Title
Nucleotide sequence and transcription of a trypomastigote surface antigen gene of Trypanosoma cruzi

Journal
Molecular and Biochemical Parasitology, 46(2)

ISSN
0166-6851

Authors
Fouts, David L
Ruef, Barbara J
Ridley, Peter T
et al.

Publication Date
1991-06-01

DOI
10.1016/0166-6851(91)90043-6

Molecular and Biochemical Parasitology, 46 (1991) 189–200

ADONIS 016668519100155A

MOLBIO 01524

Nucleotide sequence and transcription of a trypomastigote surface antigen gene of *Trypanosoma cruzi*

David L. Fouts, Barbara J. Ruef, Peter T. Ridley[*1], Ruth A. Wrightsman[*2], David S. Peterson[*3] and Jerry E. Manning

Department of Molecular Biology and Biochemistry, University of California, Irvine, CA, U.S.A.

(Received 21 September 1990; accepted 27 November 1990)

In previous studies we identified a 500-bp segment of the gene, TSA-1, which encodes an 85-kDa trypomastigote-specific surface antigen of the Peru strain of *Trypanosoma cruzi*. TSA-1 was shown to be located at a telomeric site and to contain a 27-bp tandem repeat unit within the coding region. This repeat unit defines a discrete subset of a multigene family and places the TSA-1 gene within this subset. In this study, we present the complete nucleotide sequence of the TSA-1 gene from the Peru strain. By homology matrix analysis, fragments of two other trypomastigote specific surface antigen genes, pTt34 and SA85-1.1, are shown to have extensive sequence homology with TSA-1 indicating that these genes are members of the same gene family as TSA-1. The TSA-1 subfamily was also found to be active in two other strains of *T. cruzi*, one of which contains multiple telomeric members and one of which contains a single non-telomeric member, suggesting that transcription is not necessarily dependent on the gene being located at a telomeric site. Also, while some of the sequences found in this gene family are present in 2 size classes of poly(A)$^+$ RNA, others appear to be restricted to only 1 of the 2 RNA classes.

Key words: *Trypanosoma cruzi*; Trypomastigote surface antigen gene

Introduction

Trypanosoma cruzi is a flagellated parasitic protozoan and the causative agent of Chagas' disease, a serious health hazard throughout much of Central and South America [1]. The parasite has several morphologically different forms during its life cycle, which is divided between the insect vector and the mammalian host. In the insect vector the dividing form of the parasite is the non-invasive epimastigote. Upon migration of the epimastigote to the hindgut of the insect, it transforms to the non-dividing but invasive trypomastigote stage.

In the mammalian host the parasite circulates in the bloodstream as the infective trypomastigote. Upon penetration of the host cell it transforms to the intracellular amastigote, which is the dividing form of the parasite in the mammalian host. Since the trypomastigote is the major invasive form of the parasite in the mammalian host and is exposed to the host defensive mechanisms, much attention has been focused on the surface proteins of this stage of the parasite as potential protective immunogens against infection. Particular attention has been given to surface glycoproteins of 85 kDa, primarily because they constitute the major surface glycoproteins found on the surface of the invasive trypomastigote [2–4] and have been implicated in the

Correspondence address: Jerry E. Manning, Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92717, U.S.A.

*Present addresses: [1]Allergan, Irvine, CA, U.S.A.; [2]Department of Biological Sciences, Saddleback College, Mission Viejo, CA, U.S.A.; [3]National Institutes of Allergy and Infectious Disease, Bethesda, MD, U.S.A.

Note: Nucleotide sequence data reported in this paper have been submitted to the GenBank™ sequence data base with the accession number M58466.

Abbreviations: aa, amino acid; nt, nucleotide; TSA-1, trypomastigote specific surface antigen.

cell penetration process [5,6].

Although an 85-kDa trypomastigote surface glycoprotein has not yet been purified and characterized, genes which encode portions of trypomastigote specific 85-kDa surface antigens have been identified and partially characterized [7–10]. Examination of the genomic organization and RNA transcription products of these genes show them to share certain common features. Each of the genes have sequence homology with trypomastigote poly(A)$^+$ RNA of length 3.4–3.9 kb long, and each is a member of a large multigene family. Differences have been observed, however, regarding the number of members of the gene families which are being co-expressed. Our previous studies [7,8] identified a 4-member subset of a large 85-kDa gene family in which each member of the subset contains a 27-bp sequence that is tandemly repeated within the gene. Only 1 member of the subset is telomeric and it serves as the template for an abundant trypomastigote specific poly(A)$^+$ RNA while the other 3 members are either not transcribed or transcribed at very low levels. This observation suggests that not all members of the gene family are being co-expressed and that those members which are expressed might be located at telomeric sites. In contrast, another study has shown that several members of an 85-kDa gene family, both telomeric and non-telomeric, are being expressed simultaneously in both amastigotes and trypomastigotes [10]. Since the pattern of expression of these 2 gene families seems quite different, the question arises as to whether they represent two distinctly different 85-kDa gene families.

We report here the complete nucleotide sequence of the telomeric member of the subfamily defined by the presence of the 27-bp repeat unit. Comparison of the gene sequence with partial sequence data from other genes observed to encode an 85-kDa trypomastigote surface antigen(s) [9,10] indicates that these genes share extensive sequence homology, suggesting that they are members of the same 85-kDa gene family identified previously [7,8]. In addition, sequences within the major coding portion of the gene are present in 2 different size classes of RNA, while those sequences found in the 3′ non-coding region of the gene are present in only one size class of RNA.

Materials and Methods

Parasite strains and culture. The *T. cruzi* Peru strain was obtained from Stuart M. Krassner, University of California, Irvine. Clonal lines of this strain were established from individual parasites which were isolated by micromanipulation using procedures provided by James Dvorak, National Institutes of Health, Bethesda, MD. Peru clone 3 was utilized for these studies. The cloned *T. cruzi* lines Esmeraldo clone 3 and Silvio X10 clone 1 were obtained from James Dvorak. Growth and maintenance of epimastigotes and tissue-culture derived trypomastigotes of these strains are as described elsewhere [11].

Nucleic acid isolation, radiolabeling, Southern and Northern transfer, and restriction enzymes. Parasite nuclear DNA, bacterial plasmid DNA and phage λ DNA were isolated as described previously [12,13]. Agarose gel electrophoresis of DNA, Southern transfer, Northern transfer, prehybridization, hybridization and filter washing were performed as described [14–16]. DNA restriction fragments were radiolabeled with [α-^{32}P]dNTP using a nick translation kit from Bethesda Research Laboratory (Gaithersburg, MD). Synthetic oligonucleotides were end-labeled using T4 polynucleotide kinase (Boehringer-Mannheim, Indianapolis, IN) and [γ-^{32}P]ATP [12]. All restriction enzymes were purchased from Boehringer-Mannheim and used as recommended.

Isolation of cDNA and genomic clones. A cDNA library constructed in phage λgt10 using trypomastigote poly(A)$^+$ RNA [8] was screened with a 27-nucleotide (nt) synthetic oligomer representing one unit of the tandem repeat array present in the 85-kDa surface protein gene [7]. Inserts present in phage that showed hybridization were excised,subcloned and characterized by both restriction enzyme mapping and direct nucleotide sequence analysis.

A Peru genomic library was constructed in λDASH (Stratagene, La Jolla, CA) using DNA isolated from culture form trypomastigotes. Genomic DNA was partially digested with endonuclease *Sal*I, size fractionated on a 1% agarose gel, and fragments in the size range 10–20 kb were excised

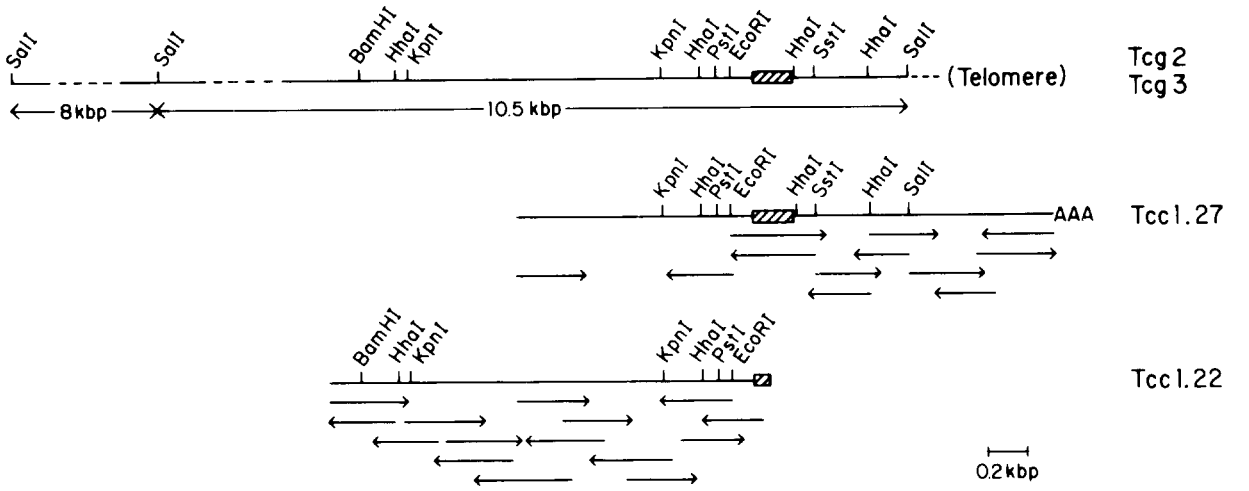

Fig. 1. Restriction map and sequencing strategy of the TSA-1 genomic clones Tcg 2 and Tcg 3 and cDNA clones Tcc 1.27 and Tcc 1.22. The hatched boxed regions denote the position of the 27-bp tandem repeat sequence.

and ligated into the *Sal*I site of the vector.

DNA sequencing. The strategy for determining the nucleotide sequence of the cDNA and genomic DNA fragments which encode TSA-1 is shown in Fig. 1. DNA sequencing was performed using the dideoxynucleotide chain termination method [17] with ^{32}P-labeled deoxynucleotide triphosphates and *T. cruzi*-derived DNAs inserted into the Bluescript vector (Stratagene, La Jolla, CA).

Results

Isolation of cDNA and genomic DNA encoding TSA-1. We previously reported the isolation and partial characterization of 25 recombinant λgt10 phage which contain the 27-bp repeat within the cDNA insert [8]. In order to select cDNAs which would be useful for determining the complete sequence of the 85-kDa trypomastigote specific surface antigen (TSA-1) gene, the nucleotide sequence of the 5′ and 3′ ends of each of the *T. cruzi* DNA inserts was ascertained. Of the 25 cDNAs examined, two, designated Tcc 1.22 and Tcc 1.27, were chosen for further analysis (Fig. 1). Tcc 1.22 was selected because the 8 bases at its 5′ terminus are identical to those at the 3′ end of the miniexon [19], indicating that it likely contains the 5′ most sequences present in the mature TSA-1 mRNA. Tcc 1.22, however, does not contain the 3′ terminus of the mature TSA-1 mRNA. Synthesis of the cDNA insert appears to have initiated within the 27-bp tandem repeat units, since sequences at the 3′ terminus of Tcc 1.22 align perfectly with the 5′ sequences present in Tcg-1, a cloned 500-bp genomic fragment containing the tandem repeat motif of the gene [7]. It is also possible that cDNA synthesis initiated 3′ downstream of the repeat motif and that only partial synthesis of the second strand of the cDNA occurred. To obtain the sequence of the 3′ portion of TSA-1, the nucleotide sequence of the DNA insert in phage Tcc 1.27 was determined. The 5′ end of Tcc 1.27 extensively overlaps with the 3′ end of the insert in Tcc 1.22, and its 3′ end terminates in a poly(A) stretch which is identical in position to 11 other cDNAs which also terminate in a stretch of A residues.

To obtain genomic DNA fragments which contain the TSA-1 gene, a detailed restriction map of the cDNA inserts in Tcc 1.22 and Tcc 1.27 was generated. The results of this analysis revealed the presence of a single *Sal*I restriction site approximately 700 bp upstream of the 3′ end of the insert in Tcc 1.27. The absence of a second *Sal*I site within the cDNA inserts suggested that a genomic *Sal*I DNA fragment should contain approximately 3 kb of the TSA-1 gene as well as additional bases 5′ upstream of the *trans*-splicing site. This prediction was confirmed by restriction enzyme analysis of total genomic DNA which revealed the presence of a *Sal*I

site approximately 10.5 kb upstream of the *Sal*I site in Tcc 1.27. To clone and isolate the genomic copy of TSA-1, a recombinant library of *Sal*I-digested genomic DNA was constructed in λDASH. Approximately 400000 recombinant phage were obtained and 150000 were screened by hybridization with the 27-nt repeat unit. Four phage showed positive signals upon subsequent rescreening. The *T. cruzi* DNA inserts in each phage were characterized by restriction enzyme mapping and partial nucleotide sequence analysis. Two of the phage contained single *Sal*I inserts of about 16 kb. Each insert also contained an internal 4.8-kb *Eco*RI fragment diagnostic of a non-telomeric, non-transcribed member of the subfamily [8]. The remaining 2 phage, designated Tcg-2 and Tcg-3 each contained *Sal*I inserts of 10.5 and 8.0 kb. Restriction mapping of total genomic DNA showed that both fragments lie adjacent within the genome with the 10.5-kb fragment mapping proximal to the telomere. The 10.5-kb *Sal*I fragment also was shown by hybridization analysis to contain sequences homologous to the 27-bp repeat. Therefore, it was excised, subcloned into the plasmid vector Bluescript and subjected to restriction enzyme analysis. As shown in Fig. 1, the restriction enzyme pattern of the 3′ region of the 10.5-kb fragment overlaps perfectly with that of the two cDNA inserts, suggesting that it is the genomic site of the TSA-1 gene.

Sequence analysis. In order to obtain information on the structure of the TSA-1 gene and the protein which it putatively encodes, the complete nucleotide sequence of the 2 cDNAs and selected regions of the 10.5-kb *Sal*I genomic DNA fragment were determined by the dideoxy chain termination method [17]. Also, the position of 6-bp restriction enzyme recognition sites predicted by either the sequence or restriction mapping was confirmed. The nucleotide sequence of the TSA-1 gene is shown in Fig. 2 and underscored by the predicted amino acid sequence of the protein.

The sequence of the cDNA differed from the genomic DNA only at 15 nucleotides at the 5′ terminus (Fig. 2), confirming our supposition that the 2 cDNAs are derived from the same gene. The first 8 nucleotides of the cDNA (i.e., −222 to −215) likely represent the synthetic *Eco*RI site introduced during construction of the library. The 8 nucleotides immediately following the *Eco*RI linker are identical to those on the 3′ terminus of the miniexon [19], suggesting that the *trans*-splicing site for miniexon addition is between nucleotides −207 and −206. Consistent with this assignment is the presence of an AG dinucleotide in the genomic DNA immediately 5′ of the putative splice junction. The 3′ end of the gene is tentatively defined by the presence of A residues at the 3′ terminus of the cDNA. Since the genomic region corresponding to this portion of the cDNA has not been cloned, a definitive assignment of the 3′ terminus cannot be made. Nevertheless, the comparison of twelve cDNAs which possess A residues at the 3′ terminus show no variation in the position of the stretch of A residues, suggesting that the site of post-transcriptional poly(A) addition is at or immediately downstream of the C residue at position 3495. Interestingly, the sequence AATAAA at position 3459 is identical to the consensus sequence AATAAA that precedes the polyadenylation site of most eukaryotic mRNAs by 12–40 nucleotides. However, since this sequence has not yet been observed in other *T. cruzi* genes [20], its presence here may be fortuitous and merely reflect the high concentration of A and T residues in the 3′ untranslated region of *T. cruzi* mRNA.

Fig. 2. The nucleotide sequence of the *T. cruzi* TSA-1 gene. The nucleotide sequence encompassing the TSA-1 gene starts at the *Eco*RI synthetic linker on the 5′ end of Tcc 1.22 and extends through the *Eco*RI synthetic linker on the 3′ end of Tcc 1.27 (Fig. 1). The proposed amino acid sequence of TSA-1 is given below the nucleotide sequence. Nucleotides 5′ upstream from the proposed translational initiation codon are given negative numbers, starting with −1. Nucleotides 3′ downstream from the start codon are given positive numbers starting with the A base in the proposed ATG translational initiation codon. Numbering of amino acids starts at the Met at nucleotide position 1. A 23-nt sequence of genomic DNA is shown above the 5′ end of the cDNA sequence. The boxed region at the beginning of the sequence denotes the *Eco*RI synthetic linker at the 5′ end of Tcc 1.22, and underlining marks the 3′ end of the miniexon sequence. The downward arrow marks the possible site of cleavage of the protein signal sequence. The 5 boxed regions starting at nucleotide 2011 denote the 27-bp tandem repeat regions. Potential *N*-glycosylation site are marked by underlining. The horizontal bracket marks the potential polyadenylation signal at the 3′ end.

```
    TTTGCTTTCCTTCAAGAAGGAAG --- Genomic DNA
-222 GGAATTCCCTATATTGAAGGAAGTTACTACCAAACGCTAAACGCTATTGTGAAATTCGAAGGTACTGAAAGGACCAACAACAGCACACACACCGCCCA -125
    EcoRI    3'end of
    linker   mini-exon

-124 CACACATATTTATATGCTCTCACGTGTTGCTGCTGTGAAGGCACCCCGCACACACAACCGTCGCCGCGTGACCGGATCCAGCGGAAGGAGGGAAGGAG -27

                              -1  1
 -26 GAGAGAGTGAGCCGCAGAGGCCCAAC  ATG TCC CGG CGT CAC TTC TAT TCT GCG GTG CTG CTT CTA CTC CTC GTC GTG ATG 54
                                MET SER ARG ARG HIS PHE TYR SER ALA VAL LEU LEU LEU LEU LEU VAL VAL MET
                                  1                                   10

  55 GTT TGC GGC GGC AGT GGA GCT GCG CAC GCT GTG GAG AGA AAC TCA GGG GAT TTG CAA CTG CCG CAG GAG ATC GCC 129
     VAL CYS GLY GLY SER GLY ALA ALA HIS ALA VAL GLU ARG ASN SER GLY ASP LEU GLN LEU PRO GLN GLU ILE ALA
          20                                  30                                  40

 130 ATG CTT GTG CCG AAT AAG ACG CAG GTG GTA CCA AAA AGT GGT GGT GAA GGC AAA GTG AAG GAT ATC TTT GCT TCA 204
     MET LEU VAL PRO ASN LYS THR GLN VAL VAL PRO LYS SER GLY GLY GLU GLY LYS VAL LYS ASP ILE PHE ALA SER
                     50                                  60

 205 CCC GCT CTC GTC CGT GCT GGT GGA GTG ATG ATT GCA TTT GTC GAA GGC CGC ACC AAA AAT AAA CTT TTT CCA GAA 279
     PRO ALA LEU VAL ARG ALA GLY GLY VAL MET ILE ALA PHE VAL GLU GLY ARG THR LYS ASN LYS LEU PHE PRO GLU
          70                                  80                                  90

 280 GTG ATT GAT CTC AGT TCT TCT GAT ATT GTT GCC GGT TAC ATC AAG GCT CCA GAG ACA TGG CAG TCT CTT GTT GCT 354
     VAL ILE ASP LEU SER SER SER ASP ILE VAL ALA GLY TRY ILE LYS ALA PRO GLU THR TRP GLN SER LEU VAL ALA
                     100                                 110

 355 GAG GTA ACC AAA GAG TAC TGG CAG GCA CAC ACT GTC CTT GAG AGT GCG AAT AAT AGT AAT CAT CGT GTG GGT GTT 429
     GLU VAL THR LYS GLU TYR TRP GLN ALA HIS THR VAL LEU GLU SER ALA ASN ASN SER ASN HIS ARG VAL GLY VAL
          120                                 130                                 140

 430 GCG AGG CTA CCC ACC GGA ATT ACA AGG GGC AAT AAA GTG TTT CTG CTA GTG GGG AGC TAT GAA GAG AGG CGT GAA 504
     ALA ARG LEU PRO THR GLY ILE THR ARG GLY ASN LYS VAL PHE LEU LEU VAL GLY SER TYR GLU GLU ARG ARG GLU
                     150                                 160

 505 ATT GAT GAT TAT ATT TGG AAG GCC GAG GCA TGG AAC ATT AAA GTG ATT GAG GGT GAG GCC ACG CAG TCC ACG GAA 579
     ILE ASP ASP TYR ILE TRP LYS ALA GLU ALA TRP ASN ILE LYS VAL ILE GLU GLY GLU ALA THR GLN SER THR GLU
          170                                 180                                 190

 580 GTT CAG CCG ACT CAA CCG ATC AAC TGG AGT GAA CCC AAA CCG CTG TTC CAA ACT GAC TCT CCT AAT AAT AAA GGT 654
     VAL GLN PRO THR GLN PRO ILE ASN TRP SER GLU PRO LYS PRO LEU PHE GLN THR ASP SER PRO ASN ASN LYS GLY
                     200                                 210

 655 GAC CTA AAG GAA TTT TTG GGT GGT GGT GGC TCA GGA ATT GTG ATG GGG AAT GGC ACA CTT GTG TTT CCC CTG ACG 729
     ASP LEU LYS GLU PHE LEU GLY GLY GLY GLY SER GLY ILE VAL MET GLY ASN GLY THR LEU VAL PHE PRO LEU THR
          220                                 230                                 240

 730 GCA AAG GAT GAA AGT AAT AAA GTT TTC TCC CTA ATC ACT TAT TCG ACG GAC GAC GGC CAA AAG TGG GAG ATA CCA 804
     ALA LYS ASP GLU SER ASN LYS VAL PHE SER LEU ILE THR TRY SER THR ASP ASP GLY GLN LYS TRP GLU ILE PRO
                     250                                 260

 805 GGG GGC GTT TCT TCT GTG GCA TGC CGT TCC CCC CGC GTC ACC GAA TGG GAG GAG GGA ACA CTT CTC ATG GTT ACT 879
     GLY GLY VAL SER SER VAL ALA CYS ARG SER PRO ARG VAL THR GLU TRP GLU GLU GLY THR LEU LEU MET VAL THR
          270                                 280                                 290

 880 TAT TGC GAG GAT GGC CGC AAG GTG TTT GAG TCG CGT GAC ATG GGG AAA ACG TGG ACG GAG GCT TTT GGG ACA CTC 954
     TYR CYS GLU ASP GLY ARG LYS VAL PHE GLU SER ARG ASP MET GLY LYS THR TRP THR GLU ALA PHE GLY THR LEU
                     300                                 310

 955 CCA GGC GTG TGG CTC AAA TCA GGT CCA GAA CTT CCG GAA GTA AGT TTG CGT GTT GAT GCT CTC ATC ACC GCG ACC 1029
     PRO GLY VAL TRP LEU LYS SER GLY PRO GLU LEU PRO GLU VAL SER LEU ARG VAL ASP ALA LEU ILE THR ALA THR
          320                                 330                                 340

1030 ATT GAG GGA AGG AAG GTC ATG CTG TAC ACT CAG AAA GTG AGG CAT TTT CTG GAA GTG GAT GAA CCC AAC GCA CTC 1104
     ILE GLU GLY ARG LYS VAL MET LEU TYR THR GLN LYS VAL ARG HIS PHE LEU GLU VAL ASP GLU PRO ASN ALA LEU
                     350                                 360

1105 CAC CTT TGG GTC ACG GAC AAC AAC CGC ACT TTT CAT CTT GGA CCG TTT TCT GTG GAC TGT GCT GAG AAT AAG ACG 1179
     HIS LEU TRP VAL THR ASP ASN ASN ARG THR PHE HIS LEU GLY PRO PHE SER VAL ASP CYS ALA GLU ASN LYS THR
          370                                 380                                 390

1180 TTT GCC AAC ACC TTG CTG TAC TCG GAT GAT GCG TTG CAC CTT TTA CAA GCG AAG GGC GAT CAT GAA AGC ACA GCC 1254
     PHE ALA ASN THR LEU LEU TYR SER ASP ASP ALA LEU HIS LEU LEU GLN ALA LYS GLY ASP HIS GLU SER THR ALA
                     400                                 410
```

```
1255 GTT TCA CTT GCC CGC CTA ACG GAG GAG CTG AAT ACG ATC AAT TCT GTC CTT AGT ACT TGG GTA CAG TTG GAC GCT 1329
     VAL SER LEU ALA ARG LEU THR GLU GLU LEU ASN THR ILE ASN SER VAL LEU SER THR TRP VAL GLN LEU ASP ALA
         420                                 430                                 440

1330 TCC TTC TCC GAG TCG TCT ATA CCC ACG GCT GGT CTG GTT GGA TTC CTG TCC AAT ACG ACG TCC AGT GGA GAC ACG 1404
     SER PHE SER GLU SER SER ILE PRO THR ALA GLY LEU VAL GLY PHE LEU SER ASN THR THR SER SER GLY ASP THR
                             450                                 460

1405 TGG ATC GAC GGG TAC CGT TGC ATG AAT GCA ACG GTG ACG AAG GCA GCG AAG GTT GAA AAT GGT TTC AAG TTC ACG 1479
     TRP ILE ASP GLY TRY ARG CYS MET ASN ALA THR VAL THR LYS ALA ALA LYS VAL GLU ASN GLY PHE LYS PHE THR
         470                                 480                                 490

1480 GGG CCT GGG TCC AGG GCA ACA TGG CCC GTA AAC AGT CGG TGG GAT ATT AAA CAG TAC GGC TTT GTG GAT TAC AAC 1554
     GLY PRO GLY SER ARG ALA THR TRP PRO VAL ASN SER ARG TRP ASP ILE LYS GLN TYR GLY PHE VAL ASP TRY ASN
                             500                                 510

1555 TTC ACT ATT GTG GCG ATG GCG ACT ATA CAC CAG GTT CCG AGT GAG AGC ACT CCT CTG CTG GGT GCG AGT CTG AGG 1629
     PHE THR ILE VAL ALA MET ALA THR ILE HIS GLN VAL PRO SER GLU SER THR PRO LEU LEU GLY ALA SER LEU ARG
         520                                 530                                 540

1630 GGC AAT AAG AGG ACG AAG TTA ATT GGT TTG TCG TAC GGT GCG GGC GGT AAG TGG GAG ACA GTG TAT GAC GGG ACA 1704
     GLY ASN LYS ARG THR LYS LEU ILE GLY LEU SER TYP GLY ALA GLY GLY LYS TRP GLU THR VAL TYR ASP GLY THR
                             550                                 560

1705 AAA ACA GTA CAG GGT GGC ACT TGG GAG CCG GGG AGA GAA TAC CAG GTG GCG CTC ATG CTG CAG GAC GGC AAC AAG 1779
     LYS THR VAL GLN GLY GLY THR TRP GLU PRO GLY ARG GLU TRY GLN VAL ALA LEU MET LEU GLN ASP GLY ASN LYS
         570                                 580                                 590

1780 GGC TTC GTG TAC GTG GAT GGT GTG CTT GTG GGG AAC CCG GCG ATG TTA CCA ACA CCT GAG GAG CGG TGG ACT GAA 1854
     GLY PHE VAL TYR VAL ASP GLY VAL LEU VAL GLY ASN PRO ALA MET LEU PRO THR PRO GLU GLU ARG TRP THR GLU
                             600                                 610

1855 TTC TCA CAT TTC TAC TTT GGG GGC GAC GAG GGA GAC AGC GGC AGC GAT GCG ACG TTG ACG GAC GTC TTT CTG TAC 1929
     PHE SER HIS PHE TYR PHE GLY GLY ASP GLU GLY ASP SER GLY SER ASP ALA THR LEU THR ASP VAL PHE LEU TYR
         620                                 630                                 640

1930 AAC CGC CCA CTG AGT GTC GGT GAA CTA AAA ATG ATC AAG GAA GTT GAA GAT AAA AAA GAA AAG GGA AGC GGT GAC 2004
     ASN ARG PRO LEU SER VAL GLY GLY LEU LYS MET ILE LYS GLU VAL GLU ASP LYS LYS GLU LYS GLY SER GLY ASP
                             650                                 660

2005 AGT GAA GAT AAA AAA GAA AGC GGT GAC AGT GAA GAC AAA AAA GAA AGC GGT GAC AGT GAA GAC AAA AAA GGA AGC 2079
     SER GLU ASP LYS LYS GLU SER GLY ASP SER GLU ASP LYS LYS GLU SER GLY ASP SER GLU ASP LYS LYS GLY SER
         670                                 680                                 690

2080 GGT GAC AGT GAA GAT AAA AAA GAA AGC GGT GAC AGT GAA GAT AAA AAA GAA AGC GGT GAC AGT GAA GAT AAA AAA 2154
     GLY ASP SER GLU ASP LYS LYS GLU SER GLY ASP SER GLU ASP LYS LYS GLU SER GLY ASP SER GLU ASP LYS LYS
                             700                                 710

2155 GGA AGC GGT GAC GGC GCA TTC ACA CCA GCG GTG TCC AAT GCC ACG ACA CAC ACA GCA GAG GAG GAG ACC GTA AAC 2229
     GLY SER GLY ASP GLY ALA PHE THR PRO ALA VAL SER ASN ALA THR THR HIS THR ALA GLU GLU GLU THR VAL ASN
         720                                 730                                 740

2230 CAG TCG GCA TCT GGA ACA TTT TCA ATT ACC GAC AGT ACT GAG GGT GAC GTG AGC TCT GAT GAG AAT GGG GAG ACG 2304
     GLN SER ALA SER GLY THR PHE SER ILE THR ASP SER THR GLU GLY ASP VAL SER SER ASP GLU ASN GLY GLU THR
                             750                                 760

2305 ACG GGA GGA GCT GAT GGG CAA GAG GAA GAT ATC CAG CCA CAG GAC GGG GAA GCA AAT GCT GCG GCA CTC GGC CTC 2379
     THR GLY GLY ALA ASP GLY GLN GLU GLU ASP ILE GLN PRO GLN ASP GLY GLU ALA ASN ALA ALA ALA LEU GLY LEU
         770                                 780                                 790

2380 GCA CTC AAA AGC AGT CTT GGA ACT TCG TCG CAG TGG GAT GGC AGC GTT GCT GGC ACC ATG CGT GAG AGC AGG GTG 2454
     ALA LEU LYS SER SER LEU GLY THR SER SER GLN TRP ASP GLY SER VAL ALA GLY THR MET ARG GLU SER ARG VAL
                             800                                 810

2455 CTG CTG CCA TCG CTG TTC CTT CTG TTG GGA CTC TGG GGG TTT GCG GCT CTG TGA GGAGTGTGCGGCCTCAGTGTGGGACAC 2535
     LEU LEU PRO SER LEU PHE LEU LEU LEU GLY LEU TRP GLY PHE ALA ALA LEU  *
         820                                 830                 835

2536 TTGCGCCCTCCCCCACACCGACGCACATCCGGTTATGAATTACTCTCCGCAATAAACAATACCTTTTTTTTTTATTATTGCTTGTTGCCTTGGGGCAAC 2634

2635 CACTATGAACTTTTTTTCTTTTCTACTTTTCCCTTTTTAATTGTTTTTTGCCGCACTCTTCCTGTGAGGGGTGCCGTGTGTTTCCCGCCCAACTGCTCC 2733

2734 ACTCGCACACCCACCGACACGCTCATGTCGACGGCCCTGTGCGCCTCGCCTGCTGCACGCATGAGGAGGGGCCGCGACGAACGGCGCGCTGAGGGAGCC 2832

2833 GTGACGTGTGGGCGTGCGACGAGGGCCGGTGGGGCGGAGCACACATTTCTTTCCCTCTTGTTTTGTTTTTGGCGGTCAGGGCGGGCAGACACCCGCAGC 2931

2932 GCTACACACACACAGCCACACCCTGCCGCCGTGTTTTTCTTTTTTTTTTTTTCGTGCGCTCATACTTCCAATGTAAAGACACAACTGCCGTCATGTCTG 3030

3031 CGTGCTTCTCTCAGCCCCCACACATACACACGAGTGACTGTGCTCTGTAGGCCGCACTCACGCACGCCGATGGGAGGGCTGAGGGGAGCAGCAACACGC 3129

3130 GGATGACCCGTGAAATATAAACGATGTGCGGACACGTGGAGGGTCAGACGGTTGAAAGACACGTAAACATATTTTTAATTTGTGGACAAAATATGCAGA 3228

3229 CAATTTTTGGAACAACATGGTAGTTAGGTTCTTTTTTTGTGTTTGTGTTTTTTCTTTTTTTTTTTTTGCGTGTGTGTGAAACAGTAATAGAGGACGACAA 3327

3328 TTTTTTTTTTTTTTTTGCGTGTGTGTGGAACAGTAATAGAGGACAACCTTTTTCTTTTTTAATAATTATTTTGTGTGAGGGGGTTAATATATGATGAAA 3426

3427 GGGAATTGTTTTTTCTAAATTATTTTGAATAAATTTTTTATGAATTGCCGCTGTTTGGTCTCGAGTACCAAAAAAAAAA 3505
```

TABLE I
Comparison of sequences 5′ of potential translation initiation sites for TSA1 with consensus initiation sequences of eukaryotic protein coding genes and the initiation sequence for other *T. cruzi* protein coding genes

Consensus sequence[a]		C	C	A/G	C	C	ATG
TSA-1	First ATG[b]	T	T	T	A	T	ATG
	Second ATG[c]	C	C	A	A	C	ATG
	Third ATG[d]	T	C	G	T	G	ATG
Hsp-85[e]		C	A	A	A	G	ATG
IF8[f]		A	C	G	A	G	ATG
pTC-FUS1[g]		A	A	A	C	C	ATG

[a]The rules used for assignment of consensus are described elsewhere [23].
[b]Nucleotides −116 to −109 in Fig. 2.
[c]Nucleotides −5 to 3 in Fig. 2.
[d]Nucleotides 47 to 54 in Fig. 2.
[e]The ATG initiation codon and the 5 nt immediately 5′ to this initiation site in the *T. cruzi* 85-kDa heat shock protein, Hsp-85 [13].
[f]The ATG initiation codon and the 5 nt immediately 5′ to this initiation site in the *T. cruzi* Ca^{2+} binding protein gene, IF8 [21].
[g]The ATG initiation codon and the 5 nt immediately 5′ to this initiation site in the *T. cruzi* ubiquitin gene, pTc-FUS [22].

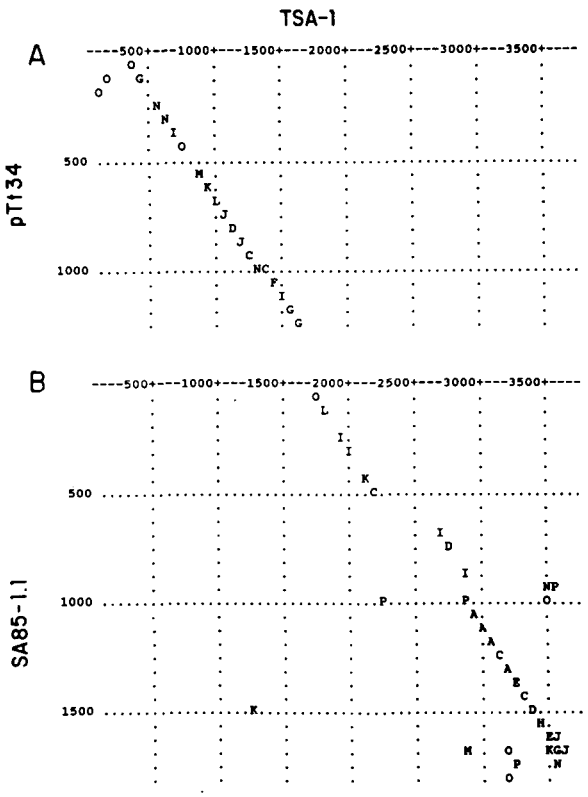

Fig. 3. Dot matrix comparison of TSA-1 and (A) pTt34 and (B) SA85-1.1 nucleotide sequences. Windows of 31 nt are sequentially compared and a letter scored for any 21 identical nt (DNA/Protein Sequence Analysis System, International Biotechnologies, Inc., New Haven, CT). The letters represent varying degrees of homology among the 31 nt being scored. The letter A represents a match value of 100%, the letter B a value of 99–98%, with a progressive decrease in % homology to the letter S which represents 64–65% homology, or 21 of 31 nucleotides having an identical match.

Identification of the protein coding region. Translation of the TSA-1 DNA sequence reveals the presence of a single long open reading frame. Within this reading frame the first ATG is 112 bp downstream of the 5′ end at position −111 (Fig. 2). Translation starting at this site would end at nt position 2507 and yield an 872-amino acid (aa) protein of 94513 Da. A second and third ATG triplet are found in the TSA-1 transcript at nt positions 1 and 52, and are in the same reading frame as the first ATG. Translation starting at the second or third ATG would result in proteins of 835 aa and 818 aa, respectively, with predicted M_rs of 90429 and 88430. Each of the predicted M_rs are larger than the apparent M_r of 85000 previously reported [4–7,9,10], suggesting that processing of the primary translation product might be occurring. As discussed below, analysis of the putative translation product is in keeping with this suggestion. Examination of the 5′ sequences flanking the three potential translation start sites shows that only those sequences upstream of the second ATG strongly match the generalized eukaryotic consensus sequence proposed by Kozak [23] (Table I). Therefore, we have tentatively assigned the second ATG triplet as the start codon.

Nucleotide homology between TSA-1, pTt34 and SA85-1.1. Two DNA fragments, pTt34 [9] and SA85-1.1 [10], which represent parts of genes which encode 85-kDa trypomastigote specific surface antigens, have been cloned and characterized. A homology matrix analysis of nucleotide identities between TSA-1 and both pTt34 and SA85-1.1 is shown in Fig. 3. Position homology between the 2 sequences is scored by placement of a symbol if 21 out of 30 nt (i.e., 70%) showed identities in a sequential scan of the DNA sequence. In the comparison

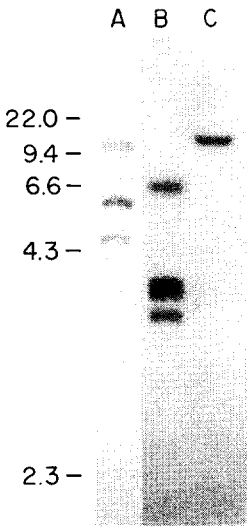

Fig. 4. A Southern blot containing 10 μg per lane of trypomastigote nuclear DNA digested with *Eco*RI from the following strains of *T. cruzi* was hybridized with ^{32}P-labeled 27-nt synthetic oligomer. (A) Peru DNA; (B) Esmeraldo DNA; (C) Silvio X10 DNA. Numbers in kb on the margin refer to the migration of ^{32}P-labeled *Hin*dIII fragments of λ phage DNA.

with pTt34, it is clear that considerable homology exists with TSA-1 in the predicted coding region in the 5′ end of the gene. Most of the scored homologies lie on a single continuous linear axis with only a few regions being identified elsewhere within the sequence. The high degree of homology observed in this alignment (i.e., about 68%) indicates that this portion of the coding region of the 2 genes has been maintained in length with few, if any large deletions or gene rearrangements.

In the comparison with SA85-1.1, most of the homology exists within the extreme 3′ end of the predicted coding region and the contiguous noncoding region of TSA85-1 with only limited homology being evident elsewhere within the coding region. The 2 regions of high homology observed within the predicted coding region of TSA-1 occur within regions of 43 bp (nt 1901–1944) and 84 bp (nt 2424–2508) which show 91% and 70% homology, respectively. Not all of the scored homologies lie on a linear array, suggesting that these regions of the 2 genes have been conserved but some rearrangements have occurred in the form of deletions and/or insertions. Similar results were obtained when the nucleotide sequence of TSA-1 was compared with the sequence of 2 other members of the SA85-1 gene family, SA85-1.2 and SA85-1.3 (data not shown).

Strain-specific expression of the TSA-1 gene subfamily. Our previous results with the Peru strain have shown that the 27-bp repeat unit in the TSA-1 gene defines a subfamily of a larger 85-kDa gene family. To determine whether the repeat motif which defines this subfamily was present in strains of *T. cruzi* other than Peru, a Southern blot containing total genomic DNA from the Peru, Esmeraldo and Silvio X10 strains was hybridized with the 27-nt repeat unit (Fig. 4). Hybridization was observed to 4 *Eco*RI restriction fragments in both the Peru and Esmeraldo DNA and to a single *Eco*RI fragment in Silvio X10 DNA, indicating that the TSA-1 subfamily is present in all 3 strains of the parasite.

To determine whether members of the subfamily in the Esmeraldo and Silvio X10 strains were also transcribed, the 27-nt repeat unit was hybridized to Northern blots containing poly(A)$^+$ RNA from each of the strains. As shown in Fig. 5A, the 27-nt repeat unit hybridized to RNA of 3.7 kb from both the Peru and Esmeraldo strains and to RNA size 3.4 kb from the Silvio X10 strain, suggesting that at least 1

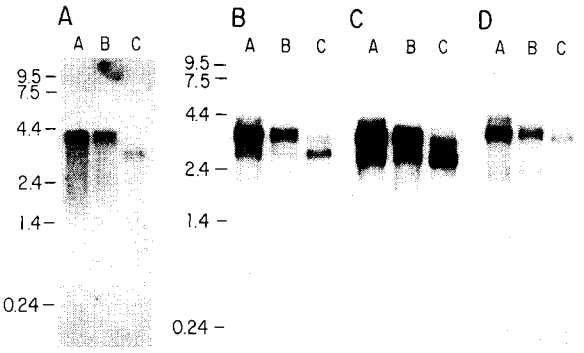

Fig. 5. Identification of poly(A)$^+$ RNA from trypomastigotes of three strains of *T. cruzi* complementary to various regions of TSA-1. In the Northern blots in inserts A–D, lane A contains 2 μg of poly(A)$^+$ RNA from the Peru strain; lane B contains 2 μg of poly(A)$^+$ RNA from the Esmeraldo strain; and lane C contains 2 μg of poly(A)$^+$ RNA from the Silvio X10 strain. Numbers on the right refer to the migration of RNA size standards (Bethesda Research Laboratories, Gaithersburg, MD). The Northern blots were hybridized with (A) ^{32}P-labeled 27 nt oligomer; (B) [^{32}P]Tcc 1.22 and [^{32}P]Tcc 1.27; (C) ^{32}P-labeled *Bam*HI/*Eco*RI fragment containing nt −51 to 1857 of TSA-1 (Fig. 2); (D) ^{32}P-labeled *Sal*I/*Eco*RI fragment containing nt 2761 to 3505 of TSA-1 (Fig. 2).

member of this subfamily is being expressed in each of the 3 strains.

The 85-kDa gene family is expressed in more than 1 size class of poly(A)$^+$ RNA. Our previous studies suggested that in the Peru strain sequences homologous to members of the TSA-1 gene family may be present in two size classes of poly(A)$^+$ RNA, of average length 3.7 and 3.4 kb [8]. The repeat motif, however, was observed only in the 3.7-kb class of RNA. In view of the current observation that in the Silvio X10 strain the repeat motif is present only in the 3.4-kb class of RNA, and the fact that our previous studies included hybridization data using only a limited region of the TSA-1 gene (i.e., nucleotides 1852–2342 in Fig. 2), we have extended these studies to include sequences 5′ upstream and 3′ downstream of this 490 bp region. As shown in Fig. 5B, the full length transcript of TSA-1 hybridized to 2 differently sized RNAs in each of the 3 strains. In the Peru and Esmeraldo strains, TSA-1 shows a strong hybridization signal with poly(A)$^+$ RNA of 3.7 kb and a less intense signal with poly(A)$^+$ RNA of 3.4 kb. However, in contrast to the hybridization profile observed with the Peru and Esmeraldo RNAs, the length of the larger RNA species in Silvio X10 (i.e., 3.6 kb) is slightly less than observed in the other 2 strains, and the hybridization signal of the 3.4-kb RNA species is considerably more intense than that observed with the 3.6-kb RNA.

To further define which regions of TSA-1 are represented in the two poly(A)$^+$ RNA size classes, selected regions of the TSA-1 gene were hybridized to Northern blots containing poly(A)$^+$ RNA from the 3 strains. As shown in Fig. 5C, the 1.8-kb *Bam*HI/*Eco*RI fragment which contains most of the 5′ end of the gene (i.e., nucleotides –51 to 1857 in Fig. 2) hybridized to both RNA classes in all 3 strains. As previously observed with the full length TSA-1 gene, the intensity of the hybridization signal was greatest with the 3.7-kb RNA species from Peru and Esmeraldo strains and the 3.4-kb RNA species from the Silvio X10 strain. In contrast, the *Sal*I/*Eco*RI fragment which contains 744 bp of the 3′ end of the gene (i.e., nucleotides 2761–3505 in Fig. 2) hybridized only to the 3.6 and 3.7 kb RNA (Fig. 5D). These results and those reported previously [8] indicate that sequences found 5′ of the repeat motif in the TSA-1 transcript share homology with 2 different size classes of RNA, while both the repeat motif and sequences 3′ of the repeat are present only in one size class of RNA.

BAL 31 nuclease sensitivity of the TSA-1 homologues. Of the 4 *Eco*RI restriction fragments which contain the 27-bp repeat motif in the Peru strain, 1 has been shown to be telomeric in location and has been implicated as the site of transcription of the TSA-1 RNA [8]. The remaining 3 fragments are not located near a telomere and appear to be either transcriptionally silent or transcribed at a very low level. We have examined the possibility that 1 or more of the *Eco*RI fragments containing the repeat motif in the Esmeraldo strain, and the single *Eco*RI fragment in the Silvio X10 strain, are telomeric by the technique of preferential sensitivity to digestion with BAL 31 nuclease

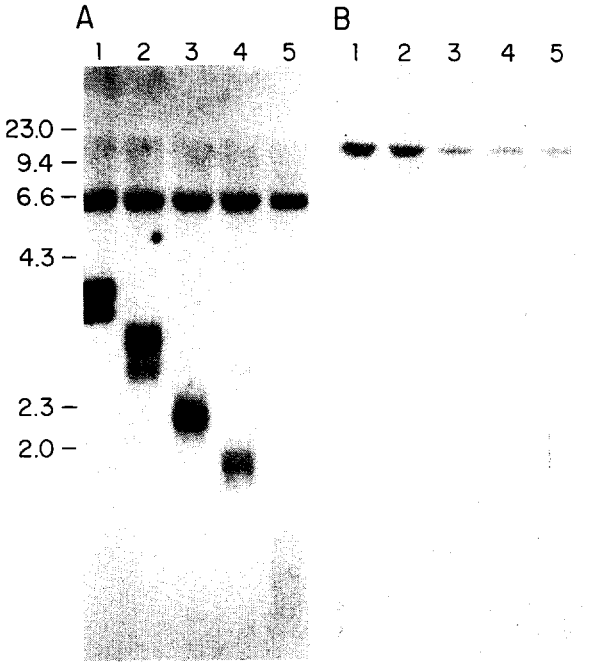

Fig. 6. Nuclease BAL 31 sensitivity of genomic sequences with homology to the 27-bp repeat unit. Aliquots of trypomastigote genomic DNA from the Esmeraldo strain (A) and the Silvio X10 strain (B) were digested for increasing times with BAL 31 nuclease, restricted with *Eco*RI, electrophoresed on an agarose gel and probed with the ^{32}P-labeled 27-nt oligomer. Numbers in kb on the margin refer to DNA size markers as in Fig. 4.

[24]. As shown in Fig. 6, 3 of the 4 *Eco*RI fragments in the Esmeraldo strain are preferentially sensitive to BAL 31 digestion, which suggests that these three members of the subfamily are telomeric in location. In contrast, the single *Eco*RI fragment in the Silvio X10 strain is not preferentially sensitive to BAL 31 nuclease, which suggests that the subfamily in Silvio X10 has no telomeric member.

Discussion

Extensive searches of data banks and the literature have failed to reveal any strong homology between TSA-1 and other genes or proteins whose biological function is known. Therefore, we do not have any substantial clues as to the function of TSA-1. However, the search did reveal that TSA-1 has extensive sequence homology with two other cDNA fragments which have been shown to encode 85 kDa trypomastigote specific surface antigens, pTt34 and SA85-1.1 [9,10]. Although SA85-1.1 was believed to be a member of a gene family other than that containing either pTt34 or TSA-1 [10], it now seems likely that TSA-1, pTt34 and SA85-1.1 are members of the same multigene family. Therefore, it is quite possible that the differences observed in the physical properties of the 85-kDa surface antigen [4–6] are due to diversity within this single gene family.

Although the functions of the major 85-kDa surface glycoproteins of *T. cruzi* are not known, biological and physical properties of these glycoproteins have been reported [2–6]. It is therefore worthwhile to determine how certain features of the predicted protein sequence of TSA-1 relate with these properties. The N-terminus contains a hydrophobic region which is compatible with an N-terminal signal peptide [25], with a reasonable candidate signal peptide processing site being present at residue 29 [26,27]. The protein contains no hydrophobic region at the COOH-terminus which would serve as a transmembrane domain, but it does possess a hydrophobic stretch of amino acids at the COOH-terminus that could serve as a processing site for a phosphatidylinositol linkage [28]. As noted previously [7], the antigen contains a 9-amino-acid repetitive peptide sequence proximal to the COOH-terminus of the protein. Of the 9 residues in the repeat unit, 4 are charged, making this region of the protein ostensibly hydrophilic. Accordingly, computer analysis shows that the region of the protein containing the repeat motif is extensively hydrophilic and would likely represent an area of high antigenicity. Consistent with this interpretation, synthetic oligopeptides containing 2 repeat units arranged in a head-to-tail tandem array are recognized by antibodies from Chagasic patients and mice infected with *T. cruzi* (Wrightsman and Manning, unpublished). Finally, the protein contains several potential sites for attachment of *N*-linked glycosyl moieties. Each of these properties are consistent with the observation that the 85-kDa protein is a highly antigenic surface glycoprotein.

Transcription of TSA-1. Our previous results with the Peru strain have shown that of the four members of the TSA-1 gene family which are defined by the presence of the 27-bp repeat motif, only the single telomeric member appears to be transcribed into poly(A)$^+$ RNA [8]. While this also appears to be true for the Esmeraldo strain (Fouts and Manning, unpublished), it is clearly not the case for the member of the subfamily found in the Silvio X10 strain. BAL 31 analysis of Silvio X10 genomic DNA reveals the presence of a single member of the subfamily which is located at a non-telomeric site. Northern blot analysis indicates that this member is transcribed into an abundant poly(A)$^+$ RNA, thus indicating that members of this subfamily need not necessarily be located at a telomeric site in order to be transcriptionally active.

It is clear that at least two different poly(A)$^+$ RNA size classes share sequence homology with TSA-1 (Fig. 5) and that the presence of the 27-bp repeat unit in these 2 size classes of RNA differs among strains. In the Peru and Esmeraldo strains, the sequences within the coding region of the TSA-1 gene which are 5′ upstream of the 27-bp repeat are represented in both the 3.7- and 3.4-kb RNAs, while the repetitive sequence and those sequences downstream of the repeat share homology only with the 3.7-kb poly(A)$^+$ RNA. Conversely, in the Silvio X10 strain the repeat motif is present only in the smaller 3.4-kb RNA while those sequences 3′ of the repeat motif in TSA-1 are present only in the larger 3.6-kb RNA. It is very clear, therefore, that the sequences present 3′ downstream of the repeat motif in the Peru and Esmeraldo gene(s) are not pre-

sent in the Silvio X10 transcript.

Although we do not yet understand the relationship between the two classes of RNA which encode this gene family, 2 possibilities present themselves with regards to the origin of the transcripts. One is that the 3.4-kb RNA is a processed form of the 3.7-kb transcript. Alternatively, the 2 RNA classes could originate from different genes within the family. Although we cannot formally exclude the possibility of processing, we favor the suggestion that the 2 classes of RNA arise independently by transcription of different genes. This is based on the observation that although the 2 RNA classes differ by only 300 bp in the Peru and Esmeraldo strains, the 3.7-kb RNA shares homology with sequences found throughout the 1.5-kb 3′ terminus of TSA-1, while the 3.4-kb RNA shows no homology with this 1.5-kb region of the gene. It is difficult to imagine, therefore, how the 3.7-kb transcript can be processed to yield the 3.4-kb RNA. The question also arises as to whether the different size RNAs might provide alternate properties to the protein. Several feasible possibilities present themselves and are currently being investigated.

Acknowledgements

We thank Barbara Granger for assistance in maintenance and growth of parasite cultures, Stuart Kahn and Harvey Eisen for thoughtful suggestions and sharing unpublished data, and Jim Massey for technical assistance. This work was supported by Public Health Service grant AI18873 from the National Institutes of Health.

References

1 Brener, Z.A. (1973) Biology of *Trypanosoma cruzi*. Annu. Rev. Microbiol. 27, 347–383.

2 Zingales, B., Andrews, N.W., Kuwajima, V.Y. and Colli, W. (1982) Cell surface antigens of *Trypanosoma cruzi*: possible correlation with the interiorization process in mammalian cells. Mol. Biochem. Parasitol. 6, 111–124.

3 Katzin, A.M. and Colli, W. (1983) Lectin receptors in *Trypanosoma cruzi*, and *N*-acetyl-D-glucosamine-containing surface glycoprotein specific for the trypomastigote stage. Biochim. Biophys. Acta 727, 403–411.

4 Andrews, N.W., Katzin, A.M. and Colli, W. (1984) Mapping of surface glycoproteins of *Trypanosoma cruzi* by two-dimensional electrophoresis. Eur. J. Biochem. 140, 599–604.

5 Alves, M.J.M., Abuin, G., Kuwajima, V.Y. and Colli, W. (1986) Partial inhibition of trypomastigote entry into cultured mammalian cells by monoclonal antibodies against a surface glycoprotein of *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 21, 75–82.

6 Abuin, G., Colli, W., de Souza, W. and Alves, M.J.M. (1989) A surface antigen of *Trypanosoma cruzi* involved in cell invasion (Tc-85) is heterogeneous in expression and molecular constitution. Mol. Biochem. Parasitol. 35, 229–237.

7 Peterson, D.S., Wrightsman, R.A. and Manning, J.E. (1986) Cloning of a major surface antigen gene of *Trypanosoma cruzi* and identification of a nonapeptide repeat. Nature 322, 566–568.

8 Peterson, D.S., Fouts, D.L. and Manning J.E. (1989) The 85-kd surface antigen gene of *Trypanosoma cruzi* is telomeric and a member of a multigene family. EMBO J. 8, 3911–3916.

9 Takle, G.B., Young, A., Snary, D., Hudson, L. and Nicholls, S.C. (1989) Cloning and expression of a trypomastigote-specific 85-kilodalton surface antigen gene from *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 37, 57–64.

10 Kahn, S., Van Voorhis, W.C. and Eisen, H. (1990) The major 85 kD surface antigen of the mammalian form of *Trypanosoma cruzi* is encoded by a large heterogeneous family of simultaneously expressed genes. J. Exp. Med. 172, 589–597.

11 Beard, C.A., Wrightsman, R.A. and Manning, J.E. (1985) Identification of monoclonal antibodies against the trypomastigote stage of *Trypanosoma cruzi* by the use of iminobiotinylated surface polypeptides. Mol. Biochem. Parasitol. 16, 199–212.

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

13 Dragon, E.A., Sias, S.R., Kato, E.A. and Gabe, J.D. (1987) The genome of *Trypanosoma cruzi* contains a constitutively expressed, tandemly arranged multicopy gene homologous to a major heat shock protein. Mol. Cell. Biol. 7, 1271–1275.

14 Thomas, P.S. (1980) Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA 77, 5201–5205.

15 Rozek, C.E. and Davidson, N. (1983) *Drosophila* has one myosin heavy-chain gene with three developmentally regulated transcripts. Cell 32, 23–34.

16 Beard, C.A., Wrightsman, R.A. and Manning, J.E. (1988) Stage and strain specific expression of the tandemly repeated 90 kDa surface antigen gene family in *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 28, 227–234.

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

18 Hall, R., Hyde, J.E., Goman, M., Simmons, D.L., Hope, I.A., Mackay, M., Scaife, J.H., Merkli, B., Richle, R. and Stocker, J. (1984) Major surface antigen gene of a human malarial parasite cloned and expressed in bacteria. Nature 311, 379–382.

19 McCarthy-Burke, C., Taylor, Z.A. and Buck, G.A. (1989)

Characterization of the spliced leader genes and transcripts in *Trypanosoma cruzi*. Gene 82, 177–189.

20 Zwierzynski, T.A., Widmer, G. and Buck, G.A. (1989) In vitro 3′ end processing and poly(A) tailing of RNA in *Trypanosoma cruzi*. Nucleic Acids Res. 17, 4647–4660.

21 Gonzalez, A., Lerner, T.J., Heucas, M., Sosa-Pineda, B., Nogueira, N. and Lizardi, P.M. (1985) Apparent generation of a segmented mRNA from 2 separate tandem gene families in *Trypanosoma cruzi*. Nucleic Acids Res. 13, 5787–5804.

22 Swindle, J., Ajioka, J., Eisen, H., Sanwal, B., Jacquenmot, C., Browder, Z. and Buck, G.A. (1988) The genomic organization and transcription of the ubiquitin genes of *Trypanosoma cruzi*. EMBO J. 7, 1121–1127.

23 Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. 12, 857–872.

24 De Lange, T. and Borst, P. (1982) Genomic environment of the expression-linked extra copies of genes for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. Nature 299, 451–453.

25 von Heijne, G. (1985) Signal sequences: the limits of variation. J. Mol. Biol. 184, 99–105.

26 Perlman, D. and Halvorson, H.O. (1983) Putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. 167, 391–409.

27 von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. 14, 4683–4690.

28 Ferguson, M.A.J. and Williams, A.F. (1988) Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. Annu. Rev. Biochem. 57, 285–320.