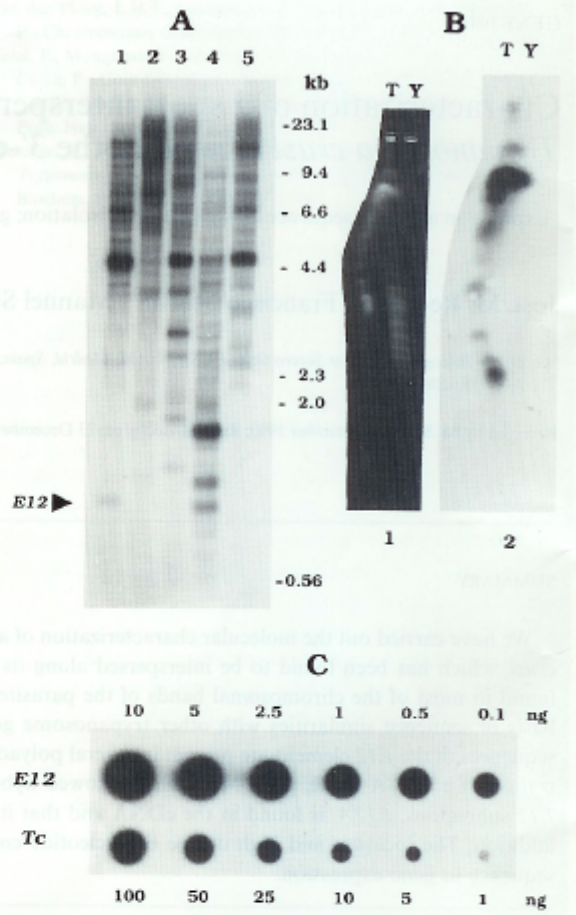

Fig. 1. Genomic organization, distribution and copy-number determination of the *E12* repeated element. (**A**) Southern hybridization of the DNA from *T. cruzi* parasites probed with the *E12* element. Lanes: 1, *Eco*RI; 2, *Bam*HI; 3, *Hin*dIII; 4, *Pst*I; 5, *Sma*I. The arrow indicates the position of the E12 DNA fragment. **Methods:** *T. cruzi* DNA was isolated as previously described (Requena et al., 1988). DNA samples (2 μg) were digested using the endonucleases listed above. DNA fragments were fractionated in a 0.8% agarose gel, transferred to a nylon membrane and probed with the [α-^{32}P]dATP-nick-translated insert of pE12 clone. Hybridizations were carried out at 42°C for 16 h in a solution containing 50% formamide/6 × SSC/1% SDS/60 μg denatured herring sperm DNA per ml. The last wash was done in stringent conditions in 0.1 × SSC/0.1% SDS at 65°C for 1 h. (**B**) Distribution of the E12 DNA sequences on *T. cruzi* chromosomes. Lanes: T, *T. cruzi*; Y, *Saccharomyces cerevisiae*. Panels: 1, ethidium bromide-stained gel; 2, Southern blot of the gel hybridized, as indicated above, with the *E12* probe. **Methods:** The preparation of the agarose blocks, parasite pellets and the OFAGE separation were carried out as described by Van der Ploeg et al. (1984). The pulse frequency was 90 s using 1% agarose gels. Southern blotting was achieved in the usual way (Maniatis et al., 1982) with a preliminary acid soak (0.25 M HCl, 30 min) to facilitate the transfer of large fragments. (**C**) *T. cruzi* genomic content of the *E12* repeated element. Known amounts of *E12* DNA fragments (*E12*) and *T. cruzi* DNA (Tc) were denatured by addition of 2 vols. of 0.5 M NaOH and incubation at 55°C for 10 min. Samples were dot-blotted onto a nylon membrane and hybridized with a *E12* probe as stated above. After exposure to the film the radioactivity bound to each DNA spot was quantitated by liquid scintillation counting. The amount of the probe sequence in the trypanosome DNA samples was estimated from a graph of the counts present in the *E12* spots.

(b) The nt sequence of the *E12* repeated element

In order to characterize the *E12* repetitive element of the *T. cruzi* genome the nt sequence of the *E12* element was determined by the dideoxy chain-termination method (Sanger et al., 1977) using the modifications introduced for sequencing of double-stranded DNA with the Sequenase kit (US Biochemical, Cleveland, OH, USA). A computer-assisted comparison of the *E12* nt sequence (Fig. 2a) with sequences from databanks indicated that there are three regions with similarities to previously reported *T. cruzi* genomic sequences (Fig. 2b). *E12A*, defined from nt 1 to nt 322 of the *E12* sequence, has an 81% of sequence identity with nt 4444 to 4712 (reverse and complementary) of the genomic sequence containing the *T. cruzi* neuraminidase-encoding gene (Pereira et al., 1991). A TAA-rich region, comprised by 17 repeats of the TAA motif, is located from nt 475 to 525. (TAA)*n*-rich motifs have been previously described in intergenic regions of both *T. brucei* (Glass et al., 1986) and *T. cruzi* (Requena et al., 1989) Hsp70 gene clusters and in sequences flanking several intrachromosomal variant surface glycoproteins (VSG)-encoding genes of *T. brucei* (Aline et al., 1985) which may be implicated in the gene conversion mechanism. Also, (TAA)*n* sequences has been implicated in the transcription reinitiation site that can mediate the expression of the *VSG MITat 117 TC* gene (Campbell et al., 1984). A third region, *E12B*, spanning from nt 1020 to 1123 has 90% of sequence identity with the first 104 nt of the *E13* repetitive element of *T. cruzi* (Requena et al., 1992).

(c) Transcription of the *E12* repeated element

In order to know whether the *E12* element forms part of transcribed molecules a Northern blot of polyadenylated and non-polyadenylated RNA from *T. cruzi* epimastigotes was probed with the ^{32}P-labelled *E12* fragment (Fig. 3). A prominent RNA band of about 5 kb together with three other bands of 1.3, 1.6 and 1.9 kb were observed in blots containing polyadenylated RNA samples (lane A^+). Labelled bands were absent from non-polyadenylated RNAs (lane A^-). In order to directly determine the position and the nt sequence of the *E12* element present in specific mRNAs a search for *E12* containing transcripts was performed on a *T. cruzi* cDNA library. A cDNA clone, pSPFM55A, was highly labelled by the ^{32}P-nick translation of *E12* fragment. Fig. 4a shows the restriction map of this cDNA, which is probably full-length because it has an estimated size of about 4.9 kb and the internal 0.8-kb *Sph*I fragment (Fig. 4) of the cDNA hybridizes in Northern blots with a mRNA species of about 5 kb (data not shown).

(d) Sequence comparisons

The entire *E12* element is not present in *SPFM55A* cDNA, since it only hybridizes with the 0.5-kb *Aat*II fragment located in the 3′ end (Fig. 4a). In order to determine the nt sequence of *SPFM55A* homologous to *E12*, both strands of the *Aat*II fragment of pSPFM55A were sequenced by the dideoxy chain-termination method. Fig. 4b shows that the homology between the nt sequence of the 3′-end of the cDNA with subregion *E12A* and with

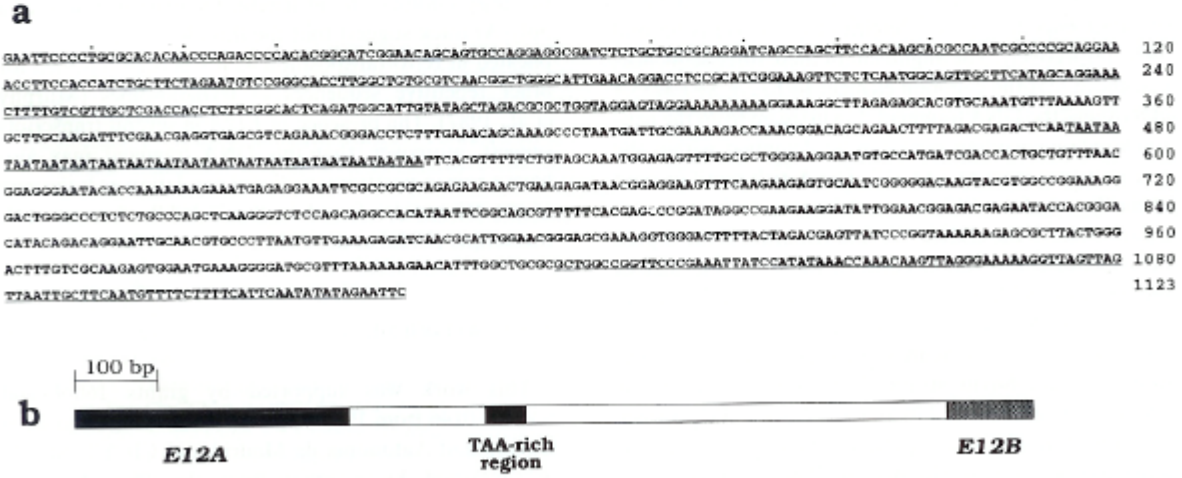

Fig. 2. Molecular characterization of *E12* element. **(a)** The nt sequence of the *T. cruzi E12* repeated element. Regions with nt identity to previously reported *T. cruzi* genomic sequences are underlined. **(b)** Schematic representation of the *E12* element. **E12A**, sequence with similarity to a DNA fragment located downstream to the *T. cruzi* neuraminidase-encoding gene (Pereira et al., 1991); TAA-rich region composed of 17 repeats of the TAA trinucleotide; **E12B**, sequence with similarity to a previously characterized repeated *E13* element from *T. cruzi* (Requena et al., 1992). The sequence reported in this paper have been deposited in the GenBank data base (accession No. L22304).

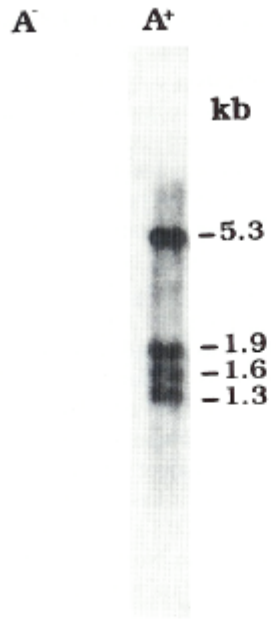

Fig. 3. *E12*-containing transcripts. RNA was isolated from epimastigotes of the Maracay strain of *T. cruzi* by the method of Chomczynski and Sacchi (1987). Polyadenylated RNAs were isolated on oligo(dT)-cellulose (Boehringer-Mannheim, Barcelona, Spain) according to the supplier's method. The poly(A)$^-$ RNA (10 μg, lane A$^-$) and of poly(A)$^+$ RNA (2 μg, lane A$^+$) were separated on 1% agarose gel containing 2.2 M formaldehyde in 1× MOPS buffer (40 mM MOPS/5 mM Na-acetate/1 mM EDTA pH 7.5). The RNA was transferred to Hybon-N membranes (Amersham Iberica, Madrid, Spain). Hybridization conditions are described in the legend to Fig. 1. The *Eco*RI insert of the pE12 clone was used as probe. The size of the hybridization bands in kb are shown on the right margin.

a region located downstream from the *T. cruzi* neuroaminidase-encoding gene (Pereira et al., 1991) is 95 and 79%, respectively. The alignment of these three sequences evidenced that the common element to *E12A*, *SPFM55A* cDNA and the DNA fragment downstream from the neuroaminidase-encoding gene can be delimited within the *E12* sequence from nt 7 to 317. The A-rich motif located at the 3′-end of the *E12A* at nt 317 serves as the polyadenylation site of the cDNA. We find remarkable the degree of sequence conservation between the *E12A* element of the E12 isolated from the Maracay strain, the *E12A* element of the *SPFM55A* cDNA and the *E12A* element found in the genomic neuroaminidase clone 7F, isolated from the San Silvio strain (Pereira et al., 1991). A common feature to *E12A* in all these cases is the presence of a poly(A) track similar to that found in RIME repeated sequences of *T. brucei* (Hasan et al., 1984) associated with retrotransposable elements (Murphy et al., 1987). Since the pattern of hybridization bands observed after digestion of DNA with several restriction enzymes and hybridization with *E12A* is similar to that obtained when the entire *E12* element was used as probe it is likely that *E12A* is the most repetitive fraction of the entire element (data not shown).

(e) Conclusions

(*1*) A highly repetitive sequence, named *E12*, from the *T. cruzi* genomic DNA has been found to be dispersed among most of the chromosomal bands representing about 2.5% of the total DNA.

(*2*) Sequence analysis of the repeated element has shown that, on the basis of sequence homology with previously reported genomic DNA fragments of this parasite, it has a complex structure in which three subdomains can be defined. Subregion *E12A* has about 80% of sequence identity with a sequence located downstream from the neuraminidase-encoding gene (Pereira et al., 1991). Subregion *E12B* has about 90% of sequence identity with another highly repeated element of *T. cruzi* (Requena et al., 1992). The third subregion is defined by a succession of 17 TAA repeats.

(*3*) The *E12* element is transcribed and it is present in several poly(A)$^+$ transcripts. A cDNA clone (pSPFM55A), containing 95% of sequence identity with the *E12A* subregion, was isolated. The *E12A* element defines the polyadenylation site of the transcript.

(*4*) The *E12A* element is markedly conserved since comparison of the sequence of this element with that of two other sequences located in different genetic environments revealed strong similarities in both sequence identity and length. Data are shown suggesting that the *E12A* subregion is the most highly repetitive fraction of the *E12* element. The structural characteristics of *E12A* such as the copy number, the nt length, the existence in 3′ of a poly(A) track suggest that the this element could be implicated in retrotransposition as it has been reported for other repetitive sequences of trypanosomes (Murphy et al., 1987). The presence of *E12A* in several transcripts suggest, moreover, that this DNA fragment may also have a role in posttranscriptional regulation as it the case of the B1 repeats found in the 3′-*UTR* of mouse pIL transcripts (Vidal et al., 1993).

ACKNOWLEDGEMENTS

This work was supported by grants 160/9 and SAF93–0146 from Plan Regional de Investigación de la Comunidad Autónoma de Madrid and CICYT. M.C.L. is supported by a grant from the Plan Nacional (BIO-90–0786). The support of LETI, S.A. and the institutional grant of the Fundación Ramón Areces are also gratefully acknowledged. M.S. is supported by a doctoral fellowship from Comunidad Autónoma de Madrid.

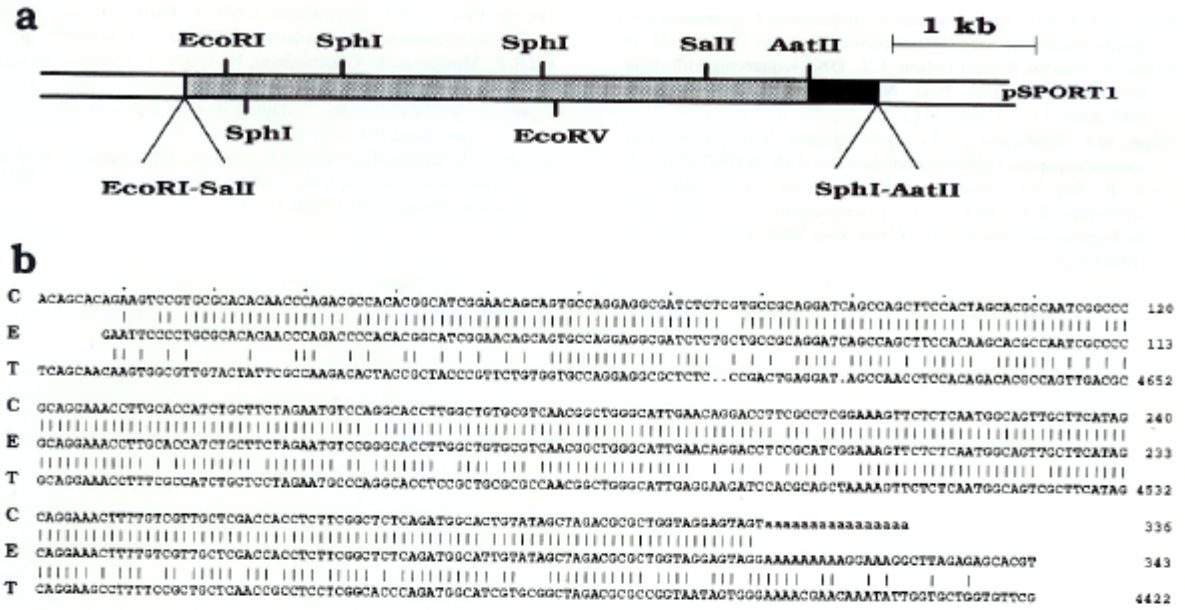

Fig. 4. Analysis of *E12A* element. (a) Restriction map of the pSPFM55 cDNA clone. The cDNA library was constructed from *T. cruzi* poly(A)$^+$ RNA using the Superscript Plasmid System for cDNA synthesis and the Plasmid Cloning kit (BRL, Gaithersburg, MD, USA). The direction of the MCS (polylinker) sequence of the pSPORT1 plasmid, used for the construction of the library, is indicated. The black box indicates the region of the *SPFM55* cDNA that hybridizes with *E12*. (b) Comparison of the nt sequence of the repeated *E12A* element. The best alignment was obtained between the 336 nt of the 3'-end of the *SPFM55* cDNA (C), the first 343 nt of the *E12* element (E) and nt 4768 to 4422 of the *T. cruzi* genomic clone 7F (T, Trbtenaa; Pereira et al., 1991). The poly(A) tail found in the *SPFM55* cDNA has been indicated in lower-case letters. Identical nt are indicated by vertical bars.

REFERENCES

Aline Jr., R., MacDonald, G., Brown, E., Allison, J., Myler, P., Rothwell, V. and Stuart, K.: (TAA)n within sequences flanking several intrachromosomal variant surface glycoprotein genes in *Trypanosoma brucei*. Nucleic Acids Res. 13 (1985) 3161–3177.

Campbell, D.A., Van Bree, M.P. and Boothroyd, J.C.: The 5'-limit of transposition and upstream barren region of a trypanosome VSG gen: tandem 76 base-pair repeats flanking (TAA)n. Nucleic Acids Res. 12 (1984) 2759–2774.

Chomczynski, P. and Sacchi, M.: Single-step method of RNA extraction by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162 (1987) 156–159.

Deininger, P.L. and Daniels, G.R.: The recent evolution of mammalian repetitive DNA elements. Trends Genet. 2 (1986) 76–80.

de Mendonça-Lima, L. and Traub-Cseko, Y.M.: A new repetitive DNA sequence from *Trypanosoma cruzi*. Mem. Inst. Oswaldo Cruz 86 (1991) 475–475.

Dover, G.A.: Victims of perpetrators of DNA turnover? Nature 342 (1989) 347–348.

Epplen, J.T.: Diagnostic applications of repetitive DNA sequences. J. Int. Fed. Clin. Chem. 4 (1992) 228–234.

Glaichenhaus, N. and Cuzin, F.: A role for ID repetitive sequences in growth- and transformation-dependent regulation of gene expression in rat fibroblasts. Cell 50 (1987) 1081–1089.

Glass, D.J., Polvere, R.I. and Van der Ploeg, L.H.T.: Conserved sequences and transcription of the hsp70 gene family in *Trypanosoma brucei*. Mol. Cell. Biol. 6 (1986) 4657–4666.

Gonzalez, A., Prediger, E., Huecas, M.E., Nogueira, N. and Lizardi, P.M.: Minichromosomal repetitive DNA in *Trypanosoma cruzi*: its use in a high-sensitivity parasite detection assay. Proc. Natl. Acad. Sci. USA 81 (1984) 3356–3360.

Hasan, G., Turner, M.J. and Cordingley, J.S.: Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei*. Cell 37 (1984) 333–341.

Lanar, D.E., Levy, L.S. and Manning, J.E.: Complexity and content of the DNA and RNA in *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 3 (1981) 327–341.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Murphy, N.B., Pays, A., Tebabi, P., Coquelet, H., Guyaux, M., Steinert, M. and Pays, E.: *Trypanosoma brucei* repeated element with unusual structural and transcriptional properties. J. Mol. Biol. 195 (1987) 855–871.

Pereira, M.E.A., Mejia, J.S., Ortega-Barria, E., Matzilevich, D. and Prioli, R.P.: The *Trypanosoma cruzi* neuraminidase contains sequences similar to bacterial neuraminidases, YWTD repeats of the low density lipoprotein receptor, and type III modules of fibronectin. J. Exp. Med. 174 (1991) 179–191.

Requena, J.M., Jimenez-Ruiz, A., Soto, M., Lopez, M.C. and Alonso, C.: Characterization of a highly repeated interspersed DNA sequence of *Trypanosoma cruzi*: its potential use in diagnosis and strain classification. Mol. Biochem. Parasitol. 51 (1992) 271–280.

Requena, J.M., Lopez, M.C., Jimenez-Ruiz, A., de la Torre, J.C. and Alonso, C.: A head-to-tail tandem organization of hsp70 genes in *Trypanosoma cruzi*. Nucleic Acids Res. 16 (1988) 1393–1406.

Requena, J.M., Lopez, M.C., Jimenez-Ruiz, A., Morales, G. and Alonso, C.: Complete nucleotide sequence of the hsp70 gene of *T. cruzi*. Nucleic Acids Res. 17 (1989) 797–797.

Requena, J.M., Soto, M. and Alonso, C.: Isolation of *Trypanosoma cruzi* specific nuclear repeated DNA sequences. Biol. Res. 26 (1993) 11–18.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Singer, M.F.: SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell 28 (1982) 433–434.

Sloof, P., Bos, J.L., Konings, A.F.J.M., Menke, H.H., Borst, P., Gutteridge, W.E. and Leon, W.: Characterization of satellite DNA in *Trypanosoma brucei* and *Trypanosoma cruzi*. J. Mol. Biol. 167 (1983) 1–21.

Van der Ploeg, L.H.T., Cornelissen, A.W.C.A., Barry, J.D. and Borst, P.: Chromosomes of kinetoplastida. EMBO J. 3 (1984) 3109–3115.

Vidal, F., Mougneau, E., Glaichenhaus, N., Vaigot, P., Darmon, M. and Cuzin, F.: Coordinated posttranscriptional control of gene expression by modular elements including *Alu*-like repetitive sequences. Proc. Natl. Acad. Sci. USA 90 (1993) 208–212.

Wincker, P., Roizes, G. and Goldenberg, S.: Characterization of a *Trypanosoma cruzi* specific nuclear repeated sequence. Mol. Biochem. Parasitol. 41 (1990) 147–152.