Title

Isolation and characterization of the TSA/trans-sialidase family gene from the Silvio strain of Trypanosoma cruzi

Journal

Molecular and Biochemical Parasitology, 62(1)

ISSN

0166-6851

Authors

Ruef, Barbara J
Manning, Jerry E

Publication Date

1993-11-01

DOI

10.1016/0166-6851(93)90184-y

Molecular and Biochemical Parasitology, 62 (1993) 117–120

MOLBIO 2064

Short Communication

Isolation and characterization of the TSA/trans-sialidase family gene from the Silvio strain of *Trypanosoma cruzi*

Barbara J. Ruef** and Jerry E. Manning*

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

(Received 27 April 1993; accepted 29 July 1993)

Key words: *Trypanosoma cruzi*; Trypomastigote surface antigen

The major surface glycoproteins of the bloodstream trypomastigote stage of Trypanosoma cruzi are encoded by a large multigene family [1–10], several members of which are expressed simultaneously and encode glycoproteins of 85–160 kDa. We have previously isolated, characterized, and obtained the complete nucleotide sequence of TSA-1, a surface antigen gene from the Peru strain of *T. cruzi* belonging to the 85-kDa multigene family [1,2,5]. TSA-1 contains a 27-bp tandem repeat unit in the coding region which defines a family of 4 members in the Peru strain and one member in the Silvio X10 strain [2]. TSA-1 is the major expressed gene of the 4 Peru subfamily members and is found at a telomeric location on the chromosome [2] with the other members being found at a more internal location(s). A similar study with the Silvio strain showed that the solitary family member was insensitive to Bal 31 nuclease suggesting a chromosomal location more internal than that of TSA-1. A second difference between the TSA-1 families in these two strains is that the TSA-1 gene in Peru encodes a 3.7-kb A^+ RNA while the TSA-S1 gene in Silvio encodes a 3.1-kb A^+ RNA. The question arises as to whether the difference in length of these RNAs is due to altered processing of their primary transcript or whether the genes which encode these RNAs are structurally different. While genes which encode RNAs found in the 3.7-kb size class have been cloned and characterized, members of the gene family which encode RNAs in the 3.1-kb class have not been examined. Therefore, we have isolated and characterized the Silvio family member to determine the basis for the difference between these two size classes of RNA.

Since our previous studies showed that the 27 nucleotide repeat sequence in Silvio genomic DNA was present on a 13-kb *Eco*RI genomic DNA fragment, a genomic library was constructed in the Bluescript plasmid vector using purified *Eco*RI fragments of size 10–20 kb. The library was screened with a radiolabeled oligonucleotide representing the 27-nt repeat unit. Two bacterial recombinants rescreened positive. *Eco*RI digests of the recombinant plasmid DNA isolated from the

*Corresponding author.

**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 accession number L13844.

Abbreviations: aa, amino acid; nt, nucleotide; TSA-1, Peru trypomastigote surface antigen; TSA-S1, Silvio trypomastigote surface antigen.

two positives revealed a 13-kb insert in each. Restriction enzyme analysis showed the *Eco*RI fragments to have identical restriction patterns, again suggesting the presence of one family member in Silvio.

In order to isolate a cDNA fragment expressed from this family member, a cDNA library was constructed in phage λgt10 and screened with the 27-nt repeat unit oligonucleotide. Thirteen positive phage were identified. DNA was isolated from each recombinant phage, digested with *Eco*RI, and the size of *T. cruzi* inserts determined by electrophoresis on agarose gels. The largest insert, a 2.5-kb *Eco*RI fragment in λ phage TSA-S1, was excised and subcloned into the Bluescript vector for further analysis. The restriction enzyme map of the 2.5-kb insert matched that of the genomic *Eco*RI fragment, confirming that the TSA-S1 protein is encoded by the 13-kb genomic *Eco*RI fragment.

The complete nucleotide sequence of the 2.5-kb insert in λTSA-S1 was determined as well as approximately 2.0 kb of genomic DNA upstream of the 5′ terminus of the cDNA. The 3′ end of the insert contains a stretch of 17 A residues which we assign as the poly(A) tail since they do not occur in the genomic DNA sequence. Translation of the TSA-S1 gene sequence in each of the 6 possible reading

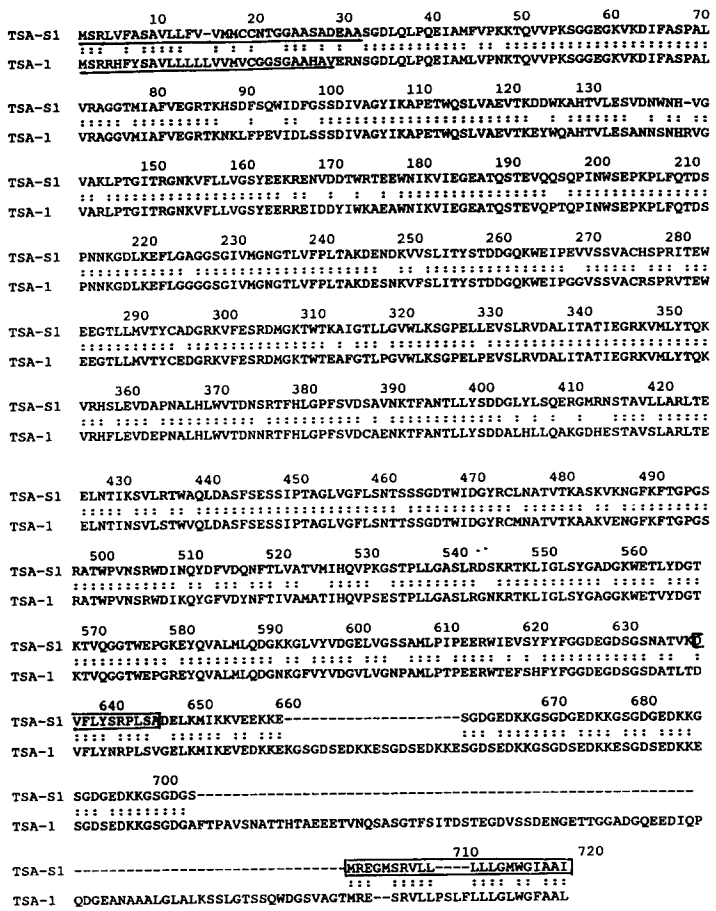

Fig. 1. Amino acid sequence alignment of TSA-S1 and TSA-1. The numbers above the sequence represent the aa residue number in TSA-S1. A colon between the two aligned sequences signifies identity, while a blank space signifies a lack of perfect aa identity. The putative signal peptides are underlined in each sequence. The boxed regions in the TSA-S1 sequence represent the two conserved regions described in the text.

frames revealed only one large open reading frame (ORF). The translational start site for this ORF is the first ATG codon in the reading frame whose immediate upstream sequence meets the requirements of the Kozak consensus sequence [11]. The translational stop site for this ORF is at nucleotide 2170 yielding a protein sequence of 723 amino acids.

The predicted amino acid sequence of the TSA-S1 gene was directly compared to that of TSA-1. The alignment of the two sequences (Fig. 1) shows a high degree of similarity (i.e. 83%) through the 9-aa repeat array. However, the aa sequences downstream of the repeat units in TSA-1 are absent in TSA-S1 with the exception of the 21 amino acid residues immediately upstream of the translational stop site. The COOH terminus of the coding region represents one of the two regions conserved in all 85-kDa multigene family members characterized to date [2–8] and is the proposed processing site for a phosphatidylinositol linkage [12]. The other conserved region found in all family members examined is a 12-aa region which lies immediately upsteam of the repeat array in TSA-S1 [4,10]. Interestingly, the predicted signal sequence at the amino terminus of TSA-1 shows only 35% identity with the predicted amino terminus of TSA-S1 (Fig. 1). Nevertheless, the amino terminus of TSA-S1 has all of the characteristics of a signal sequence [13,14] indicating that the TSA-S1 protein likely is found on the surface of the parasite.

A comparison of the nucleotide sequence of the TSA-S1 and TSA-1 genes show a pattern of identities in keeping with that observed in the alignment of the predicted aa residues (data not shown). TSA-S1 differs from TSA-1 by 336 nucleotides in the coding region immediately 3′ of the repeat array and 399 nucleotides in the 3′ untranslated region. Therefore, the 3′ end regions in TSA-S1 and TSA-1 differ in both length and sequence. Probes containing sequences from the 3′ end of either TSA-1 or TSA-S1, but not common to both, hybridized to only one of the two size classes of mRNA. The TSA-1 sequences hybridized only to the 3.7-kb size class and the TSA-S1 sequences hybridized only to the 3.1-kb size class (data not shown). It appears, therefore, that these two RNA size classes result from differences in the structure of the genes from which they are transcribed rather than from differential processing.

Acknowledgements

This work was supported by Public Health Service grant AI18873 from the National Institutes of Health. We thank Barbara Granger for assistance in maintenance and growth of parasite cultures.

References

1 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.
2 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.
3 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.
4 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.
5 Fouts, D.L., Ruef, B.J., Ridley, P.T., Wrightsman, R.A., Peterson, D.S. and Manning, J.E. (1991) Nucleotide sequence and transcription of a trypomastigote surface antigen gene of *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 46, 189–200.
6 Kahn, S., Colbert, T.G., Wallace, J.C., Hoagland, N.A., and Eisen, H. (1991) The major 85-kDa surface antigen of the mammalian-stage forms of *Trypanosoma cruzi* is a family of sialidases. Proc. Natl. Acad. Sci. USA 88, 4481–4485.
7 Takle, G.B. and Cross, G.A.M. (1991) An 85-kilodalton surface antigen gene family of *Trypanosoma cruzi* encodes polypeptides homologous to bacterial neuraminidases. Mol. Biochem. Parasitol. 48, 185–198.
8 Pereira, M.E., Mejia, J.S., Ortega-Barria, E., Matzilevich, D. and Prioli, R.P. (1991) The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neuraminidase, YWTD repeats of the low density lipoprotein receptor, and type III modules of fibronectin. J. Exp. Med. 174, 179–191.
9 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.
10 Takle, G.B., O'Connor, J., Young, A.J. and Cross, G.A.M. (1992) Sequence homology and absence of mRNA defines a pseudogene member of the *Trypanosoma cruzi* gp85/sialidase multigene family. Mol. Biochem. Parasitol. Mol. Biochem. Parasitol. 56, 117–128.
11 Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. 12, 857–872.
12 Ferguson, M.A.J. and Williams, A.F. (1988) Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. Annu. Rev. Biochem. 57, 285–320.
13 von Heijne, G. (1985) Signal sequences: the limits of variation. J. Mol. Biol. 184, 99–105.
14 von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. 14, 4683–4690.