ELSEVIER

Molecular and Biochemical Parasitology 63 (1994) 109–120

MOLECULAR AND BIOCHEMICAL PARASITOLOGY

Expression and evolution of members of the *Trypanosoma cruzi* trypomastigote surface antigen multigene family

Barbara J. Ruef[1], Bradley D. Dawson, Devansu Tewari, David L. Fouts, Jerry E. Manning*

Department of Molecular Biology and Biochemistry, University of California, Irvine CA 92717, USA

(Received 11 October 1993; accepted 22 October 1993)

Abstract

The trypomastigote specific surface antigens of *Trypanosoma cruzi* are encoded by a supergene family which includes the TSA family. The TSA family is characterized by the presence of a 27-bp tandem repeat array in the coding region. Here, we report the characterization and analysis of the three TSA family members in the Esmeraldo strain of the parasite. In this strain 2 distinct telomeric members are expressed abundantly as 3.7-kb mRNAs, while the remaining member is located at an internal chromosomal site and is expressed at less than 2% of the level seen for the telomeric members. Based on hybridization to DNA separated by PFGE, 3 chromosomes of sizes 1.8 Mb, 0.98 Mb, and 0.90 Mb each contain one of the telomeric members. In addition, the two smaller chromosomes also contain the single internal member. Since both chromosomes contain similar TSA family members, and vary only slightly in size, we suggest that they are homologues. Comparisons of the nucleotide sequences of the different members of the family show that the internal gene differs from the telomeric genes primarily in sequences found 3′ of the repeat array. These comparisons also reveal that the three genes are analogous, supporting the hypothesis that short segments between the family members are exchanged by gene conversion events. We propose that similar conversion events between members of different gene families may generate some of the diversity found within the supergene family.

Key words: Trypanosoma cruzi; Gene expression; Gene conversion

1. Introduction

Studies on the surface glycoproteins of the bloodstream trypomastigote stage of the parasitic protozoan *Trypanosoma cruzi* indicate that several of these molecules may be involved in processes of parasite recognition and penetration of the host cell [1–8]. Additional studies [9] have shown that many, and possibly all, of these trypomastigote surface glycoproteins are encoded by a supergene

*Corresponding author. Tel.: 714-856-5578; Fax: 714-856-8551.

Abbreviations: aa, amino acids; nt, nucleotide; PFGE, pulse field gel electrophoresis; TSA, trypomastigote surface antigen.

[1] *Present address:* Seattle Biomedical Research Institute, Seattle, WA, USA

Note: Nucleotide sequence data reported in this paper have been submitted to the GenBank™ data base with the accesion numbers UO2613, UO2614, UO2615.

SSDI 0166-6851(93)E0178-B

family. Members of the superfamily share about 30% amino acid identity and all possess a partially conserved sequence (VTVxNVfLYNR) near the COOH terminus of the protein. The family members also contain partial or complete copies of the motif SxDxGxTW. Interestingly, this motif was first identified in bacterial neuraminidases [10], suggesting that genes encoding neuraminidase may be found within the supergene family.

Indeed, the SAPA/TCNA (shed-acute phase antigen/*Trypanosoma cruzi* neuraminidase) gene family has been shown to contain most of the neuraminidase/trans-sialidase activities of the parasite. While criteria to definitively assign individual gene families within the superfamily have not been firmly established, it has been possible to define at least 2 gene families by the presence of tandemly repeated amino acid motifs. The SAPA/TCNA gene family shares a repeat motif of 12 amino acids [11], and one or more members encode the unique 160-kDa sialic acid-transferring enzyme, trans-sialidase [12,13]. Members of the first gene family identified in *T. cruzi* share a different repeat motif of 9 amino acids and encode trypomastigote-specific surface antigens (TSA) of 85–110 kDa [14].

In previous studies we demonstrated that the gene TSA-P1 in the Peru strain contains a tandemly repeated 27-bp sequence within the coding region which defines this 85-kDa gene family [15]. The repeat unit hybridizes to 4 genomic *Eco*RI fragments in the Peru and Esmeraldo strains of the parasite, and to one *Eco*RI fragment in the Silvio strain [16]. Bal 31 nuclease studies revealed that, of the four fragments defined by *Eco*RI digests of Peru and Esmeraldo DNA, one fragment in Peru and 3 fragments in Esmeraldo were Bal 31 nuclease sensitive, indicating that these fragments are located at or near a telomere. The remaining family members in Peru and Esmeraldo, and the solitary Silvio member, were insensitive to Bal 31 nuclease suggesting these fragments are found at internal chromosomal locations.

This report focuses on the relationship between the members of the TSA gene family in the Esmeraldo strain in order to elucidate the mechanisms underlying the evolution of the gene family and the larger superfamily. Based on the data presented, we hypothesize that concerted evolution of the TSA family is occurring, most likely by the process of gene conversion, and we propose that diversity within the superfamily may be generated by conversion events between members of different gene families.

2. Materials and methods

2.1. 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 were as described elsewhere [17].

2.2. Pulsed field gel electrophoresis (PFGE). DNA was prepared from late exponential phase epimastigotes in low-melting-point agarose (InCert Agarose, FMC Bioproducts) at a concentration of 2×10^9 cells ml^{-1} as described [18]. Electrophoresis was performed in 1.0% agarose gels immersed in $0.5 \times$ TBE buffer (90 mM Tris-borate, 2.5 mM EDTA, pH 8.0) with a BioRad CHEF DR II unit (BioRad) for 24–48 h at 14°C using 200 volts with a switch time of 60 and 90 s. DNA molecular weights were estimated using standards prepared from *Saccharomyces cerevisiae*. Pulse-field gel electrophoresis (PFGE) gels were stained with ethidium bromide, photographed, destained for 20 min and blotted to nylon membranes following depurination with 0.25 M HCl.

Isolation of individual DNA bands from PFGE gels was accomplished by excision of the gel region and electroelution in the CHEF DR II system for 24 h using the conditions described above. The DNA was further purified by treatment with proteinase K followed by extraction with phenol-chloroform.

2.3. Nucleic acid isolation and analysis. Parasite nuclear DNA and poly A^+ RNA, phage lambda DNA, and bacterial plasmid DNA were isolated as described previously [19]. Agarose gel electrophoresis on 1.0% agarose was used to separate nucleic acids. Southern transfer to nylon membranes, prehybridization, hybridization, and wash conditions were carried out as previously reported. Hybridization with the 27 nucleotide repeat unit was performed at 37°C in 30% formamide. Hybridization of oligonucleotide probes specific to the three Esmeraldo genes was performed at 42°C in 30% formamide.

2.4. Oligonucleotide synthesis, radiolabeling, and restriction enzymes. Oligonucleotides used for probes or as sequencing primers were synthesized on the Gene Assembler Plus (Pharmacia, Piscataway, NJ) according to the manufacturer's instructions. Synthetic oligonucleotides were radiolabeled by end-labeling using [γ-^{32}P]ATP and T4 polynucleotide kinase. DNA restriction fragments were radiolabeled using a nick translation kit (BRL, Gaithersburg, MD) as recommended with [α-^{32}P]dCTP incorporation. All restriction enzymes were purchased from Boehringer-Mannheim and used as recommended.

2.5. Genomic and cDNA library construction and isolation of recombinant phage. Esmeraldo nuclear DNA was digested with the selected restriction enzymes and ligated into the λFIX replacement vector (Stratagene, La Jolla, CA). A cDNA library was constructed in phage λgt10 using trypomastigote poly A^+ RNA from the Esmeraldo strain and cDNA synthesis kits from Pharmacia and BRL. Recombinant phage from the genomic and cDNA libraries were screened by plating the phage, transferring to nitrocellulose filters, and screening with a radiolabeled 27-mer representing one unit of the repeat array present in the coding region of the TSA-P1 gene from the Peru strain. Positive λ phage were plaque purified and *T. cruzi* DNA inserts from each phage were excised and subcloned into the Bluescript plasmid vector (Stratagene) for nucleotide sequencing and restriction enzyme analysis.

2.6. DNA sequencing and sequence analysis. DNA sequencing was performed using the dideoxy chain termination method [20] using a T7 polymerase sequencing kit from Pharmacia as recommended. All sequencing reactions were carried out on *T. cruzi* derived DNA fragments inserted into the Bluescript plasmid using [α-^{32}P]dATP for incorporation. Nucleotide sequence data was compiled and analyzed using the IBI Pustell Sequence Analysis programs for the IBM computer (New Haven, CT), and the Clustal V programs [21].

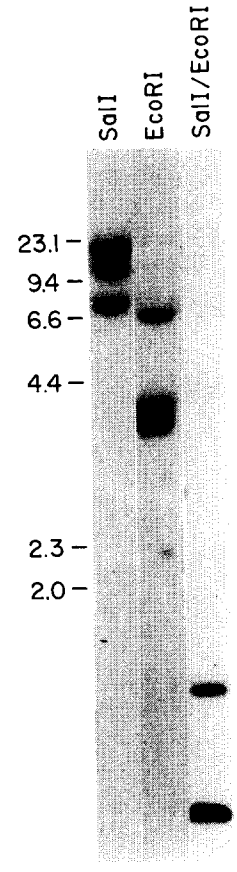

Fig. 1. Southern blot analysis of genomic Esmeraldo strain DNA. Nuclear DNA from Esmeraldo strain *T. cruzi* was digested with *Sal*I, *Eco*RI, and a double digest of *Sal*I/*Eco*RI restriction enzymes, blotted and hybridized with a [γ-^{32}P]ATP end-labeled synthetic oligonucleotide corresponding to one repeat unit of the 27 nucleotide repeat array. Numbers to the left are sizes of λ *Hin*dIII markers in kb.

3. Results

3.1. Genomic organization of the TSA family in Esmeraldo. Previous studies identified 4 genomic *Eco*RI restriction fragments from the Esmeraldo strain which hybridize with the 27 nucleotide repeat unit [15]. Three of these *Eco*RI fragments, of sizes 3.2, 3.4 and 3.6 kb, are sensitive to Bal 31 nuclease digestion and are inferred to be located at telomeric sites. The fourth fragment, of size 6.0 kb, shows no sensitivity to Bal 31 nuclease digestion and is likely located at an internal chromosomal site. Since direct cloning of the telomeric genes as *Eco*RI fragments is not possible, cloning of DNA fragments containing all or part of these four *Eco*RI fragments was accomplished using an approach similar to that previously used for the cloning of the telomeric and internal members of the TSA genes in the Peru strain [16].

In the Peru strain, a single *Sal*I restriction site separates the telomere from the *Eco*RI site found within the telomere associated gene, TSA-P1. A second *Sal*I site is located several kilobases upstream of the 5′ splice site of TSA-P1, allowing the entire coding region of the gene to be cloned as a single *Sal*I fragment. To determine whether a similar pattern of restriction enzyme sites is present in the telomeric genes of the Esmeraldo strain, genomic DNA was digested with either *Sal*I, *Eco*RI or *Sal*I/*Eco*RI, Southern blotted and hybridized with the 27 nucleotide repeat unit. As shown in Fig. 1, 3 *Sal*I fragments of sizes 9, 13 and 20 kb, and 4 *Eco*RI fragments of sizes 3.2, 3.4, 3.6 and 6.0 kb hybridized. In the *Sal*I/*Eco*RI digest, only 2 fragments of sizes 0.9 and 1.7 kb are seen. Since both fragments are smaller than any of the *Eco*RI fragments, each telomeric gene must possess a *Sal*I site between the *Eco*RI site and the telomere. Also, since only 3 *Sal*I fragments hybridized, it is probable that at least 2 of the four members defined by *Eco*RI digestion share the same *Sal*I restriction pattern.

A *Sal*I genomic library was constructed in order to clone the *Sal*I fragments by digesting Esmeraldo nuclear DNA with *Sal*I, purifying fragments of size 9–23 kb, and ligating into the λFIX replacement vector. 200 000 recombinant phage were screened with the repeat unit and 11 recom-

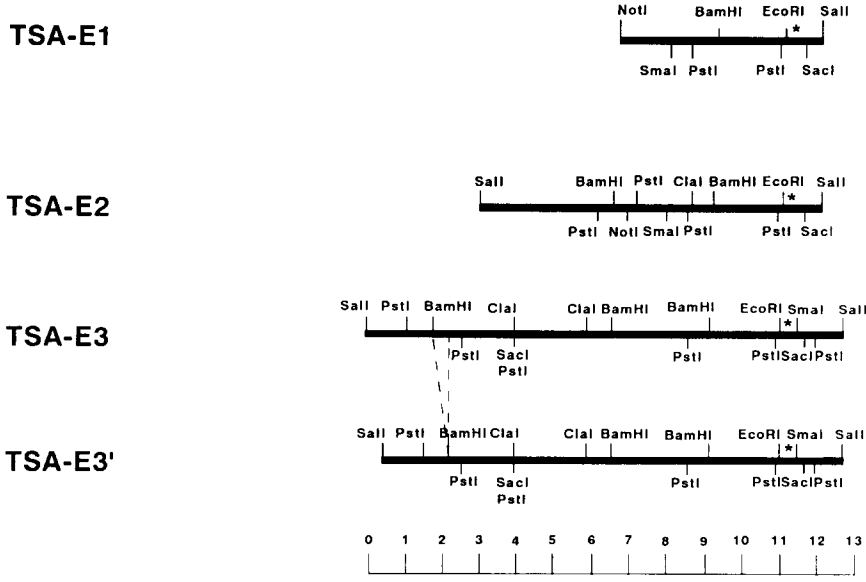

Fig. 2. Restriction maps of the TSA family members in the Esmeraldo strain of *T. cruzi*. TSA-E1 was cloned originally as a 20-kb *Sal*I fragment, but was subcloned and restriction mapped as a 5-kb *Sal*I/*Not*I fragment. TSA-E2 was cloned and mapped as a 9-kb *Sal*I fragment. Two isolates of TSA-E3 were cloned and upon restriction analysis were found to differ by a single 400-bp deletion/insertion. The location of the deletion/insertion is represented as dashed lines between TSA-E3 and TSA-E3′. The location of the 27 nucleotide repeat array is shown as an asterisk (*). The scale is shown below in kb.

binants rescreened positive. *Sal*I digests of the recombinant phage DNA isolated from the 11 positives revealed 1 containing a 9.0-kb insert, 2 containing a 20-kb insert, and the remaining 8 containing a 13-kb fragment. Restriction enzyme analysis of the two 20-kb inserts showed the *Sal*I fragments to have identical restriction patterns. Restriction enzyme analysis of the eight 13-kb inserts showed 2 variants which differ by a single 400-bp deletion/insertion (Fig. 2).

3.2. Chromosome mapping and analysis. In order to determine the relationship between the *Sal*I and *Eco*RI fragments, mapping of these fragments to chromosomal size molecules was undertaken. Fig. 3A shows PFGE of chromosome size DNA molecules from the Esmeraldo strain. Approximately 15 DNA bands resolved ranging in size from 0.40 to 2.2 megabases (Mb). Hybridization of the 27 nucleotide repeat unit to a Southern blot of this gel showed positive signals with DNA fragments of size 0.90, 0.98 and 1.8 Mb. In order to assign the *Eco*RI and *Sal*I fragments to their respective chromosomes, the gel regions containing the three DNA bands were excised. The DNA in each band was isolated, digested with *Eco*RI or *Sal*I, electrophoresed on an agarose gel and Southern blotted to nylon membranes. Hybridization of the Southern blot with the 27 nucleotide repeat unit (Fig. 3B) shows that the 3.6-kb *Eco*RI fragment is

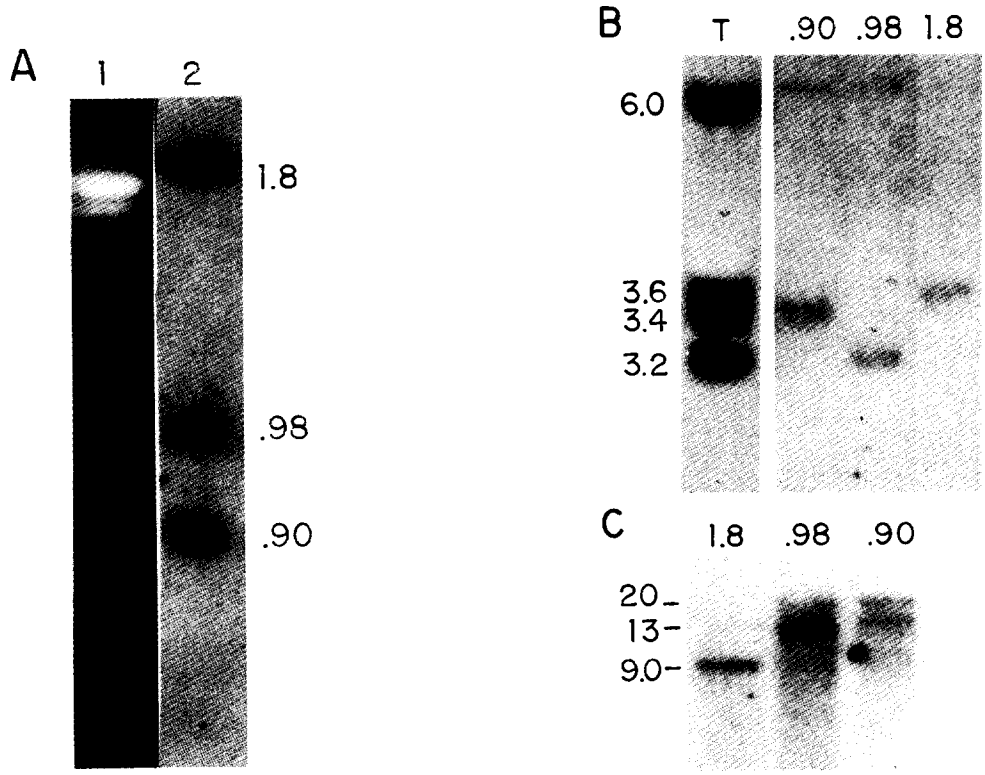

Fig. 3. Chromosomal distribution of members of the TSA-1 subfamily in the Esmeraldo strain of *T. cruzi*. (A) Lane 1 depicts chromosome sized molecules separated by PFGE for 48 h and stained with ethidium bromide. Lane 2 shows the same gel after Southern transfer and hybridization with [γ-^{32}P] end-labeled 27 nucleotide repeat unit oligomer. The numbers to the right indicate size in Mb of the chromosome sized molecules that hybridized. (B) Lane T is a Southern blot of total genomic DNA, restricted with *Eco*RI, and hybridized with the end-labeled 27 nucleotide repeat unit. The remaining lanes contain DNA isolated from the individual chromosomes (indicated by the number in Mb on top of each lane) that were seen to hybridize with the repeat unit. DNA was excised from the gel, restricted with *Eco*RI, Southern transferred, and hybridized with the repeat unit. (C) Southern blot of chromosomal DNA excised from the gel, digested with *Sal*I, and hybridized with the repeat unit. Numbers to the left of (B) and (C) represent the size (in kb) of the DNA fragments seen to hybridize with the repeat unit.

found in the 1.8 Mb chromosome, the 6.0-kb *Eco*RI fragment is present in both the 0.90 and 0.98 Mb chromosomes, and the 3.2- and 3.4-kb fragments are found in the 0.98 and 0.90 Mb chromosomes, respectively. The results of the *Sal*I digestion of DNA isolated from the three chromosomes (Fig. 3C) shows that the 9.0-kb *Sal*I fragment is present in the 1.8 Mb chromosome, and that the 13-kb and 20-kb *Sal*I fragments are present on both the 0.90 and 0.98 Mb chromosomes.

The chromosomal map shown in Fig. 4 was constructed based on these results. The 1.8 Mb chromosome contains the 9.0-kb *Sal*I fragment and the 3.6-kb telomeric *Eco*RI fragment. The 0.98 Mb and 0.90 Mb chromosomes contain both the 20-kb and 13-kb *Sal*I fragments as well as the 6.0-kb internal *Eco*RI fragment. The maps of the two chromosomes differ in the *Eco*RI fragments found at each telomere, since the 0.98 Mb chromosome contains the 3.2-kb *Eco*RI fragment whereas the 0.90 Mb chromosome contains the 3.4-kb *Eco*RI fragment. The assignment of the 13-kb *Sal*I fragment to an internal chromosomal location is based upon direct nucleotide sequence analysis of it as well as the 6.0-kb *Eco*RI fragment, as discussed below. Also, we have tentatively depicted the 0.98 Mb and 0.90 Mb chromosomes as homologues based upon their similarity in size and the fact that they possess similar TSA family members and restriction enzyme patterns.

3.3. Nucleotide and amino acid sequence. The nucleotide sequence of the TSA family members contained in each of the *Sal*I fragments and the 6.0-kb *Eco*RI fragment were determined using synthetic oligonucleotide primers. Alignment of the nucleotide sequence of the three genes is shown in Fig. 5. The positions of 6-bp restriction enzyme recognition sites predicted by restriction mapping or sequencing were confirmed. The nucleotide sequence of the family member present in the 6.0-kb *Eco*RI fragment showed 100% identity with the family member present in the 13-kb *Sal*I fragment, providing direct evidence for assignment of the 13-kb *Sal*I fragment to an internal chromosomal site (Fig. 4). Translation of the gene sequences in each of the 6 possible reading frames revealed only one large open reading frame (ORF) in each sequence. The remaining 5 reading frames each contain numerous stop codons distributed throughout the predicted amino acid sequences. The putative initiation codon for the ORFs is at nucleotide 1 in all three sequences (Fig. 5), and

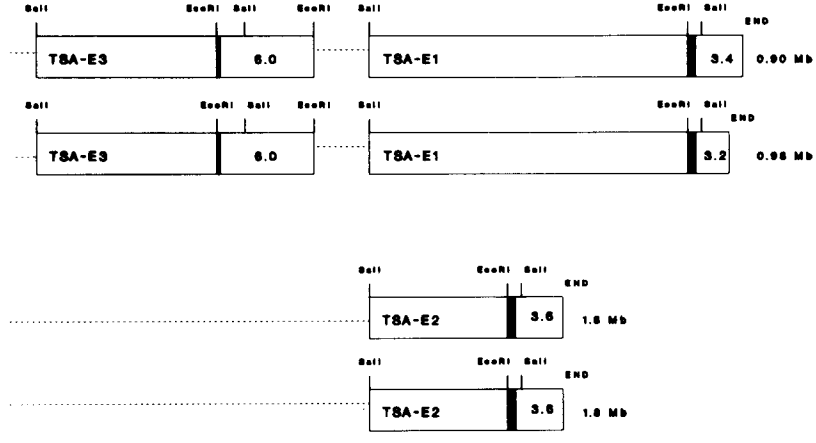

Fig. 4. Schematic representation of the Esmeraldo chromosomes containing the 27-bp repeat unit. Two sets of homologues are shown with the TSA family members mapped to each. The dashed lines represent the undefined distance between the TSA-E3 and TSA-E1 genes. The boxed regions represent a composite fragment including the *Sal*I genomic fragments and the *Eco*RI fragments containing the three subfamily members. The vertical lines within the boxed regions represent the repeat units. The numbers at the 3′ terminus of each gene refer to the size, in kb, of the *Eco*RI fragments assigned to each chromosome. END denotes the telomere of each chromosome.

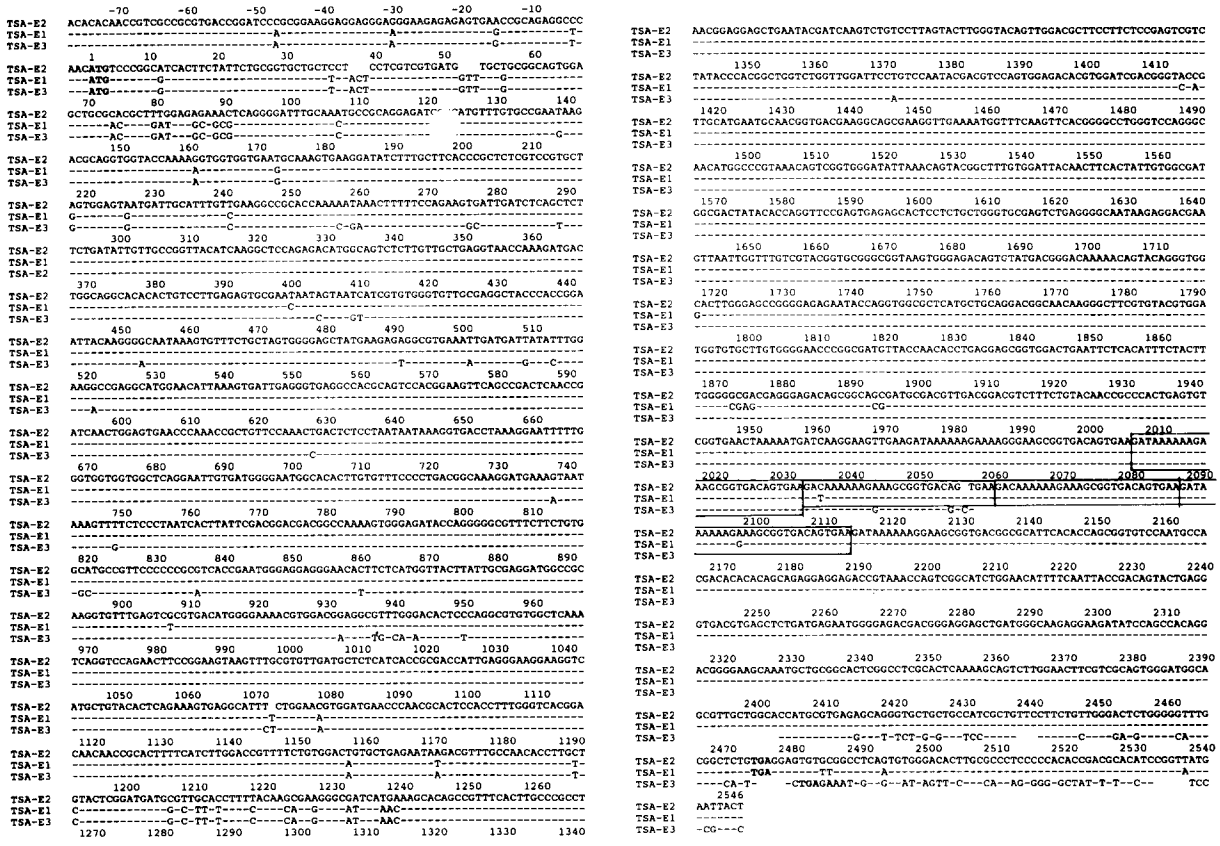

Fig. 5. Comparison of the nucleotide sequence of the three Esmeraldo TSA family members. Nucleotides 5′ upstream from the proposed translational initiation codon are given negative numbers, starting with −1. Nucleotides 3′ downstream from the start codon are given positive numbers starting with base A in the proposed ATG translational initation codon. Indicated in bold letters are the intiating codon (ATG) and the terminating codon (TGA). The boxed regions starting at nucleotide 2005 denote the 27-bp tandem repeat regions. The dotted line indicates the sequence of either TSA-E1 or TSA-E3 that does not differ from TSA-E2.

the sequence immediately upstream of the ATG codons for each gene meets the requirements of a eukaryotic translation start site [22,23]. Translation of each of the three genes would result in proteins with predicted M_rs of 89 168, 89 433 and 77 874 for genes TSA-E1, TSA-E2 and TSA-E3, respectively. These values are in agreement with those reported for other members of the gene family [16,24,25] and suggest that all copies are genes and not pseudogenes [26]. Consistent with this interpretation is the observation that the N-terminus of each gene contains a hydrophobic region which is compatible with an N-terminal signal peptide [Fig. 6; 27], as well as a hydrophobic stretch of amino acids at the COOH-terminus of each protein that could serve as a processing site

Amino Terminal Sequences:

```
TSA-E2   MSRHHFYSAVLLLL-VVM-CCGSGAAHALERN
         ::: :::::::::: ::: : ::::: : :
TSA-E1   MSRRHFYSAVLLLLLVVMVCGGSGAATADEAA
         ::::::::::::::::::::::::::::::::
TSA-E3   MSRRHFYSAVLLLLLVVMVCGGSGAATADEAA
```

Carboxyl Terminal Sequences:

```
TSA-E2   GTMRE--SRVLLPSLFLLLGLWGFAAL
         :::::  ::::::::::::::::::::
TSA-E1   GTMRE--SRVLLPSLFLLLGLWGFAAL
         : :::  ::::      :::: :: ::
TSA-E3   GSMREGMSRVL---PLLLLGMWGIAAIY
```

Fig. 6. Comparison of the predicted amino acid sequences at the amino terminus and carboxyl terminus of the three TSA genes. The aa sequences shown are those predicted by translation of nucleotides 1–90 and 2398–2472 for TSA-E2, nucleotides 1–96 and 2405–2478 for TSA-E1, and nucleotides 1–96 and 2059–2133 for TSA-E3. Colons represent aa identity.

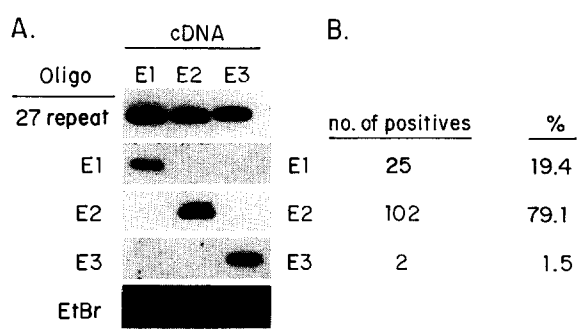

Fig. 7. Expression of the three members of the TSA-1 subfamily in the Esmeraldo strain of *T. cruzi*. (A) cDNA clones corresponding to the three subfamilily members, TSA-E1, TSA-E2, and TSA-E3 were restricted with *Eco*RI/*Sal*I, Southern transferred, and hybridized with [γ-^{32}P]ATP end-labeled 27 nucleotide repeat unit or end-labeled oligonucleotide specific for each individual member. The sequences of the member specific oligonucleotides are as follows: E1, 5′-CATTGAGGCCAAACACTCCTGACAG-3′; E2, 5′-CACTGAGGCCGCACACTCCTCACAG-3′; E3, 5′-CCCACATTTCATTGGGATGTTCAGTG-3′. The last panel is an ethidium bromide stained gel showing the relative amount of each band on the blot. (B) A λgt10 cDNA library made from Esmeraldo strain trypomastigote poly A$^+$ RNA was screened with the 27 nucleotide repeat unit oligomer and 129 plaques rescreened positive. These were then rescreened with oligomers E1, E2, and E3, and the results are shown as the number of positives that rescreen with the specific oligomer along with the percent of the total transcripts of the family that each represents.

for a phosphatidylinositol linkage [28].

3.4. Transcription of the family. To determine whether each of these three genes is actively transcribed and to assess their relative abundance in the stable A$^+$RNA population, two λgt10 cDNA libraries were constructed and 250 000 recombinants from each library were screened with the repeat unit. 129 plaques rescreened positive and, of these, 14 were selected for restriction enzyme and nucleotide sequence analysis. DNA from each was isolated, restricted with *Eco*RI/*Sal*I, electrophoresed in agarose, Southern blotted and hybridized with the 27 nucleotide repeat. Thirteen of the DNAs showed positive hybridization to a fragment of length 0.9 kb while one DNA showed hybridization to fragment of length 0.96 kb. Direct nucleotide sequence analysis of these DNA fragments indicated that the 0.9-kb fragment was an *Eco*RI/*Sal*I fragment identical in sequence to either gene TSA-E1 or TSA-E2. The 0.96-kb fragment, however, was an *Eco*RI fragment from gene TSA-E3 which terminated in a stretch of 15 A residues that have no counterpart in the genomic sequence. To determine the transcriptional origin of the remaining cDNA isolates, oligonucleotide probes representing unique regions in each of the three genes were synthesized and hybridization conditions were determined such that each probe hybridized only to its homologous gene (Fig. 7). Of the 129 plaque purified recombinants, 25 hybridized with the TSA-E1 specific oligo, 102 hybridized with the TSA-E2 specific oligo, and 2 hybridized with the TSA-E3 specific oligo. Comparison of the relative abundance of each of the three members in the mRNA population shows that the internal member, E3 is represented at less than 2% of the level of that seen for the telomeric

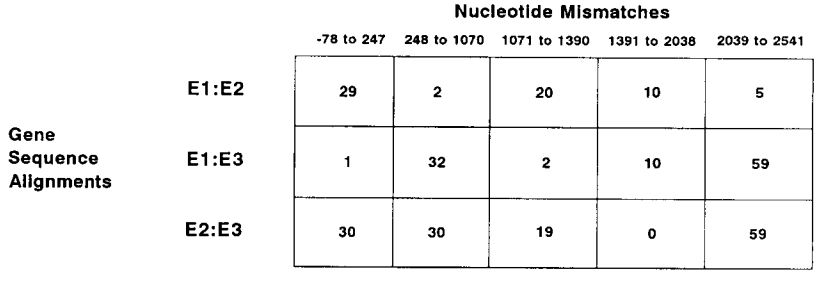

Fig. 8. Nucleotide sequence mismatches in blocks of sequence from the three TSA family members in the Esmeraldo strain of *T. cruzi*. The columns represent regions of sequence with the numbering system based on the nucleotide sequence of TSA-E2 (see Fig. 4). The numbers in each cell are the total number of nucleotide seqence mismatches in that region between the two aligned family members. The boundaries between the regions were chosen as the midpoints between nonidentities which denote a shift in maximal sequence similarity from one gene pair to a different gene pair.

members. Also, there is a five-fold difference in the levels of the telomeric members with TSA-E2 more highly represented than TSA-E1.

4. Discussion

In our previous studies, restriction enzyme analysis of genomic DNA indicated that the Esmeraldo strain of *T. cruzi* contained 4 members of the TSA-1 multigene family, 3 of which are located at telomeres. We have now determined the nucleotide sequence, chromosomal location and transcript abundance of each of these members. A primary finding of this study is the chromosomal organization of the gene family. The single family member located at an internal chromosomal site, TSA-E3, is present on 2 chromosomes of similar size, 0.98 and 0.90 Mb, each of which also contains a telomeric family member. The two telomeric members are distinguished because the COOH termini of their coding regions are found on distinct *Eco*RI restriction fragments (i.e., 3.4 and 3.2 kb), while the entire coding region of each telomeric member is present within either a 13.0- or 13.4-kb *Sal*I fragment. A fine scale restriction enzyme map of these two *Sal*I fragments showed that they are identical with the exception of a 0.4-kb insertion/deletion. In addition, partial nucleotide sequence analysis of the family member present in each *Sal*I fragment showed no differences in their nucleotide sequence. These findings suggest that the chromosomes containing the two telomeric *Eco*RI fragments are homologues, with each homologue containing one telomeric and one internal family member. This interpretation is in keeping with previous studies which indicate that *T. cruzi* is a diploid organism of which the homologues often differ in size [29] and is also consistent with previous observations in the African trypanosome, *Trypanosoma brucei*, which show that restriction fragments containing telomeres with the same surface antigen gene may vary in length due to the nature of the replication of chromosomal ends [30,31].

Transcription analysis of the family in Esmeraldo shows that A^+ RNAs from the internal gene, TSA-E3, are present at a level $<2\%$ of that seen for the telomeric members. This is also the case in the Peru strain where a similar study revealed that of 27 cDNAs which contained the repeat unit, only one represented a transcript from the internal gene(s). An internal location is not strictly correlated with the absolute level of transcription, since the single member in the Silvio strain is internal and transcripts of this member are detected in Northern blots of A^+ RNA [16]. Likewise, studies on related members of the multigene family demonstrated that genes from telomeric and internal locations can be expressed simultaneously [24]. However, within the TSA family it is clear that when both internal and telomeric genes are present in the same genome, transcripts from internal member(s) are much less abundant than transcripts from telomeric member(s). In addition, in the Esmeraldo strain a five-fold variation in the level of transcripts from the two telomeric members is seen, reaffirming our previous observation that the level of expression of different members of the gene family may vary and that such variation can be associated with the chromosomal location of the family members, possibly reflecting position effect.

Nucleotide and amino acid sequence comparisons show that the internal member differs from the two telomeric members primarily in the number of repeat units and in those sequences 3′ downstream of the repeat array. TSA-E1 and TSA-E2 each contain 4 tandemly repeated 27-bp sequences which are flanked by partial repeats, while TSA-E3 has only one complete repeat unit which is flanked by partial repeats. Most striking is the observation that the 345 bp immediately downstream of the degenerate repeat array in the telomeric genes (i.e. nucleotides 2056–2401, Fig. 5) are absent in the internal gene, TSA-E3. The simplest explanation for this observation is that the 345 bp have been deleted from the internal gene during evolution of the gene family. Consistent with this view is the observation that the remaining 23 amino acids in TSA-E3 show significant identity (i.e. 64%, Fig. 6) with the carboxyl terminus of both TSA-E1 and TSA-E2. It is curious, however, that the percent of sequence identity is significantly decreased from the $>91\%$ identity observed throughout the remainder of the protein. Also,

the abrupt loss of nucleotide identity between TSA-E3 and the telomeric genes immediately 3′ of the TGA stop codons is surprising (Fig. 5), particularly when viewed in comparison to the >92% sequence identity observed throughout the coding region of the genes as well as the non-translated region upstream of the ATG initiation codon. One reasonable explanation for these observations may be found in the genetic mechanisms hypothesized to facilitate maintenance of multigene families [32,33]. Gene conversion can best explain both the maintenance of a family of related genes as well as mediate diversification of particular genes within a gene family in several organisms [see reviews 34–38]. If gene conversion events are occurring between the different TSA family members, certain features of the sequence structure of the TSA genes would be predicted. In particular, gene conversion involving the internal gene TSA-E3 and the telomeric genes could account for the number of repeat units in TSA-E3 being less than the number present in either TSA-E1 or TSA-E2 if the 3′ most region of heteroduplex formation occurred within a repeat unit. Also, if genetic exchange between E3 and either E1 or E2 does not occur downstream of the repeat units, then sequences downstream of the repeat units in E3 may diverge independently of E1 and E2 provided the primary constraint is retention of the biological function encoded by that region of the gene. Secondly, if blocks of sequences rather than entire genes are being converted, patchy homology might be observed among the three genes and maximal sequence identity between the genes would be obtained if the genes were treated as analogues. As shown in Fig. 8, this appears to be the case. Starting at 78 bp upstream of the ATG initiation codon and extending to bp 255, TSA-E1 has greater identity with TSA-E3 than with TSA-E2. However, at bp 256 TSA-E1 begins to show a greater similarity with TSA-E2 than with TSA-E3 and this similarity extends through bp 1070 at which a shift back to the E1-E3 similarity is observed. At bp 1412 maximum identity again changes, with E3 being more identical to E2 than with E1. This similarity extends through bp 2042, at which E2 again becomes most similar to E1, and this similarity extends through the non-tranlated region of the two genes. This pattern of sequence similarity suggests that these genes are analogues, and is consistent with gene conversion events between the family members.

There is now substantial evidence that the natural propagation of *T. cruzi* is clonal, and that sexual reproduction is either absent or so rare as to leave no trace in the population structure of the parasite [39, 40]. Thus, gene conversion in *T. cruzi* most likely takes place during mitotic growth. While somatic conversion events generally occur at a frequency lower than that observed for meiotic exchange, they do have important biological consequences in many organisms. In particular, somatic gene conversion in chickens serves to introduce diversity by the exchange of short segments between family members of the immunoglobulin λ light chain locus [41, 42], and to provide diversification of the variable surface glycoprotein genes in *Trypanosoma equiperdum* and *Trypanosoma brucei* [43–45].

In summary, the results presented herein show that individual members of the TSA gene family do not diverge independently of each other, and therefore over evolutionary time the homogeneity of this family will not necessarily disappear by a slow accumulation of mutations in individual family members through genetic drift. At present, it is not possible to speculate on whether conversion-driven evolution of gene families, both tandemly repeated as well as dispersed, is a general phenomenon in this parasite. If this is the case, however, a pattern of sequence identity like that observed for the TSA family might be expected to be found within the SAPA/TCNA trans-sialidase multigene family as well as other families within the superfamily. Also, if gene conversion between members of *T. cruzi* gene families depends on tracts of sequence identity for initiation of recombination via homology-dependent strand transfer, as is the case for the chicken immunoglobulin genes, genetic diversity within the supergene family could be enhanced by exchange of gene segments between families which share even limited sequence identity. Such events would be expected to generate members of the superfamily which are dimorphic in that they would contain regions which have high

sequence similarity to one family, while immediately adjacent regions of the gene would exhibit substantially less similarity. Since these new constructs likewise would be presented to the host defense mechanisms, their rate of fixation within the population may be accelerated.

5. Acknowledgements

We thank B. Granger for technical assistance and J. Lucchesi, A. James and W. Fitch for many helpful discussions. This research was supported by grant AI18873 from the National Institutes of Health.

6. References

[1] 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.

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

[3] 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.

[4] Alves, M.J.M., Albuin, 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.

[5] Piras, M.M., Hendriquez, D. and Piras, R. (1987) The effect of fetuin and other sialoglycoproteins on the in vitro penetration of *Trypanosoma cruzi* trypomastigotes into fibroblastic cells. Mol. Biochem. Parasitol. 22, 135–143.

[6] Albuin, G., Colli, W., De Souza, W. and Alves, M.J.J. (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] Schenkman, S., Jiang, M.S., Hart, G.W. and Nussenzweig, V. (1991) A novel cell surface trans-sialidase of *Trypanosoma cruzi* generates a stage-specific epitope required for invasion of mammalian cells. Cell 65, 1117–1125.

[8] Schenkman, S., Kurosaki, T., Ravetch, J.V. and Nussenzweig, V. (1992) Evidence for the participation of the Ssp-3 antigen in the invasion of nonphagocytic mammalian cells by *Trypanosoma cruzi*. J. Exp. Med. 175, 1635–1641.

[9] Campatella, O., Sanchez, D., Cazzulo, J.J. and Frasch, A.C.C. (1992) A superfamily of *Trypanosoma cruzi* surface antigens. Parasitol. Today 8, 378–381.

[10] Roggentin, P., Rothe. B, Kaper, J.B., Galen, J., Lawrisuk, L., Virm, E.R. and Schauer, R. (1989) Conserved sequences in bacterial and viral sialidases. Glycoconj. J. 6, 349–353.

[11] Ibanez, C.F., Affranchino, J.L., Macina, R.A., Reyes, M.B., Leguizamon, S., Camargo, M.E., Aslund, L., Pettersson, U. and Frasch, A.C. (1988) Multiple *Trypanosoma cruzi* antigens containing tandemly repeated amino acid squence motifs. Mol. Biochem. Parasitol. 30, 27–33.

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

[13] Parodi, A.J., Pollevick, G.D., Mautner, M., Buschiazzo, A., Sanchez, D.O. and Frasch, A.C. (1992) Identification of the gene(s) coding for the trans-sialidase of *Trypanosoma cruzi*. EMBO J. 11, 1705–1710.

[14] 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.

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

[16] Fouts, D.L., Ruef, B.J., Ridley, P.T., Wrightsman, R.A. and Manning, J.E. (1991) Nucleotide sequence and transcription of a trypomastigote surface antigen gene of *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 46, 189–200.

[17] 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.

[18] Engman, D.M., Reddy, L.V., Donelson, J.E. and Kirchhoff, L.V. (1987) *Trypanosoma cruzi* exhibits inter- and intra-strain heterogeneity in molecular karyotype and chromosomal gene location. Mol. Biochem. Parasitol. 22, 115–123.

[19] Beard, C.A., Saborio, J.L., Tewari, D., Krieglstein, K.G., Henschen, A.H. and Manning, J.E. (1992) Evidence for two distinct major protein components, PAR1 and PAR2, in the paraflagellar rod of *Trypanosoma cruzi*; complete nucleotide sequence of PAR2. J. Biol. Chem. 267, 21656–21662.

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

[21] Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1991) CLUSTAL V: improved software for multiple sequence alignment. Comp. Appl. Biosci. 8, 189–191.

[22] Kozak, M. 1991, Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem.

266, 19867–19870.
[23] Cavener, D.R. and Ray, S.C. (1991) Eukaryotic start and stop translation sites. Nucleic Acids Res. 19, 3185–3192.
[24] Kahn, S., Van Voorhis, W.C. and Eisen, H. (1990) The major 85 kD surface antigen gene of the mammalian form of *Trypanosoma cruzi* is encoded by a large heterogeneous family of simultaneously expressed genes. J. Exp. Med. 172, 589–597.
[25] Takle, G.B. and Cross, G.A. (1991) An 85-kilodalton surface antigen gene family of *Trypanosoma cruzi* encodes polypeptides homologous to bacterial neuraminidases. Mol. Biochem. Parasitol. 48, 185–198.
[26] Takle, G.B., O'Connor, J., Young, A.J. and Cross, G.A.M. (1992) Sequence homology and absence of mRNA defines a possible pseudogene member of the *Trypanosoma cruzi* gp85/sialidase multigene family. Mol. Biochem. Parasitol. 56, 117–128.
[27] Von Heijne, G. (1985) Signal sequences: the limits of variation. J. Mol. Biol. 184, 99–105.
[28] Cross, G.A.M. (1990) Glycolipid anchoring of plasma membrane proteins. Annu. Rev. Cell Biol. 6, 1–39.
[29] Henrikkson, J., Aslund, L., Macina, R.A., Franke de Cassulo, B.M., Cazzulo, J.J., Frasch, A.C.C. and Petersson, U. (1990) Chromosomal localization of seven cloned antigen genes provides evidence for diploidy and further demonstration of karyotype variability in *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 42, 213–223.
[30] Bernards, A., Michels, P.A.M., Lincke, C.R. and Borst, P. (1983) Growth of chromosome ends in multiplying trypanosomes. Nature 303, 592–597.
[31] Van der Ploeg, L.H.T., Liu, A.Y.C. and Borst, P. (1984) Structure of the growing telomeres of trypanosomes. Cell 36, 459–468.
[32] Dover, G. (1982) Molecular drive: a cohesive mode of species evolution. Nature 299, 111–117.
[33] Arnheim, N. (1983) Concerted evolution of multigene families, pp. 38–61 in *Evolution of Genes and Proteins*, edited by M. Nei and R.K. Koehn. Sinauer Associates Inc., Sunderland, Massachusetts.
[34] Roeder, G.S. and Stewart, S.E. (1988) Mitotic recombination in yeast. Trends Genet. 4, 263–267.
[35] Maizels, N., 1989 Might gene conversion be the mechanism of somatic hypermutation of mammalian immunoglobulin genes? Trends Genet. 5, 4–8.
[36] Roth, C., C. Jacquemot, C., Giroud, C., Bringaud, F., Eisen, H. and Baltz, T. (1991) Antigenic variation in *Trypanosoma equiperdum*. Res. Microbiol. 142, 725–730.
[37] Van der Ploeg, L.H.T., Gottesdiener, K. and Lee, M. G-S. (1992) Antigenic variation in African trypanosomes. Trends Genet. 8, 452–456.
[38] Thompson, C.B. (1992) Creation of immunoglobulin diversity by intra chromosomal gene conversion. Trends Genet. 8, 416–422.
[39] Tibayrenc, M., Ward, P., Moya, A. and Ayala, A.F. (1986) Natural populations of *Trypanosoma cruzi*, the agent of Chagas' disease, have a complex multiclonal structure. Proc. Natl. Acad. Sci. USA 83, 115–119.
[40] Tibayrenc, M. and Ayala, F.J. (1988) Isozyme variability in *Trypanosoma cruzi*, the agent of Chagas' disease. Evolution 42, 277–292.
[41] Reynaud, C.A., Dahan, A., Anques, V. and Weill, J.C. (1989) Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. Cell 59, 171–183.
[42] McCormach, W.T., Tjoelker, L.W. and Thompson, C.B. (1991) Avian B-cell development; generation of an immunoglobulin repertoire by gene conversion. Annu. Rev. Immunol. 9, 219–241.
[43] Longacre, S. and Eisen, H. (1986) Expression of whole and hybrid genes in Trypanosoma equiperdum antigenic variation. EMBO J. 5, 1057–1063.
[44] Roth, C., Bringaud, F., Layden R.E., Baltz, T. and Eisen, H. (1989) Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes. Proc. Natl. Acad. Sci. USA 86, 9375–9379.
[45] Pays, E. (1989) Pseudogenes, chimaeric genes and the timing of antigen variation in African trypanosomes. Trends Genet. 5, 389–391.