SHORT COMMUNICATION

A mucin like gene different from the previously reported members of the mucin like gene families is transcribed in *Trypanosoma cruzi* but not in *Trypanosoma rangeli*

Teresa Abate/+, Mónica Rincón*, Zoraida Díaz-Bello, Lilian Spencer, Alexis Rodríguez-Acosta**

Instituto de Medicina Tropical "Dr. Félix Pifano" *Escuela de Biología, Facultad de Ciencias, Universidad Central de Venezuela, Apartado 47.706, Caracas 1041-A, DF, Venezuela **Departamento de Biología Celular, Universidad Simón Bolívar, Sartenejas, Caracas, Venezuela

Trypanosoma cruzi *expresses mucin like glycoproteins encoded by a complex multigene family. In this work, we report the transcription in* T. cruzi *but not in* T. rangeli *of a mucin type gene automatically annotated by the* T. cruzi *genome project. The gene showed no nucleotide similarities with the previously reported* T. cruzi *mucin like genes, although the computational analysis of the deduced protein showed that it has the characteristic features of mucins: a signal peptide sequence, O-glycosylation sites, and glycosylphosphatidylinositol (GPI) anchor sequence. The presence in this gene of N- terminal and C- terminal coding sequences common to other annotated mucin like genes suggests the existence of a new mucin like gene family.*

Key words: mucin like gene - *Trypanosoma cruzi* - *Trypanosoma rangeli*

Trypanosoma cruzi, the etiologic agent of Chagas disease, expresses mucin like glycoproteins rich in Thr, Ser, and Pro residues that cover the parasite cell surface (reviewed in Acosta-Serrano et al. 2001). These glycoconjugates are highly O-glycosylated molecules anchored to the plasma membrane via a glycosylphosphatidylinositol (GPI) moiety. They represent the major acceptors of sialic acid of the cell-surface trans-sialidase and are involved in parasite infectivity and protection from the immune response of the host. The mucin like genes encode proteins with conserved signal peptides and GPI anchor addition sites, but with highly variable central regions containing the sites for GlcNAc addition. The hundred of genes up to now characterized have been arranged into two separate families on the basis of sequence homology (reviewed in Frasch 2000): the TcMUC family, that contains two subfamilies (TcMUC I and TcMUC II) according to the presence or absence of tandem arrays of a variable number of repeat units, and the TcSMUG family, divided into two groups (Small and Large) according to the size of their encoded mRNAs (Di Noia et al. 2000). It has been recently demonstrated that the TcMUC II group of genes codes for the majority of mucins of the trypomastigote stage and that the C-terminal peptide of TcMUC II mucins elicits strong antibody responses in patients with Chagas disease (Buscaglia et al. 2004).

On the other hand, nothing has been reported on mucins in *T. rangeli*. This is an hemoflagelate species that is considered non-pathogenic for humans but represents a complication for the serological diagnosis of Chagas disease, because the cross reactivity with *T. cruzi* can produce false positives in the immunological tests used. Serological cross-reactivities can also confuse epidemiological studies of these infections (D'Alessandro & Saravia 1999, Grisard et al. 1999, Guhl & Vallejo 2003). It is still necessary to identify the distinctive characteristics that allow a differential diagnosis and/or epidemiological studies to avoid the unnecessary application of drug treatments that can cause severe side effects in patients with the misdiagnosed disease (Vasquez et al. 1997).

As there is evidence for the close evolutive relationship between *T. cruzi* and *T. rangeli* (Stevens et al. 2001), we decided to use the subtraction approach for the isolation and characterization of genes expressed by *T. cruzi* and not by *T. rangeli*. We report here the obtention and molecular characterization of a *T. cruzi* specific cDNA clone that showed a high homology percent with a *T. cruzi* gene automatically annotated by the *T. cruzi* genome project as a putative mucin associated surface protein (MASP). Given the reported antigenicity of C-terminal peptide of trypomastigote mucins coded by the TcMUCII group of genes (Buscaglia et al. 2004), this work might contribute to the potential development of a differential diagnosis method for Chagas disease in countries where co-infection with *T. rangeli* represents a problem.

The subtraction approach used was the representa-

Financial support: Fonacit grant S1-96001924, Consejo de Desarrollo Científico y Humanístico, grant A-09-34-3619-95, Fondo para la Promoción a la Investigación Científica y Tecnológica de la Facultad de Medicina de la Universidad Central de Venezuela
+Corresponding author. E-mail: abatet@ucv.ve
Received 10 December 2004
Accepted 21 June 2005

tion of differential expression (RDE) method (Krieger & Goldenberg 1998), which takes advantage of the existence of the mini-exon (a sequence common to all 5' ends of kinetoplastids mRNAs) and the polyA tail, in order to obtain amplicons representing the total polyadenylated RNAs. The method is based on the generation by PCR of sequences unique to a *T. cruzi* cDNA population (the tester amplicons) after removal of common sequences by subtractive hybridization with a *T. rangeli* cDNA population (the driver amplicons). We used *T. cruzi* EP trypomastigotes (a *T. cruzi* I isolate) from VERO infected cells and *T. rangeli* Palma 2 epimastigotes cultured in LIT medium. After two rounds of subtraction, we obtained two clearly defined cDNA bands, and the negative control (a subtractive hybridization using the same *T. cruzi* amplicons as tester and driver) gave no amplification products (Fig. 1A). The second round cDNAs were cloned in plasmid pBK-CMV (Stratagene) and several clones were randomly selected and analyzed by agarose gel electrophoresis and sequencing. The insert sizes ranged between 300-600 bp (results not shown).

The sequence analysis revealed that one of the clones (named TcML8, GenBank accession number AJ634679) is a 432 bp long cDNA, with the mini-exon and polyA sequences. This clone is transcribed in *T. cruzi* epimastigotes and trypomastigotes but not in *T. rangeli* epimastigotes (Fig. 1C). A Southern blot analysis of the genomic *T. cruzi* and *T. rangeli* DNAs digested with several restriction enzymes showed the TcML8 hybridization with multiple genomic bands, revealing the repetitive nature of TcML8 and its presence in both genomes (Fig. 1D), although it is not transcribed in *T. rangeli* as just mentioned. TcML8 showed a 92% identity with a *T. cruzi* gene (GeneDB Temporary systematic ID: Tc00.1047053503875.30) coding for a hypothetical protein belonging to the trypanosomal mucin-like glycoproteins (Fig. 2). We will refer here to this coding sequence as TcMUC8. As expected, TcML8 was also 92% identical to the *T. cruzi* contig region that contains TcMUC8 (GeneDB Tcruzi.chrunknown.4796). In the contig, there exists the AG 3' acceptor site for *trans* splicing immediately upstream of the 5' end of the TcML8 insert sequence, as expected because of the use of the mini-exon primer for the TcML8 RT-PCR cloning, but also there is an A-rich sequence immediately downstream of the 3' end (Fig. 2B). The high identity percent between TcML8 and TcMUC8 and the presence of this A-rich stretch in the coding region of TcMUC8 suggests that TcML8 is not a full length clone, but a partial TcMUC8 cDNA. Since the cDNA amplicons for subtraction hybridization were obtained by RT-PCR using the mini-exon and oligodT as primers, and because of the low temperature necessary to perform the reverse transcription reaction, the polyA sequence in the cDNA probably corresponds to this internal A-rich region within the TcMUC8 mRNA. This suggestion is supported by the existence of an EST (expressed sequence tag) coming from a *T. cruzi* epimastigote normalised cDNA library (GeneBank accession number AI035168) that is 99% homologous to the TcMUC8 sequence and includes the A-rich region (Fig. 2). The possibility that TcML8 could also be a pseudogene cannot be excluded. A family of oligoA-terminated, transcribed DNA elements in the *T. cruzi* genome has been reported (Verbisck et al. 2003). In any case, with a 92% homology, TcML8 is undoubtly related to TcMUC8, so we extended

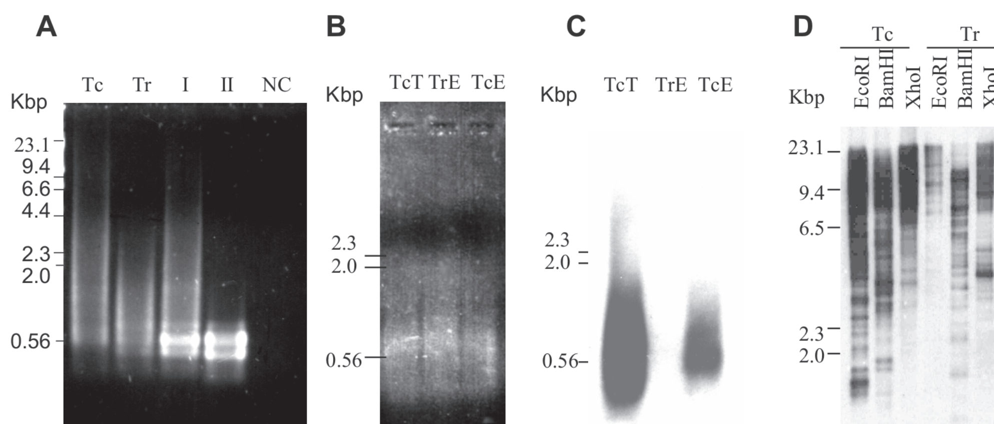

Fig. 1: agarose gel electrophoretic analysis of *Trypanosoma rangeli* subtracted *T. cruzi* cDNAs and TcML8 transcription and genomic distribution. Total RNA from both parasites was purified with TRIZOL (Gibco-BRL) and cDNA amplicons were obtained by RT-PCR using oligodT-*Bam*HI and mini-exon-*Bam*HI primers (Krieger & Goldenberg 1998). A: cDNA amplicons (0.3 µg) from the tester (Tc), the driver (Tr), the cDNAs obtained after the first (I) and second (II) rounds of subtraction and the negative control (NC) were fractionated by electrophoresis in a 1% agarose gel. For subtraction, tester amplicons were digested with *Bam*H1 and ligated to a *Bam*H1 adapter. Tester and driver amplicons were hybridized with a tester-driver ratio of 1:10. After filling in the cohesive ends, tester specific sequences were PCR amplified with an adaptor-specific primer. The second round of subtraction was carried out with another set of adapters and a tester-driver ratio of 1:100. We followed the detailed guidelines described by Pastorian et al. (2000) for adaptor ligation, tester-driver ratios, subtractive hybridization and PCR amplifications; B: cDNA amplicons (0.5 µg) from *T. cruzi* trypomastigotes (TcT), *T. rangeli* epimastigotes (TrE), and *T. cruzi* epimastigotes (TcE) were subjected to electrophoresis in a 1% agarose gel; C: the cDNA amplicons showed in B were blotted onto a nylon membrane and hybridized with a digoxigenin labeled TcML8 probe according to the instructions of the DIG High Prime DNA Labelling and Detection Starter Kit II (Roche); D: *T. cruzi* (Tc) and *T. rangeli* (Tr) genomic DNAs were purified with DNAzol (Gibco-BRL), digested with EcoRI, BamHI, and XhoI, electrophoresed in 1% agarose gel, transferred to a nylon membrane and hybridized with a digoxigenin labeled TcML8 probe according to the instructions of the DIG High Prime DNA Labelling and detection Starter Kit II (Roche). The TcML8 sequence has no internal sites for the enzymes used. Kbp: kilobase pairs.

our sequence analysis to the TcMUC8 gene.

As mentioned above, the existence of an homologous EST from the *T. cruzi* epimastigote stage demonstrates that TcMUC8 is transcribed in this parasite. On the other hand, the Southern blot analysis of RT-PCR amplicons using TcML8 as probe confirmed that TcMUC8 is not transcribed in *T. rangeli* epimastigotes (Fig. 1C). A computational analysis of TcMUC8 deduced protein using the SignalP 3.0 server (http://www.cbs.dtu.dk/ services/ SignalP), NetOGlyc 3.1 Server (http://www.cbs.dtu.dk/services/NetOGlyc/) and DPGI program (http://us. expasy.org/) showed that it has the characteristic features of mucins: a signal peptide sequence, O-glycosylation sites and GPI anchor sequence (Fig. 3A). Although TcMUC8 has no threonine repeats, it is a serine rich protein, like the group II of TcMUC genes characterized by Di Noia et al. (1998). In its aminoacidic composition, Glu, Ser, Pro, and Thr (Thr:36, Ser:58, Glu:49, Pro:36) predominate, with a Thr/Ser ratio of around 1:2. In spite of this, TcMUC8 showed no significant similarity at the nucleotidic level with the previously reported mucin like genes (Di Noia et al. 1995, 1998, 2000) not even at the conserved N and C terminal coding regions that the reported *T. cruzi* mucins have in common.

A bioinformatic analysis suggests that TcMUC8 might belong to a new mucin like gene family. A search for homologies with the BLASTn program from NCBI showed TcMUC8 partial homology with several apparently unrelated sequences: pyrimidine synthesis gene cluster region of *T. cruzi* (GeneBank accession number AB017765), a syntaxin 7-like protein gene (reverse complement, GenBank accession number AY344240), three regions (reverse complement) of *T. cruzi* unknown genes (GeneBank accession number AF525766), four regions of *T. cruzi* cosmid C71 mucin-like protein genes and C71 surface protein (sp1) genes (GeneBank accession number AY298908) and two regions of *T. cruzi* cosmid C2 mucin-like protein and gp85-like protein (gp-85) genes and unknown genes

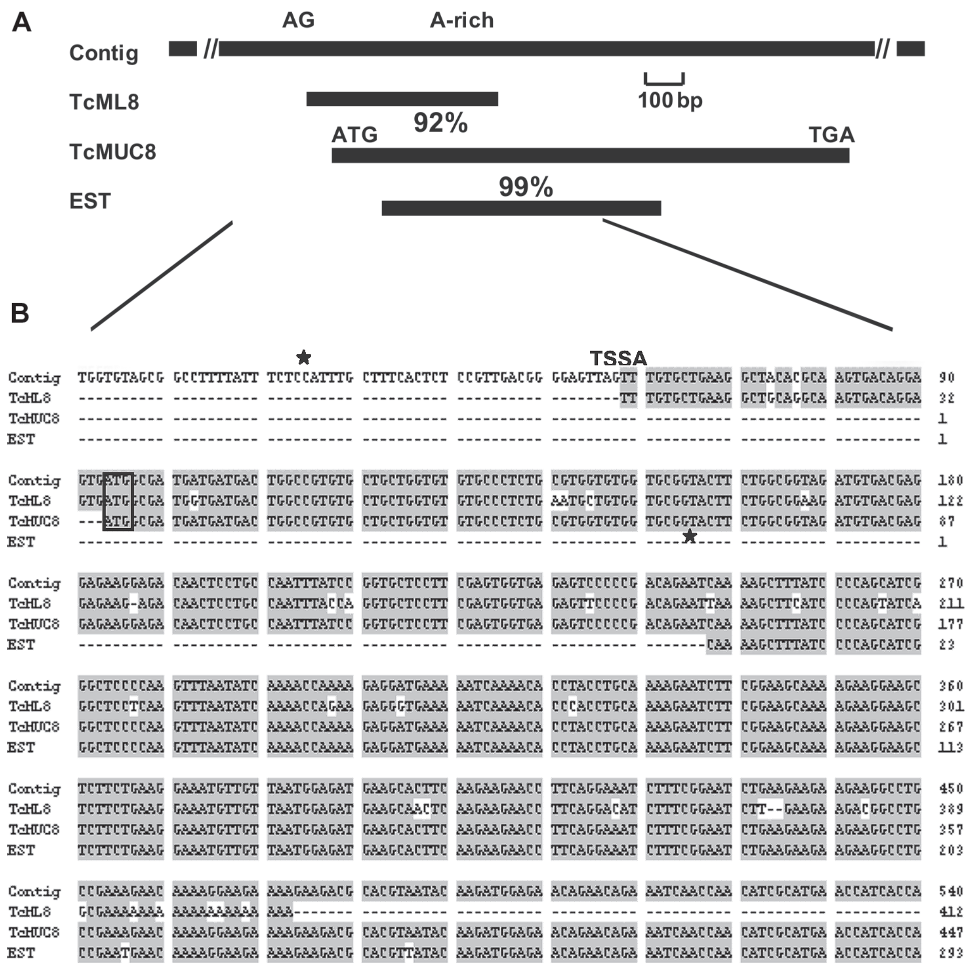

Fig. 2: schematic representation and comparison of Contig 4796, TcML8, TcMUC8, and *Trypanosoma cruzi* epimastigote expressed sequence tag (EST) sequences. A: schematic map of these sequences, showing the percentage of identity, the AG dinucleotide acceptor site for *trans* splicing, the A-rich sequence, and the TcMUC8 ORF (open reading frame); B: multiple aligment of the same sequences. Aligments were done by Clustal W program with BioEdit Sequence Aligment Editor, Isis Pharmaceuticals Inc. The ATG initiation codon is boxed, the *trans* splicing acceptor site is shown (TSAS) and the A-rich sequence is underlined. ★ indicate the start and end of the TcRE8 repetitive element. Gaps (-) were introduced to maximize the alignments, identical residues are shaded in gray and the nucleotide position is indicated in the right margin.

(GeneBank accession number AY298909). Interestingly, in all cases the homology was restricted to the first 63 pb of TcMUC8 and to the last 88 bp at its 3' end. On the other hand, two searches for TcMUC8 homologies with the BLASTn program against the NCBI unfinished *T. cruzi* database using only the first 63 bp or the last 88 bp of TcMUC8, showed a 90-97% homology percent with all the output sequences (aligment: 500). For some genomic clones the homologies were observed at several regions within the same clone. These sequences homologous only to the ends of TcMUC8 were putative mucin-like genes/ pseudogenes as reavealed by a BLASTn search against the *T. cruzi* GeneDB predicted gene database. The multiple aligment of some of these mucin like coding genes showed that they share the N-terminal coding region and the C-terminal-GPI anchor coding region (Fig. 3B, C). These observations allow us to suggest the existence of a new putative mucin like gene family. It is known that the TcMUC gene family has a diversity though this is not yet well understood (Di Noia et al. 1995, Freitas-Junior et al. 1998).

A more detailed sequence analysis made by comparing the TcMUC8 homologous genomic sequences among themselves allowed the identification of a 130 bp long repetitive sequence that included the first 98 bp of TcML8 at its 3' end (Fig. 2B). We named the repeat element TcRE8. In a BLASTn search against NCBI unfinished *T. cruzi* database there are by far much more sequences 100% homologous to the first 98 bp of TcML8 than to the complete TcML8 sequence. This fact probably explains the repetitive pattern observed in the Southern blot experiment (Fig. 1D). It also might explain the wide hybridiza-

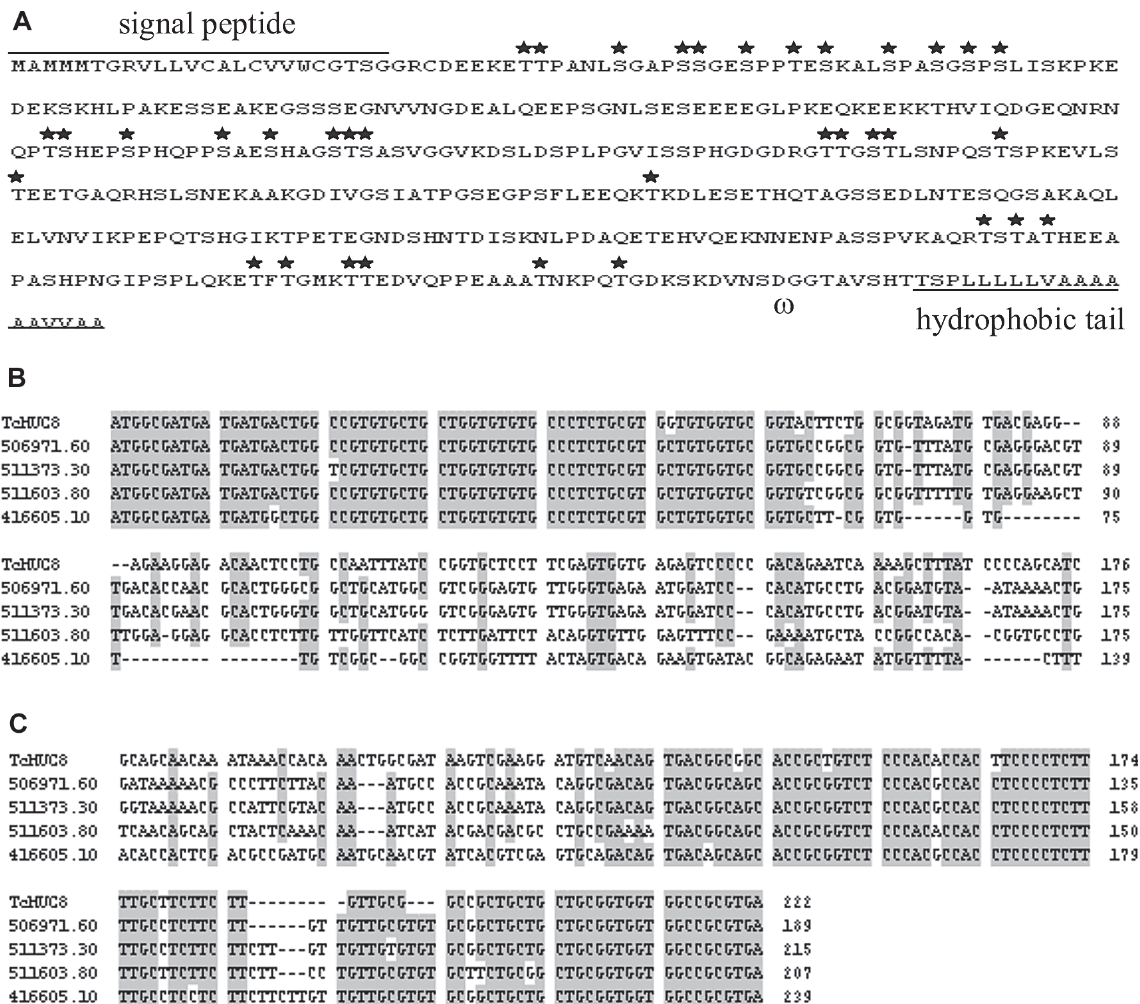

Fig. 3: structure of the TcMUC8 deduced protein and comparison of nucleotide sequences from TcMUC8 and other annotated *Trypanosoma cruzi* putative mucin like genes. A: aminoacidic sequence of TcMUC8 deduced protein showing the putative signal peptide (identified by using the SignalP 3.0 server at http://www.cbs.dtu.dk/ services/SignalP) and the w cleavage site and hydrophobic tail of the putative GPI anchor sequence (DPGI program at http://us.expasy.org/). ★ indicate the potential O-glycosylation sites according to the NetOGlyc 3.1 program (http://www.cbs.dtu.dk/services/NetOGlyc/); B: clustal W alignment of the N-terminal coding sequences of TcMUC8 and some *T. cruzi* mucin like coding sequences from the GeneDB predicted gene database; C: clustal W alignment of the C-terminal coding sequences of TcMUC8 and some *T. cruzi* mucin like coding sequences from the GeneDB predicted gene database. All sequence names correspond to the GeneDB temporary systematic ID without the preceding Tc00.1047053 common to all them, except for TcMUC8. Gaps are indicated by a dash (-). Identical residues are shaded in gray and the nucleotide position is indicated in the right margin.

tion observed in the Southern blot of RT-PCR amplicons (Fig. 1C), since this would be the result expected if the putative mucin like genes annotated in the *T. cruzi* Genome Project (genes that contain this repetitive sequence) were transcribed. This group of related genes has members sized from 0.555 Kp (GeneDB ID Tc00.104705350 6615.100) to 1.03 Kb (GeneDB ID Tc00.1047053503973.120). However, we do not know if this repetitive element is also present in other genomic contexts.

It was not possible to find sequences as direct or reverse repeats flanking TcRE8. The copy number estimation for TcRE8 in the *T. cruzi* genome was 1482, as determined by the equation proposed by Agüero et al. (2000), using 40 Mb as the haploid *T. cruzi* genome, a total number of GSS (Genome Survey Sequences) of 12036 and a TcRE8 size of 130bp. The AG dinucleotide at position 31 of TcRE8 and immediately upstream of the TcML8 5' end (as can be seen in the aligment with the contig 4796, Fig. 2B) is used as its 3' acceptor site for trans splicing, since TcML8 was cloned using the miniexon and polyA sequences in a RT-PCR reaction. It remains to be proven if TcRE8 is a repetitive element present in other genomic contexts or if it should be considered only as a part of some of the genes/pseudogenes constituting the very large and divergent mucin-like multigene family.

In conclusion, we have demonstrated by a combination of experimental and bioinformatic analysis, that the TcMUC8 gene automatically annotated by the *T. cruzi* genome project is transcribed in *T. cruzi* but not in *T. rangeli*, possesses the characteristic features of a mucin type gene but showed no nucleotidic similarities with previously described members of the *T.* cruzi mucin like gene families, and has N-terminal and C-terminal coding sequences homologous to other annotated mucin type genes. All this allows us to suggest the existence of a new mucin gene family. Since TcMUC8 gene is not transcribed in *T. rangeli* we are currently studying its expression and potential use in the epidemiology and specific diagnosis of Chagas disease in Venezuelan patients.

ACKNOWLEDGEMENTS

To Dr Clara Martínez for important comments.

REFERENCES

Acosta-Serrano A, Almeida IC, Freitas-Junior LH, Yoshida N, Schenkman S 2001. The mucin-like glycoprotein super-family of *Trypanosoma cruzi*: structure and biological roles. *Mol Biochem Parasitol 114*: 143-150.

Agüero F, Verdún RE, Frasch AC, 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.

Buscaglia CA, Campo VA, Di Noia JM, Torrecilhas AC, De Marchi CR, Ferguson MA, Frasch AC, Almeida IC 2004. The surface coat of the mammal-dwelling infective trypomastigote stage of *Trypanosoma cruzi* is formed by highly diverse immunogenic mucins. *J Biol Chem 279*: 15860-15869.

D'Alessandro A, Saravia NG 1999. *Trypanosoma rangeli*. In HM Gilles, *Protozoal Diseases*, Arnold Press, London, p. 398-412.

Di Noia JM, D'Orso I, Aslund L, Sanchez DO, Frasch AC 1998. The *Trypanosoma cruzi* mucin family is transcribed from hundreds of genes having hypervariable regions. *J Biol Chem 273*: 10843-10850.

Di Noia JM, D'Orso I, Sanchez DO, Frasch AC 2000. AU-rich elements in the 3'-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. *J Biol Chem 275*: 10218-10227.

Di Noia JM, Sanchez DO, Frasch AC 1995. The protozoan *Trypanosoma cruzi* has a family of genes resembling the mucin genes of mammalian cells. *J Biol Chem 270*: 24146-24149.

Frasch AC 2000. Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. *Parasitol Today 16*: 282-286.

Freitas-Junior LH, Briones MR, Schenkman S 1998. Two distinct groups of mucin-like genes are differentially expressed in the developmental stages of *Trypanosoma cruzi*. *Mol Biochem Parasitol 93*: 101-114.

Grisard EC, Steindel M, Guarnieri AA, Eger-Mangrich I, Campbell DA, Romanha AJ 1999. Characterization of *Trypanosoma rangeli* strains isolated in Central and South America: overview. *Mem Inst Oswaldo Cruz 94*: 203-209.

Guhl F, Vallejo GA 2003. *Trypanosoma (Herpetosoma) rangeli* Tejera, 1920: an updated review. *Mem Inst Oswaldo Cruz 98*: 435-442.

Krieger MA, Goldenberg S 1998. Representation of differential expression (RDE): a new approach to study differential gene expression in trypanosomatids. *Parasitol Today 14*: 163-166.

Pastorian K, Hawel L 3rd, Byus CV 2000. Optimization of cDNA representational difference analysis for the identification of differentially expressed mRNAs. *Anal Biochem 283*: 89-98.

Stevens JR, Noyes HA, Schofield CJ, Gibson W 2001. The molecular evolution of Trypanosomatidae. *Adv Parasitol 48*: 1-56.

Vasquez JE, Krusnell J, Orn A, Sousa OE, Harris RA 1997. Serological diagnosis of *Trypanosoma rangeli* infected patients. A comparison of different methods and its implications for the diagnosis of Chagas' disease. *Scand J Immunol 45*: 322-330.

Verbisck NV, dos Santos MR, Engman DM, Angel Chiurillo M, Ramirez JL, Araya JE, Mortara RA, da Silveira JF 2003. A novel reiterated family of transcribed oligo(A)-terminated, interspersed DNA elements in the genome of *Trypanosoma cruzi*. *Mem Inst Oswaldo Cruz 98*: 129-133.