Trypanosoma cruzi ribosomal protein S4: characterization of its coding locus, analysis of transcripts, and antigenicity of the protein


Departmento de Biología Molecular y Biotecnología *Departamento de Inmunología, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Apartado postal 70-228 CP04510, México, D. F., México

Two allelic genomic fragments containing ribosomal protein S4 encoding genes (rpS4) from Trypanosoma cruzi (CL-Brener strain) were isolated and characterized. One allele comprises two complete tandem repeats of a sequence encoding an rpS4 gene. In the other, only one rpS4 gene is found. Sequence comparison to the accessed data in the genome project database reveals that our two-copy allele corresponds to a variant haplotype. However, the deduced aminoacid sequence of all the gene copies is identical. The rpS4 transcripts processing sites were determined by comparison of genomic sequences with published cDNA data. The obtained sequence data demonstrates that rpS4 genes are expressed in epimastigotes, amastigotes, and trypomastigotes. A recombinant version of rpS4 was found to be an antigenic: it was recognized by 62.5% of the individuals with positive serology for T. cruzi and by 93.3% of patients with proven chronic chagasic disease.

Key words: kinetoplastid - trypanosomatids - Trypanosoma cruzi - ribosome - gene structure

---

Trypanosoma cruzi is a protozoan parasite that causes Chagas disease, an endemic infection in South and Central America that affects approximately 11 million people (Guzman-Bracho 2001). T. cruzi species has been classified into two major well defined genomic groups: T. cruzi I and T. cruzi II (Souto et al. 1996). T. cruzi I has been associated with the sylvatic transmission cycle and infection of marsupials whereas T. cruzi II has been associated with the domestic transmission cycle and infection of placental mammals. The natural history of the disease involves three different phases. The initial acute infection (symptomatic in only 1% of the cases) is characterized by febrile symptoms and inflammation at the site of entry of the parasite. This stage is followed by an indeterminate (subacute) stage in which symptoms are not apparent. Finally, after a period of years to decades, the chronic stage of the infection may manifest itself in up to 30% of individuals. A variety of clinical presentations may be seen including digestive tract anomalies (megasphagus and megacolon) and cardiac enlargement and malfunction (Teixeira 1987). The variability of symptoms has yet to be correlated with specific parasite or host genetic markers, although it is likely that both will affect the outcome of infection.

The ribosomal protein S4 (rpS4) is a basic type protein located at the interface between ribosomal subunits (Uchiumi et al. 1986). In agreement with this is the observation that the rpS4 can be cross-linked to the eukaryotic initiation factor eIF-3 (Westermann et al. 1983). Our early interest in this protein lies with report in which the yeast ribosomal protein S7, homologous to the mammalian rpS4 was found to be an essential protein (Synetos et al. 1992). From the sequence analysis of two epimastigote cDNA clones it was deduced that T. cruzi rpS4 is a 273 amino acid, conserved basic protein (Hernández et al. 1998). The gene is expressed as a 1 kb transcript and the initial hybridization studies were consistent with the occurrence of two alleles as predicted in a diploid organism. Moreover, Southern blot analysis of pulse field chromosomal gels demonstrated that the genes were present in two homologous chromosomes of different sizes (Hernández et al. 1998).

T. cruzi ribosomes are known to elicit both humoral and cellular immune responses (Teixeira & Santos-Buch 1974, 1975). Furthermore, T. cruzi ribosomes and ribosomal proteins have been associated with the induction of heart dysfunction in experimental models similar to that seen in patients with chronic disease (Teixeira et al. 1975). The acidic ribosomal P proteins have been found to be antigenic, and the humoral immune responses to these proteins have been implicated in the pathogenesis of chagasic chronic heart disease (Elies et al. 1996, Kaplan et al. 1997, Lopez Bergami et al. 2001). However, the role of other ribosomal antigens has not been evaluated.

In addition to a detailed characterization of the rpS4 encoding locus this work evaluates the antigenicity of this basic type ribosomal protein.

MATERIALS AND METHODS

Parasites - T. cruzi II (CL-Brener strain) epimastigotes kindly donated in 1995 by Dr Bianca Zingales (Universidade de São Paulo) have been maintained at 27°C in liver-infusion tryptose medium supplemented with 10% new-born calf serum.
Cloning of the rpS4 genes of T. cruzi - Genomic DNA from our T. cruzi II (CL-Brener strain) was used to construct an EcoRI I-digested λZAP II (Stratagene) library (Cevallos et al. 2003). The library was screened using the EcoRI I - Kpn I rpS4 encoding fragment from the T. cruzi I (Tulahuen strain) cDNA clone pS4-2 (Hernandez et al. 1998). Five positive clones were identified which contained two different inserts 8.5 and 7.5 Kb in size. Two clones p811 (with an 8.5 Kb insert) and p812 (with a 7.5 Kb insert) were selected for further study. The regions containing the homologous sequences were subcloned and both strands sequenced by the dye-terminator method on an ABI-PRISM 310 automated sequencer (Applied Biosystems).

Expression and purification of a recombinant rpS4 and recombinant GST - The entire rpS4 open reading frame (ORF) was amplified from the genomic clone p812 by PCR with primers containing BamH I restriction sites (S4-GST-F 5'-GGGA TCCCCA TGACCAAGAC-ACTTGCG-3' and S4-GST-R 5'-GGGA TCCTA TTTTCGTGC-CTTGGG-3'). The amplification product was subcloned in phase into the BamH I site of the bacterial expression vector pGEX-3X (Promega, Madison, Wisconsin), sequenced and expressed as a GST (glutathione S-transferase)-rpS4 fusion protein in Escherichia coli (BL21 strain). E. coli transformed with the recombinant plasmid was cultured at 37°C until the optical density reached 0.5 at 600 nm, and the synthesis of recombinant rpS4 was induced with 1 mM isopropylthio-β-galactoside at 37°C for 4 h. The resulting culture was harvested by centrifugation at 12,000 × g for 10 min at 4°C and incubated in lysozyme containing buffer [100 µg/ml in 20 mM Tris-HCl, pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA)] for 30 min at 4°C prior to sonication. A soluble fraction was obtained by centrifugation at 12,000 × g for 10 min at 4°C. Phosphate buffered saline pH 7.2 (PBS, 10 ×) and Triton X-100 were added to the soluble fraction to a final concentration of 1% Triton X and 1 × PBS. The lysate was loaded into a glutathione Sepharose 4B column, and the recombinant protein eluted according to the manufacturer’s instructions (Amersham Biosciences, London, UK). The size and purity of the recombinant protein was verified with the use of SDS-PAGE low molecular weight markers (BioRad Laboratories). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were performed at room temperature in gels (10 or 12%) containing 0.1% SDS, as previously described (Laemmli 1970). Samples were dissolved in sample buffer [1:1 (v/v); 62.5 mM Tris-HCl pH 6.8, 25% glycerol, 2% SDS, 0.01% bromophenol blue] and the electrophoresis conducted in running buffer (25 mM Tris pH 8.8, 192 mM glycine, 0.1% SDS).

Since several contaminant proteins remained in the eluted fraction, a second purification was required. The recombinant protein preparations were resolved by SDS-PAGE (10%) and the band of 57 kDa was cut from the gel and electroeluted from it. After dialysis of the eluted protein in 10 mM Tris-HCl pH 8; 100 mM NaCl, 10% glycerol at 4°C, the preparations were concentrated by ultra filtration in Centricon® 30 filter units (Millipore Corporation, Bedford, US) and stored at ~20°C for further analysis. The GST protein used as a control was expressed directly from the pGEX-3X vector and purified as described above.

**RESULTS**

Cloning and sequence analysis of the rpS4 locus - Screening of a T. cruzi II CL-Brener genomic library with a rpS4 probe identified two positive clones with inserts that corresponded in size to the two expected published
allelic fragments (Hernández et al. 1998). Restriction analysis and Southern hybridizations of the digested clones delimited, in both clones, Not I - EcoRI fragments recognized by the rps4 cDNA probe. These two regions differed in fragment size: 3.4 kb in one clone and 2.4 kb in the other. Both fragments were sequenced and the nucleotide data was deposited in the GenBank (accession nos. DQ288964 and DQ288965). Sequence analyses showed that the difference in size between the two fragments was due to the presence of two rps4 gene copies in the 3.4 kb fragment (S4-1 and S4-2) and one copy (S4-3) in the 2.4 kb fragment (see Fig. 1, alleles marked with an asterisk). At the nucleotide level the DNA sequence of both alleles was identical with the exception of a 1072 bp DNA insertion in the two-copy allele containing a 250 bp intergenic region and the extra rps4 gene copy (see the “Tulahuen-like” allele in Fig. 1). Sequence comparison with our previously reported rps4 cDNA clones from the T. cruzi I Tulahuen strain (GenBank accession nos. AF005421 and AF005904), demonstrated that the three coding regions of our CL-Brener strain were identical to both of them. In particular, the rps4-2 gene copy present in our 3.4 kb cloned genomic fragment is not only identical in its coding region but also in its non-coding untranslated regions to the Tulahuen derived rps4 cDNA clone. Therefore we have named this allele as “Tulahuen-like”.

The T. cruzi genome project was carried out in the CL-Brener strain which proved to be a genetic hybrid. The alleles in many cases have been named “Esmeraldo-like” and “non-Esmeraldo-like” haplotypes according to their suspected progenitors. When our rps4 coding region was used to search the T. cruzi Genome Database (http://www.genedb.org/genedb/cruzi/blast.jsp), two contigs containing rps4 genes were identified. Our one copy allele corresponds to the “non-Esmeraldo-like” haplotype. On the other hand, our two copy allele did not correspond in sequence to the accessed “Esmeraldo-

Fig. 1: map of the ribosomal protein S4 genomic locus. At the bottom are depicted the two alleles present in the CL-Brener strain accessed in the Genome Project Database. At the top is represented the variant allele “Tulahuen-like” found in our CL-Brener strain that replaces the “Esmeraldo-like” allele reported in the Genome Project Database. Our second allele corresponds to a “non-Esmeraldo-like”. The asterisks mark the two alleles present in our CL-Brener strain and the size of the sequenced fragments are indicated. Arrows represent open reading frames and the shaded box represents a SIRE element. The coding regions for the three copies of the rps4 gene are labeled as S4-1, S4-2 and S4-3. The rest of the ORFs are labeled with a letter and encode for a Thiamine pyrophosphokinase (a), a nucleobase transporter (b), and for a hypothetical protein (c). There are two insertions in the “Esmeraldo-like” allele not present in the “non-Esmeraldo-like” allele: a SIRE element and an extra copy of the rps4 gene with the corresponding intergenic region (marked with dashed lines).

Differences in arrow patterns represent differences in the nucleotide sequences of the respective ORF. The rps4 genes present in the “Tulahuen-like” and in the “non-Esmeraldo-like” alleles (solid arrows) are identical among them but have eight conserved nucleotide differences when compared to the rps4 copies 1 and 2 present in the “Esmeraldo-like” allele (arrows with diagonal stripes). In a similar manner, nucleotide sequences of the nucleobase transporter present in the “Tulahuen-like” and in the “non-Esmeraldo-like” alleles (arrows with vertical stripes) are identical among them but present 18 nucleotide differences, that result in five amino acid changes when compared to the “Esmeraldo-like” allele sequence (doted arrow). Untranslated regions of the rps4 genes are labeled as α and β to indicate the two types of 5’ UTR sequences of the rps4 genes and as x and y to indicate the two types of 3’ UTR sequences.
like” haplotype even though this allele also contains two copies of the rpS4 gene. Differences were found both in the coding and non-coding regions as mentioned in legend to Fig. 1.

A further analysis of the accessed genomic context of the rpS4 genes detected the presence of a short interspersed repetitive element (SIRE) as defined by Vazquez et al. (1994) in the “Esmeraldo-like” allele (Fig. 1). In addition, three ORFs with the same orientation are found in both accessed alleles (Fig. 1 a, b, c). Gene b (a nucleobase transporter) is different in the “Esmeraldo like” haplotype as detailed in Fig. 1.

Characterization of rpS4 mRNA processing sites - Productive expression of protein coding genes in trypanosomes involve the addition of a trans-spliced capped mini exon (39 nt) next to the 5′ UTR, and polyadenylation of the 3′ terminus within individual mRNAs. These two reactions occur within the intergenic regions of a primary polycistronic mRNAs transcript and are coupled and directed by a common polypyrimidine tract, which is part of the spliced leader acceptor site. Disruption of this bifunctional signal sequence affects the expression of 5′ and 3′ adjacent flanking genes (Lopez-Estrano et al. 1998, Hummel et al. 2000). Sequence analysis of non-coding regions of the rpS4 genes demonstrated that the 5′ upstream region of copies 1 and 3 was identical but significantly different from the 5′ upstream sequence of copy 2 (Fig. 1; indicated as α and β). The sequences downstream of the translation stop codon for copies 2 and 3 was similar, and in turn dissimilar to the sequence present downstream of rpS4 gene copy 1 (Fig. 1; indicated as x and y). These differences result in the presence of three types of transcripts that can be readily identified when analyzing cDNA sequence data and allowed us to characterize our previously reported Tulahuen strain derived cDNA clones as copy 2 transcripts.

To determine the sites at which transcript processing occurs for the other two gene copies, a search was performed to identify cDNA clones accessed in the NCBI and EMBL databases. A total of 32 rpS4 cDNA clones were identified, 21 derived from T. cruzi epimastigotes, 6 from amastigotes and 5 from trypomastigotes. The comparison of these sequences with the genomic DNA sequence allowed us to determine the sites where transsplicing and polyadenylation occur in each copy (see Fig. 2).

Antigenicity of ribosomal protein S4 - With the aid of a recombinant form of T. cruzi rpS4 we inferred the antibody response to the natural form of this protein in three well characterized groups: patients with Chagas chronic cardiomyopathy, asymptomatic individuals with two positive serological tests and blood bank donors (Fig. 3). Antibodies against rpS4 were common in patients with known exposure to the parasite: 20 out of 32 asymptomatic individuals (62.5%) and 14 out of 15 patients with established cardiomyopathy (93.3%). There were no false positives as none of the 53 sera negative for T. cruzi (healthy blood bank donors) was reactive to rpS4. To analyze the potential recognition of the GST fragment of the recombinant protein, all rpS4 positive sera were also tested for antibodies against the GST peptide. None of these sera were GST positive (data not shown).

Fig. 2: ribosomal protein S4 genes, flanking sequences, and mapping of mRNA processing sites. Genomic sequences for the three rpS4 gene copies are aligned (C1- C2, and C3) with their inferred progenitor origin denoted with a T, for a “Tulahuen-like” or nE for “non-Esmeraldo-like”. The rpS4 encoding regions are not depicted in full. Their position is shown in between the start and stop translation codons (bold). Nucleotides identical to the T-C1 sequence are denoted by dots. Relevant motifs from the spliced leader acceptor site (polypyrimidine tract and dinucleotide AG) are boxed. Polyadenylation sites are highlighted with a shaded box, as the last nucleotide identified different from A. Processing sites for T-C2 were derived from the cDNA sequences AF005421 and AF005904. Transsplicing site for copies T-C1 and nE-C3 were derived from sequences AA426688 and CB923834. Polyadenylation site of copy T-C1 was inferred from AA866549.
do not seem to be frequent enough as to disrupt the prevalent clonal pattern of the population (Telleria et al. 2004). T. cruzi CL-Brener, the strain used in this study, is known to be heterozygous at many loci, with different-sized homologous chromosome pairs. It is believed that its genome is a hybrid from subgroup IIb and subgroup IIc (which itself is also apparently a hybrid derived from T. cruzi I) (El Sayed et al. 2005). The presence of variant alleles within the same T. cruzi subgroup and the presence of two distinct sequence classes representative of different subgroups within particular strains has already been reported (Westenberger et al. 2005). In this work we have cloned and sequenced two alleles containing the \( rps4 \) genes in a laboratory maintained CL-Brener strain. Sequence analyses demonstrated that one of these fragments corresponds to a different allele from those described in the genome project, and that this variant is probably related to a T. cruzi I-Tulahuen strain. The finding of T. cruzi I sequences in the CL-Brener strain further supports the proposal of multiple progenitors in the evolution of this T. cruzi hybrid strain (El Sayed et al. 2005).

We have also identified the processing sites of \( rps4 \) primary transcripts by comparing the genomic sequences with published cDNA data. Analyses of the untranslated regions demonstrate that each transcript is different, but can be grouped into two types according to the 3’ UTR (copy 1 and copies 2/3). It has been demonstrated that differences in the 3’ UTR of \( \beta \)-tubulin mRNAs result in differences in mRNA stability of the specific transcripts in trypomastigotes and amastigotes (Bartholomeu et al. 2002). A working hypothesis would propose that the distinct 3’ UTR in the \( rps4 \) transcripts may participate differentially in the physiology of expression of this protein.

Finally, it is described here that \( rps4 \) is recognized by two thirds of individuals with positive serologic response to T. cruzi antigens and more than 90% of patients with proven chagasic disease. There were no false positives as none of the control sera recognized the antigen. The prevalence of reactivity to various T. cruzi proteins in patients with chronic symptomatic chagasic disease by Western blot (Sánchez et al. 2001) has been reported. In this study it was found that the prevalence of reactivity to specific protein bands varied from 23 to 100%. However, only two proteins of 32 and 42 kDa were recognized by 90% of the patients or more.

**DISCUSSION**

The taxon T. cruzi contains two well defined genomic groups T. cruzi I and T. cruzi II. T. cruzi I has been associated with the sylvatic transmission cycle and infection of marsupials. T. cruzi II consists of five related subgroups, termed IIa, IIb, IIc, IId, and Ile, and has been associated with the domestic transmission cycle and infection of placental mammals. The T. cruzi strain CL-Brener is a member of subgroup Ile and was selected for the genome sequence project because it was well characterized (El Sayed et al. 2005). Although T. cruzi lineages have mostly evolved in a clonal fashion, several studies have reported the existence of recombinant genotypes in this species, either in natural populations (Boglìolo et al. 1996, Carrasco et al. 1996, Machado & Ayala 2001) or in the laboratory (Gaunt et al. 2003). In vitro, studies have demonstrated that T. cruzi has an extant capacity for genetic exchange. In one report, meticulous quantification analyses demonstrated that biological clones of a single T. cruzi strain had between 30 and 70% more DNA than the parental stock indicating some rapid genetic mechanism for radical change in DNA content (McDaniel & Dvorak 1993). In another report two biological clones were passaged together through the entire life cycle and then recovered from the mammalian stage of the life cycle. Comparison of the progeny demonstrated fusion of parental genotypes, loss of alleles, homologous recombination, and uniparental inheritance of kinetoplast maxicircle DNA (Gaunt et al. 2003). Notwithstanding these findings, the recombination events

![Western blots against \( rps4 \). Gel purified \( rps4 \) (2.5 µg) was loaded in analytical gel slots. Sera from asymptomatic individuals, chagasic patients, and healthy controls were tested. A positive control a strip incubated with an anti-GST antibody that recognized the fusion portion of the recombinant protein was included in all blots (marked with an asterisk). Depicted molecular markers correspond to the Benchmark prestained ladder for protein (Invitrogen).](image-url)
mal protein L27 has an epitope that cross-reacts with the Sm-epitope present in small nuclear ribonucleoproteins (Perone et al. 2003). Approximately two thirds of patients with Chagas disease have antibodies against the Sm epitope present in both human and trypanosomal small nuclear ribonucleoproteins (Bach-Elias et al. 1998). The role of these antibodies in the pathogenesis of the disease remains to be determined.

Analysis of expressed sequence tags from *T. cruzi* amastigotes, the reproductive stage in humans, has shown that 9% clones of the cDNA library encoded for ribosomal proteins with more than 30 classes of ribosomal proteins being identified (Cerqueira et al. 2005). Interestingly, an immunosurvey of the same library with sera from patients with chagasic disease revealed that ribosomal proteins also represent the largest class of antigen coding genes expressed in amastigotes (DaRocha et al. 2002). However only a subset of the ribosomal proteins expressed appeared antigenic. Besides the previously reported ribosomal P proteins, the ribosomal proteins L19 and L7a were also found to be antigenic. An immunosurvey of the same library with sera of *T. cruzi* patients with Chagas disease has antibodies against the Sm-epitope present in small nuclear ribonucleoproteins (Bach-Elias et al. 1998). The role of these antibodies in the pathogenesis of the disease remains to be determined.

In *Leishmania* the rpS4 antigen (with a 91% identity with *T. cruzi* rpS4) was identified using parasite-specific T cell lines derived from an immune donor and has been proposed as a vaccine candidate (Probst et al. 2001). Further studies should establish the value of rpS4 antigen as a diagnostic tool or as a vaccine candidate.

**ACKNOWLEDGMENTS**

To Dr Imelda Lópe Villaseñor for the critical reading of the manuscript and Juliana Herrera for technical assistance.

**REFERENCES**


