SHORT COMMUNICATION

A Novel Reiterated Family of Transcribed Oligo(A)-terminated, Interspersed DNA Elements in the Genome of *Trypanosoma cruzi*

Newton Valério Verbisck, Márcia Regina Machado dos Santos, David M Engman*, Miguel Angel Chiurillo, José Luis Ramirez**, Jorge Enrique Araya***, Renato Arruda Mortara, José Franco da Silveira/+**

Departamento de Microbiologia, Imunologia e Parasitologia, Escola Paulista de Medicina, Rua Botucatu 862, 04023-062 São Paulo, SP, Brasil *Departments of Pathology and Microbiology-Immunology, Northwestern University Medical School, Chigago, Illinois, USA **Laboratorio de Genetica Molecular, Instituto de Biologia Experimental, Universidad Central de Venezuela, Caracas, Venezuela ***Departamento de Tecnologia Médica, Universidad de Antofagasta, Antofagasta, Chile

We report the molecular characterization of a novel reiterated family of transcribed oligo(A)-terminated, interspersed DNA elements in the genome of Trypanosoma cruzi. *Steady-state level of transcripts of this sequence family appeared to be developmentally regulated, since only in the replicative forms the parasite showed expression of related sequences with a major band around 3 kb. The presence of frame shifts or premature stop codons predicts that transcripts are not translated. The sequence family also contains truncated forms of retrotransposons elements that may become potential hot spots for retroelement insertion. Sequences homologous to this family are interspersed at many chromosomes including the subtelomeric regions.*

Key words: *Trypanosoma cruzi* - interspersed repetitive DNA - transcription - retroelement insertion - subtelomeric region

The genome of the protozoan parasite *Trypanosoma cruzi*, the etiological agent of Chagas disease, contains a large number of repeated sequences which may play an important role in genetic rearrangements and control of gene expression (Requena et al. 1996). DNA reassociation kinetic studies have shown that highly and middle repetitive sequences account for nearly 44% of *T. cruzi* nuclear genome (Castro et al. 1981, Lanar et al. 1981). The repetitious DNA fraction is composed by micro-and minisatellites, short and long interspersed nucleotide elements (SINE and LINE), LTR (long terminal repeat)- and non-LTR retrotransposons (Martin et al.1995, Requena et al. 1996, Araya et al. 1997, Oliveira et al. 1998, Vazquez et al. 1999, 2000, Olivares et al. 2000). Several *T. cruzi* repetitive elements, such as SINE, LTR- and non-LTR retrotransposons, can be actively transcribed into the poly(A^+) RNA fraction (Cotrim et al. 1989, Requena et al. 1994, Martin et al. 1995, Vazquez et al. 1999, 2000, Olivares et al. 2000). The role of these many repeated elements in *T. cruzi* is still unknown, and it has been speculated that they are involved in shaping and reshaping of the genome causing ectopic rearrangements, modifying and reshuffling existing genes and creating new genes. They can also affect the expression of 5' and 3' adjacent genes by affecting the trans-splicing reaction efficiency (Vazquez et al. 1994, 1999). In this report we characterized a novel sequence family of transcribed oligo(A)-terminated sequences whose members are spread throughout the *T. cruzi* genome. Although actively transcribed and processed, these sequences had premature stop codons that suggested a lack of translation.

In the course of studying genes encoding surface antigens of *T. cruzi*, a cDNA library constructed in phage λ UNI-ZAP XR with poly(A^+) mRNA extracted from intracellular amastigotes (clone Sylvio X10/4) was screened with the ^{32}P-labelled insert of a cDNA clone named Tt34c1. Clone Tt34c1 (5488 bp) contains an open reading frame (ORF) for a 85 kDa surface glycoprotein (gp85) and a large 3'-untranslated region (3'-UTR) of 2812 bp but lacks the miniexon and the poly(A) tail (Takle & Cross 1991).

According to identity hits in GenBank two types of clones were isolated in these experiments: (a) cDNAs encoding gp85-like glycoproteins (clone TcSx23), and (b) cDNAs sharing similarity with the 3'-UTR of clone Tt34c1 (clones TcSx12, TcSx38 and TcSx42). In this study we further characterized the recombinants TcSx12, TcSx38 and TcSx42. Nucleotide sequences reported in this paper are available in the GenBank database under accession numbers: AF510088 (TcSx12), AF510086 (TcSx38), AF510087 (TcSx42) and AF510037 (TcSx23). Recombinant inserts ranged from 1.5 to 2.1 kb in length, all contained a poly(A) tail and shared considerable identity (~ 92%) at nucleotide level with the 3'-UTR of clone Tt34c1 (Fig. 1A). Sequence comparison among recombinant cDNAs revealed several differences such as the presence of a common ~ 700 bp sequence preceding the poly(A) tail in TcSx38 and TcSx42 but not in TcSx12. In addition, clone TcSx42

This work was supported by grants from Fapesp, Capes (Brazil) and Cyted (España) to JFS, Conicit-GR 9900036 (Venezuela) to JLR. NVV was recipient of a doctoral fellowship from CNPq (Brasil).
+Corresponding author. Fax: +55-11-5571.1095. E-mail: franco@ecb.epm.br
Received 12 August 2002
Accepted 5 September 2002

A

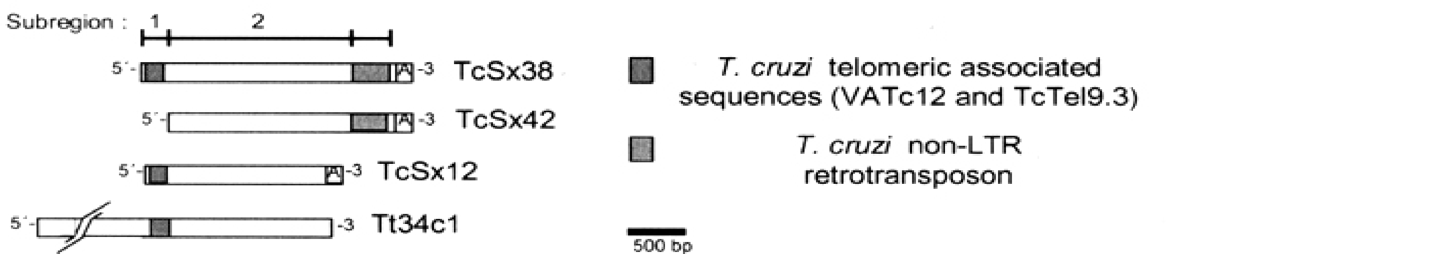

Fig. 1: sequence analysis of cDNA repeated elements. A: comparison of the cDNA sequences: TcSx38 (AF510086), TcSx42 (AF510087) and TcSx12 (AF510088) with previously reported *Trypanosoma cruzi* sequences Tt34c1 (M64836), VATc12 (AF100653), TcTel9.3 (AF100648), RS1Tc (AF208540) and RS13Tc (AF208538). GenBank accession numbers are in parentheses. Alignments were done by Clustal W with MegAlign program (DNASTAR Inc.). Ligth gray indicates a threshold for conservation of 100%; B: schematic representation of cDNAs TcSx38, TcSx42 and TcSx12 showing the different subregions. Subregion 1, sequence with similarity to the subtelomeric associated repeats VATc12 and TcTel9.3 (Chiurillo et al. 1999); Subregion 2, sequence with similarity to 3'-UTR of clone Tt34c1 (Takle & Cross 1991); Subregion 3, sequence with similarity to the non-LTR retrotransposons RS13Tc and RS1Tc (Olivares et al. 2000). Letter A within the box denotes the poly(A) tail.

has a 157-bp sequence (positions 522 to 679) that is lacking in clone TcSx38. These differences indicated that the three cDNAs originated from different sequence copies. The most striking feature of these transcripts was the presence of multiple stop codons in each frame, indicating that they are not translated.

Fig. 1 shows a sequence comparison of cDNAs TcSx12, TcSx38 and TcSx42 with other *T. cruzi* repetitive sequences reported in the literature. In TcSx38 we identified two subregions (1 and 3) that showed sequence similarity with previously reported *T. cruzi* repetitive sequences. Subregion 1, from nt 14 to nt 134, has 82.5% and 86% of sequence identity with *T. cruzi* telomere-associated sequences TcTel9.3 and VATc12, respectively (Chiurillo et al. 1999). Subregion 3, from nt 1570 to nt 2119, presents 50% and 87% of sequence identity with the non-LTR retrotransposons RS13Tc and RS1Tc, respectively (Olivares et al. 2000). In subregion 3 there is also a 41 nt sequence (nt 1592 to 1631) that displays 90% of identity with the 3' end of the E13 repeated element (Requena et al. 1992, Olivares et al. 2000). Subregions 1 and 3 flank a central region of 1434 bp (subregion 2) which is homologous to the 3'-UTR of clone Tt34c1. Clone TcSx12 contains subregions 1 and 2 but lacked subregion 3 which is part of the ~ 700 bp sequence found in TcSx38 and TcSx42, but absent in TcSx12. On the other hand, TcSx42 has subregions 2 and 3 but not subregion 1, this may have been lost during cDNA cloning. It is noteworthy that the 3'-UTR of clone Tt34c1 (Takle & Cross 1991) contains sequences with similarity with the VAT and TcTel telomere-associated sequences but not with non-LTR retrotransposons RS1Tc and RS13Tc. TcSx38 and TcSx42 contain truncated forms of retrotransposons RS1Tc and RS13Tc sequences that have been associated with a hot spot for retroelement insertion in RHS (retrotransposon hot spot) multigene family found in *T. cruzi* and *T. brucei* (Bringaud et al. 2002).

Steady-state levels of RNAs related to the recombinants were determined by Northern blot analysis using the insert of recombinant TcSx38 as a probe (Fig. 2A). The probe strongly hybridized with a ~ 3 kb transcript and less intensely with two transcripts of around 8 and 1.3 kb in epimastigotes and intracellular amastigotes. A faint hybridization signal was detected with RNAs isolated from bloodstream and metacyclic trypomastigotes. Confirming that TcSx38-related sequences are expressed in the epimastigote forms, a nucleotide sequence identity search of EST (expressed tag sequence) databases with TcSx38 sequence, revealed a high percentage identity with 23 ESTs isolated from a *T. cruzi* epimastigote cDNA library (Urmenyi et al. 1999). For comparative purposes, when the cDNA TcSx23, which encodes a gp85-like protein, was used as a probe, it hybridized with a ~ 4.5 kb mRNA present in bloodstream trypomastigotes and intracellular amastigotes (Fig. 2C). From these results we concluded that steady-state level of transcripts of TcSx38 is developmentally regulated and differs from that of TcSx23. Taken together, our findings suggested that TcSx12, TcSx38 and TcSx42 belong to a new *T. cruzi* gene family whose members are transcribed as oligo(A)-terminated sequences and processed to mature RNAs.

To further characterize genomic sequences associated to the TcSx38, a Southern blot of genomic DNA digested with several restriction enzymes was probed with the insert of this clone (Fig. 2D). The probe hybridized to multiple genomic bands, suggesting the existence of related sequences arranged in a non tandem manner. A complex hybridization pattern was also obtained for clone CL Brener, the reference clone of *T. cruzi* Genome Project (Fig. 2D). Chromosomal location of TcSx38 sequences was determined by hybridization with pulsed field gel electrophoresis (PFGE) separated chromosomal bands of clone Sylvio X10/4 (Cano et al. 1995). Implying that copies of TcSx38 are dispersed throughout the genome (rather than concentrated at a single locus), and consistent with the results from the Southern blot analysis, the probe hybridized with eight chromosomal bands of clone Sylvio X10/4 (Fig. 2E) with varying intensities. As an indication that TcSx38-related sequences belong to a multigene family, a nucleotide sequence identity search on GSS (genomic survey sequence) databases with TcSx38 sequence revealed a high percentage identity with 82 GSSs. In fact, according to the equation proposed by Agüero et al. (2000) the copy number of TcSx38-related sequences was estimated to be 73 per haploid genome. For this calculation, we used 47 Mb as the haploid *T. cruzi* genome, the total number of GSS was 21327, and the size of TcSx38-related sequences was 2119 bp.

As described above, clones TcSx38 and TcSx12 also shared sequence similarity with VATc12 and TcTel9.3 sequences located at subtelomeric regions of *T. cruzi* chromosomes (Chiurillo et al. 1999). To further confirm the presence of TcSx38 sequences at the subtelomeric regions, we have screened a *T. cruzi* telomeric library constructed in pBelo BAC vector (Chiurillo et al. 2002) with a specific probe of TcSx38 clone (nt 843 to 1767) devoid of VAT or TcTel sequences. BAC telomeric clones (n = 576) were stamped on nylon filters organized in cells of six colonies each and hybridized with TcSx38 probe. Assuming that no bias existed in the cloning and selection procedures, the number of stamped recombinants is slightly higher than a 3-fold representation of each chromosomal end. Positive and negative hybridization controls were included. Thirteen clones strongly hybridized with the probe indicating the presence of TcSx38-related sequences in the subtelomeric regions of *T. cruzi* chromosomes. The high recombination frequencies at subtelomeric regions may be favorable for the rapid generation of novel variants of this multigene family.

In conclusion, our results indicate that TcSx12, TcSx38 and TcSx42 recombinants belong to a new reiterated *T. cruzi* gene family. They are conserved in sequence, and although their transcripts are polyadenylated, the presence of frame shifts or premature stop codons preclude them from being translated into proteins. Northern blot analysis showed that the steady-state level of transcripts of this gene family is developmentally regulated and mainly expressed in the replicative forms of the parasite (epimastigotes and amastigotes). Members of this multigene family are associated with mobile elements and may be undergoing rapid evolution by recombination and sequence divergence. Our observations suggest that these

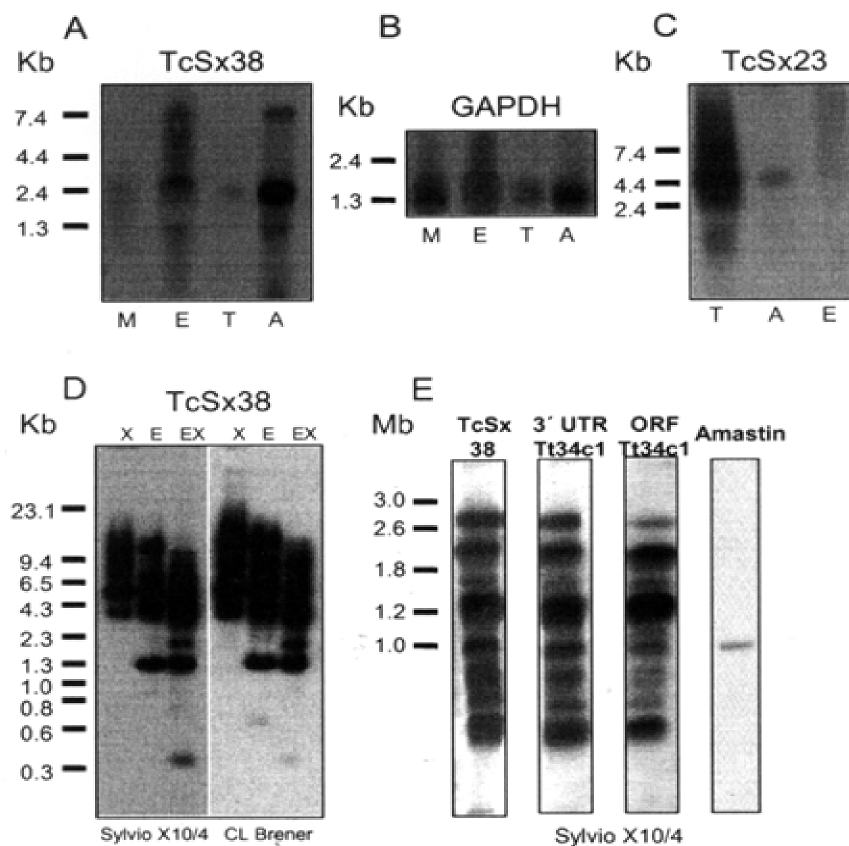

Fig. 2: transcription and genomic distribution of TcSx38 repeated element. A-C: steady-state levels of TcSx38 transcripts in the different developmental stages of *Ttypanosoma cruzi* (clone Sylvio X10/4). Ten µg of total RNA from metacyclic trypomastigotes (M), epimastigotes (E), cell culture trypomastigotes (T) and intracellular amastigotes (A) were submitted to electrophoresis on agarose/formaldehyde gel, transferred onto nylon membrane and hybridized with different probes. The following probes were used: *T. cruzi* glycosomal glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene and the inserts of cDNAs TcSx38 and TcSx23. A and B show the same blot that was sequentially hybridized to probes TcSx38 and GAPDH. Size markers (in kilobases, kb) are shown on the left; D: southern blot of restriction endonuclease digestions of *T. cruzi* (clones Sylvio X10/4 and CL Brener) genomic DNA probed with the insert of clone TcSx38. Restriction enzymes used were: X (*Xho*I), E (*Eco*RI) and XE (double digestion with *Xho*I and *Eco*RI). Size markers (in kilobases, kb) are shown on the left; E: chromosomal mapping of TcSx38-related sequences in the clone Sylvio X10/4. Chromosomal bands were separated by pulsed field gel electrophoresis (Cano et al. 1995), transferred onto nylon filters and hybridized to the following probes: the insert of clone TcSx38; two subfragments of clone Tt34c1 carrying the 3'-UTR and ORF of gp85 gene, respectively; the ORF of amastin, a *T. cruzi* amastigote stage-specific protein (Teixeira et al. 1994). Size markers (in megabases, Mb) are shown on the left. Northern and Southern blot hybridizations were carried out at 42ºC overnight in the presence of 50% formamide/5 x sodium saline citrate (5 x SSC = 0.75 M NaCl, 0.075 M sodium citrate)/5 x Denhardt's solution (0.1% Ficoll type 400, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin)/0.5% SDS/5 mM EDTA/0.1 mg tRNA per ml. Following the hybridization, the membranes were subjected to three washes (30 min each at 56ºC) in 2 x SSC containing 0.1% SDS, two additional washes at 56ºC in 0.1 x SSC containing 0.1% SDS.

cDNAs are not derived from read-through transcripts from putative promoters far upstream. Together with other reports of transcribed but not translated sequences, the present work adds more evidences to this puzzling phenomenom that may be a consequence of the all out polycistronic transcription of kinetoplastida, or else it is underlying a more important regulation mechanism for gene expression. The abundance of TcSx38 sequences in transcripts from different *T. cruzi* developmental forms could suggest that these sequences have a function in the trypanosome genome. Further studies of this sequence family are in progress aimed to unveil the possible function of these elements in the parasite.

ACKNOWLEDGEMENTS

To Dr GAM Cross and Dr SMR Teixeira for kindly providing clones Tt34c1 and amastin, respectively.

REFERENCES

Agüero F, Verdún RE, Frasch ACC, Sanchez DO 2000. A random sequencing approach for the analysis of the *Trypanosoma cruzi* genome: general structure, large gene and repetitive DNA families, and gene discovery. *Genome Res 10*: 1996-2005.

Araya JE, Cano MI, Gomes HBM, Novak EM, Requena JM, Alonso C, Levin, M, Guevara P, Ramires JL, Franco da Silveira J 1997. Characterization of an interspersed repetitive DNA element in the genome of *Trypanosoma cruzi*. *Parasitology 115*: 563-570.

Bringaud F, Biteau N, Melville SE, Hez S, El-Sayed NM, Leech V, Berriman M, Hall N, Donelson JE, Baltz T 2002. A new, expressed multigene family containing hot spot for insertion of retrolements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryot Cell 1*: 137-151.

Cano MI, Gruber A, Vazquez M, Cortés A, Levin MJ, Gonzalez A, Degrave W, Rondinelli E, Ramirez JL, Alonso C, Requena JM, Franco da Silveira J 1995. Molecular karyotype of clone CL Brener chosen for the *Trypanosoma cruzi* genome project. *Mol Biochem Parasitol 71*: 273-278.

Castro C, Craig SP, Castaneda M 1981. Genomic organization and ploidy number in *Trypanosoma cruzi*. *Mol Biochem Parasitol 4*: 273-282.

Chiurillo MA, Cano MI, Franco da Silveira J, Ramirez JL 1999. Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite *Trypanosoma cruzi*. *Mol Biochem Parasitol 10*: 173-182.

Chiurillo MA, Santos MRM, Franco da Silveira J, Ramírez JL 2002. A general improved approach for cloning and characterization of telomeres: the protozoan parasite *Trypanosoma cruzi* as model organism. *Gene* (in press).

Cotrim PC, Cummings LM, Franco da Silveira J 1989. Characterization of *Trypanosoma cruzi* genomic fragment complementary to several species-specific mRNAs but different from the spliced leader sequence. *Mem Inst Oswaldo Cruz 84*: 143-149.

Lanar DE, Levy SC, Manning JE 1981. Complexity and content of the DNA and RNA in *Trypanosoma cruzi*. *Mol Biochem Parasitol 3*: 327-341.

Martin F, Marañón C, Olivares M, Alonso C, López MC 1995. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the Ape family of DNA repair enzymes. *J Mol Biol 247*: 49-59.

Novak E, de Mello MP, Gomes HBM, Guevara P, Ramirez JL, Franco da Silveira J 1993. Repetitive sequences in the ribosomal intergenic spacer of *Trypanosoma cruzi. Mol Biochem Parasitol 60*: 273-280.

Olivares M, del Carmen Thomas M, Lopez-Barajas A, Requena JM, Garcia-Perez JL, Angel S, Alonso C, Lopez MC 2000. Genomic clustering of the *Trypanosoma cruzi* nonlong terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. *Electrophoresis 21*: 2973-2982.

Oliveira RP, Broude NE, Macedo AM, Cantor CR, Smith CL, Pena SD 1998. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. *Proc Natl Acad Sci USA 95*: 3776-3780.

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

Requena JM, Lopez MC, Alonso C 1996. Genomic repetitive DNA elements of *Trypanosoma cruzi*. *Parasitol Today 12*: 279-282.

Requena JM, Martin F, Soto M, Lopez MC, Alonso C 1994. Characterization of a short interspersed reiterated DNA sequence of *Trypanosoma cruzi* located at the 3'-end of a poly(A)+ transcript. *Gene 146*: 245-250.

Takle GB, Cross GAM 1991. An 85-kilodalton surface antigen gene family of *Trypanosoma cruzi* encodes polypeptides homologous to bacterial neuraminidases. *Mol Biochem Parasitol 48*: 185-198.

Teixeira SMR, Russell DG, Kirchhoff LV, Donelson JE 1994. A differentially expressed gene family encoding "Amastin", a surface protein of *Trypanosoma cruzi* amastigotes. *J Biol Chem 269*: 20509-20516.

Urmenyi TP, Bonaldo MF, Soares MB, Rondinelli E 1999. Construction of a normalized cDNA library for the *Trypanosoma cruzi* genome project. *J Eukaryot Microbiol 46*: 542-544.

Vazquez M, Ben-Dov C, Lorenzi H, Moore T, Schijman A, Levin MJ 2000. The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons. *Proc Natl Acad Sci USA 97*: 2128-2133.

Vazquez M, Lorenzi H, Schijman AG, Ben-Dov C, Levin MJ 1999. Analysis of the distribution of SIRE in the nuclear genome of *Trypanosoma cruzi. Gene 239*: 207-216.

Vazquez MP, Schijman AG, Levin MJ 1994. A short interspersed repetitive element provides a new 3' acceptor site for trans-splicing in certain ribosomal P2-protein genes of *Trypanosoma cruzi*. *Mol Biochem Parasitol 67*: 327-336.