Volume 9 Number 18 1981 Nucleic Acids Research

Variant surface glycoproteins of Trypanosoma brucei are synthesised with cleavable hydrophobic sequences at the carboxy and amino termini

John C.Boothroyd*, Cora A.Paynter*, George A.M.Cross*, Andre Bernards+ and Piet Borst+

*Dep. of Immunochemistry, Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, UK, and +Universiteit van Amsterdam, Afdeling Medische Enzymologie, Lab. voor Biochemie, Jan Swammerdam Institut, P.O. Box 60,000 1005 GA, Amsterdam, The Netherlands

Received 25 July 1981

ABSTRACT

cDNAs coding for the amino and carboxy termini of two trypanosome variant surface glycoproteins (VSGs) have been sequenced. The results indicate that VSGs are synthesised with hydrophobic amino-terminal leader and carboxy-terminal tail sequences which are absent from purified mature VSGs.

INTRODUCTION

The surface coat of the unicellular parasitic protozoan *Trypanosoma brucei* is composed of a dense, essentially homogeneous layer of variant surface glycoprotein (VSG) (1-3). During the course of an infection, a small fraction of the trypanosome population can switch to synthesis of an alternative, antigenically distinct VSG, thereby evading the immune response of the host. Previous studies have shown extensive diversity in the amino acid sequence of different VSGs (4-6) and that each VSG has its own gene (7,8). Activation of a VSG gene is accompanied by duplication of the gene and transposition of the duplicate to elsewhere in the DNA (8,9). To further our understanding of the mechanism of antigenic variation we have undertaken a detailed analysis of the structure and organisation of several VSGs and their genes.

We have previously reported that nucleotide sequence analysis of a cDNA copy of the mRNA for one VSG (type 117) predicts that the protein is initially synthesised with a hydrophobic carboxy-terminal tail which is absent from the purified glycoprotein (10). As an extension to these studies we have analysed additional double-stranded (ds) cDNA containing-recombinants and report here that a similar carboxy-terminal tail is predicted for a second VSG (type 221) and that the genes for both these VSGs also code for amino-terminal hydrophobic leader or signal sequences typical of membrane and secreted proteins.

MATERIALS AND METHODS

The construction and selection of recombinants containing ds cDNA inserts corresponding to the 117 and 221 VSG mRNAs have been previously described (7). Briefly, ds cDNA was synthesised using poly A^+ mRNA from four variants (including 117 and 221) of a single clone of *T.brucei* and inserted by G:C tailing into the Pst I site of pBR322. As the only known difference between two variants arising from a clone is in the VSG on their surface, recombinants which hybridised to homologous cDNA only were presumed to contain sequences corresponding to VSG mRNA. Subsequent hybrid selected translation and DNA sequence analysis has confirmed this.

On the basis of their size, relative overlap and hybridisation to short cDNA probes, recombinants containing ds cDNA inserts which extended furthest towards the 5' or 3' ends of the mRNA were chosen for further analysis. Detailed maps of restriction endonuclease cutting sites were constructed for each of these recombinants, and nucleotide sequences of the regions which code for the amino and carboxy termini of the VSGs were determined according to the method of Maxam and Gilbert (11). In addition to the four standard reactions (i.e., G, A+G, C+T and T-specific)a fifth reaction (A>C) was also carried out (11). Together with the practice of fixing and drying gels for autoradiography (10), this gave an unambiguous sequence for a particular fragment without having to sequence both strands.

RESULTS

1. Amino-terminus of VSG 117

The restriction map and sequenced region of the recombinant plasmid TcV117·5 are presented in fig.1. The nucleotide and deduced amino acid sequences for this region are shown in fig.2. Previously determined amino acid sequence data for purified VSG 117 (Variant 2 in ref.5; G.Allen, personal communication) unambiguously identified the mature amino-terminus as being at position 185. Upstream of this residue, two in-phase Met codons are observed (positions 86 and 140). The presence of a stop codon in phase and further upstream indicates that one of these Met codons must be the initiation site for translation of the VSG 117 mRNA. The leader peptide sequence resulting from translation initiating at either of these sites would include a hydrophobic region. We have no data on which initiation site is used. It has been proposed, however, that ribosomes bind to the 5' end of eukaryotic mRNAs and migrate towards the 3' end, initiating translation at the first unprotected (i.e., not involved in intra-strand base-pairing) AUG

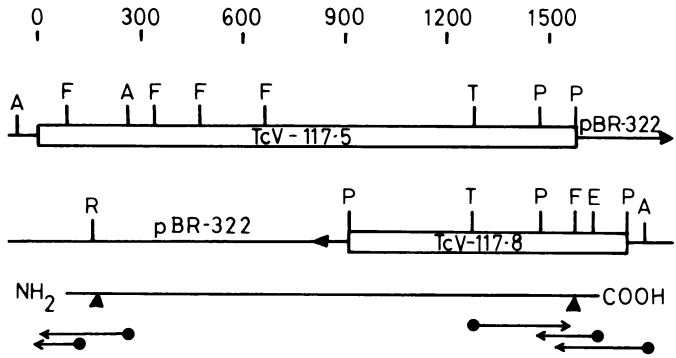

Fig.1. Physical maps of TcV117·5 and TcV117·8 showing restriction endonuclease cutting sites and regions sequenced. The two inserts were aligned by comparison of their restriction maps. Their orientation relative to the EcoRI site of pBR322 is indicated by (▶) at the end of the insert nearest to this site. Distances are in base pairs from the 5' end of the TcV117·5 insert. These data were obtained by standard mapping procedures. Beneath the maps are shown the regions of the inserts which code for the putative primary translation product. The carboxy- and amino-termini of the mature VSG are also indicated (▲). Below this are shown the sites which were 5'-^{32}P end labelled (●) and the regions where useful sequence information was obtained (→). Restriction sites are coded as follows: P, Pst I; R, EcoRI; E, Hae III; A, Hpa II; H, Hind III; T, Taq I; F, Hinf I.

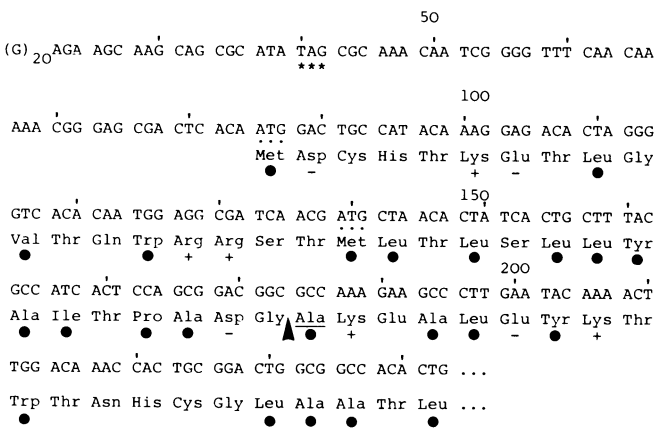

Fig.2. Summary of nucleotide and protein sequence data corresponding to the amino-terminus of VSG 117. The nucleotide sequence of the coding strand between positions 1 and 240 (see fig.1) is given together with the implied amino acid sequence. (G)20, synthetic oligo (dG)-tail used to anneal the plasmid to the dC-tailed insert: ***, in phase stop codon: ···, possible start sites for translation: +, basic residue: -, acidic residue: ●, hydrophobic residue: ▲Ala, amino-terminal residue of mature isolated glycoprotein.

codon which is encountered (12). This would suggest that the upstream AUG codon (position 86) is the initiation site for translation of the VSG 117 mRNA, since computer analysis (13) of this region does not reveal any sequences which might form stable secondary structures. This suggestion is further supported by the fact that only if translation was initiated at position 86 would there be a basic residue (Arg) preceding the hydrophobic portion of the leader, a property of many such sequences (14). It is interesting to note that if translation does start here then the resulting leader peptide would be one of the longest so far reported.

2. Amino-terminus of VSG 221

Figure 3 presents the restriction map and region of the recombinant plasmid TcV221·5 which was sequenced. Figure 4 shows the nucleotide and deduced amino acid sequences which were obtained. It was not necessary to sequence through the Hind III site at position 141-146 because this lies in the middle of a region for which the amino acid sequence of the mature VSG 221 is known (6): the predicted and known sequences agree only if this site is unique. As with VSG 117, the mature amino-terminus was identified by comparison with the known amino acid sequence of VSG 221 (6). The nucleotide sequence predicts that the primary translation product of the VSG 221 gene also would be synthesised with a basic residue (Arg) preceding the hydrophobic region of the leader sequence. Only one in-phase AUG codon is observed upstream of the mature amino-terminus, but as it is not preceded by an in-phase stop codon, translation may in fact initiate further upstream at a site

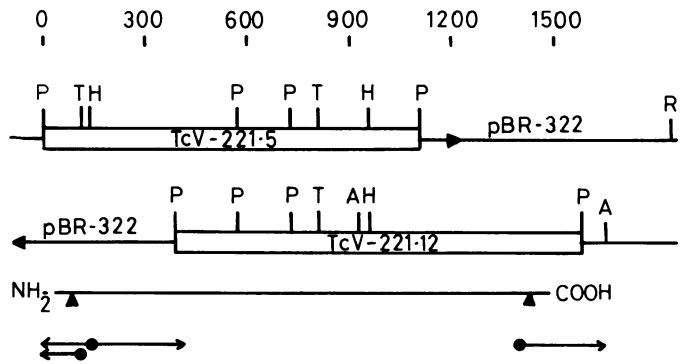

Fig.3. Physical map of TcV221·5 and TcV221·12 showing restriction endonuclease cutting sites and regions sequenced. Notation is as in fig.1 except that distances are in base pairs from the 5' end of the TcV221·5 insert.

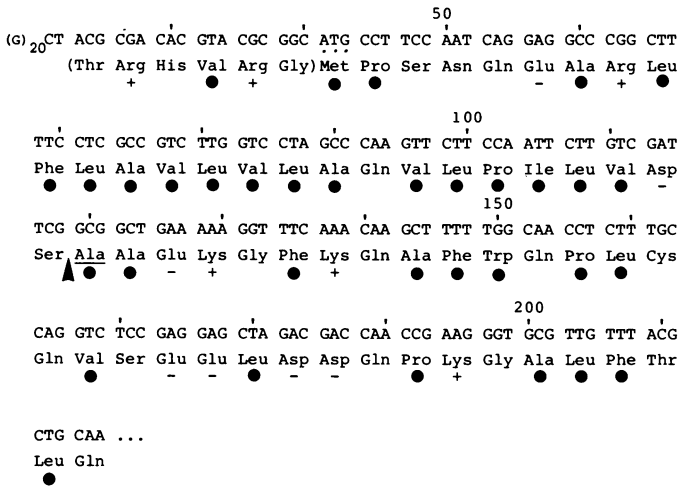

Fig.4. Summary of nucleotide and protein sequence data corresponding to the amino-terminus of VSG 221. The nucleotide sequence of the coding strand between positions 1 and 217 (see fig.3) of TcV221·5 is given together with the implied amino acid sequence. Notation as in fig.2.

not represented in the ds cDNA inserts of this recombinant.

3. Carboxy-termini of VSG 117 and 221

We have previously reported and discussed the results for the carboxy-terminus of VSG 117 (10). Figure 3 presents the restriction map and region sequenced for TcV221·12. The sequence data which were obtained are compared with TcV117·8 in fig.5. The mature carboxy-terminus of VSG 221 was identified by comparison with the amino acid sequence data published elsewhere (15). As with VSG 117, the data indicate that the primary translation product of the VSG 221 gene is synthesised with a hydrophobic carboxy-terminal tail which is absent from the purified glycoprotein. The primary structure of this tail is different in the two variants shown except for the last three residues which are identical. This reflects the high degree of homology between the two nucleotide sequences from position 1633 (of the 117 variant) to the end of the ds cDNA inserts. Neither TcV117·8 nor TcV221·12 contain sequences corresponding to the poly (A) tail or the putative polyadenylation signal AAUAA(A) found in mRNAs from vertebrates (16,17). However from nucleotide sequence analysis of additional recombinants (18), we know that the poly (A) tail starts about 7 nucleotides after the end of the 117 and 221 sequences presented in fig.5.

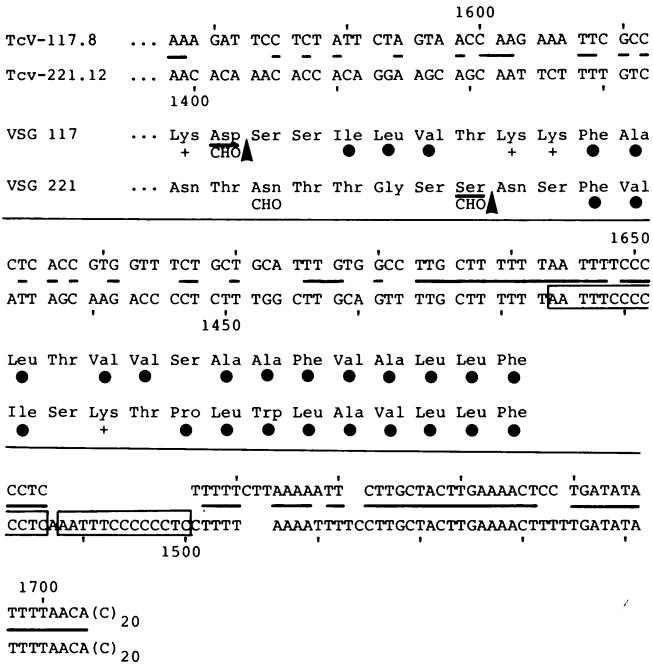

Fig.5. Comparison of nucleotide and protein sequence data corresponding to the carboxy-termini of VSG 117 and VSG 221. The nucleotide sequence of the coding strand between positions 1567 and 1704 of TcV117·8, and 1398 and 1549 of TcV221·12 are given together with the implied amino acid sequences. The two sets of sequences are aligned (with gaps, where necessary) to maximise homologies (bold underlining). Notation is as in fig.2 except the carboxy-terminal residue of the mature isolated glycoprotein is underlined and followed by (▲). CHO beneath a residue indicates attachment of a carbohydrate moiety to the mature VSG at this position. The 13bp boxed region is a direct, near-tandem repeat.

DISCUSSION

The finding that both VSG 117 and 221 are apparently synthesised with hydrophobic amino-terminal leader sequences is consistent with their status as cell surface proteins since leader sequences are thought to be the signals for transport of such proteins through the membrane (14). Apart from their hydrophobicity, there is no significant amino acid or nucleotide sequence homology between the two leaders. The predicted length of the 117 and 221 leader sequences is of the same order as that recently determined by McConnell et al for a different VSG of *T.brucei* (19). These workers have shown by tryptic peptide analyses that VSG MIT at 1.6 is synthesised *in vivo*

with a leader peptide of 30-40 amino acids. This result is further, indirect support for the upstream AUG codon being the initiation site for translation of the 117 mRNA *in vivo*.

The hydrophobic carboxy-terminal tail with which these two VSGs are initially synthesised may be acting as a "stop-transfer" signal preventing the release of the VSGs from the membranes on which they are synthesised. Such a function has been ascribed to similar sequences observed in other systems including the membrane-bound form of the μ chain of IgM (20) and the haemagglutinin of influenza virus (21). If this is the function of the carboxy-terminal tails, then the many VSGs of *T.brucei* could provide a useful model for studying the properties which are required for the stop-transfer signal. At present, a comparison of the limited number of primary structures known for integral membrane proteins suggests that the requirements for this function will be as difficult to decipher as those for the amino-terminal leader sequences.

The VSGs of *T.brucei* differ from the two examples cited above, however, in that the tails are absent from the purified protein. We do not know at what stage this cleavage takes place but it may be that the tails are essential only for the initial routing of the VSGs and that most or even all of the VSG molecules lose their tails just prior to or soon after arriving at the cell surface: other interactions could then be responsible for their retention on the surface. Interestingly, all six VSGs which have been studied (14) terminate in the mature form with a glycosylated residue at the C-terminus suggesting that the specificity of the putative proteolytic activity which removes the tail and the position of this carbohydrate moiety are in some way related.

The identification of aspartic acid (rather than the expected asparagine) as the glycosylated residue at the mature C-terminus of VSG 117 (fig.5) has been confirmed by nucleotide sequence analysis of a genomic copy of the gene (J.C.B., A.B. and S.L.Coleman, unpublished results). Glycosylated aspartate has also recently been shown to be the terminal residue of another mature VSG (type 121; J.C.B., A.B., and S.L.Coleman, unpublished results). This was anticipated from the similarity of the previously determined C-terminal amino acid sequences of VSG 117 and 121 (14).

Under standard hybridisation conditions, the cDNAs analysed here each recognise a family of related sequences in trypanosome nuclear DNA. The 117 and 221 families do not overlap and 117 and 221 cDNAs do not hybridise with each other under standard conditions (9,22; A.B., unpublished observations).

It is therefore rather unexpected to find such a striking similarity in the nucleotide sequences of TcV221·12 and TcV117·8 beyond position 1633. This homology may be a result of common functional demands relating to processing of the mRNA or protein, and/or a common origin of these sequences. Preliminary results indicate that an active VSG gene may be produced by juxtaposing one DNA segment which contains all the region coding for the amino terminal leader, the mature VSG and about half the carboxy-terminal tail with another DNA segment coding for the distal half of the carboxy-terminal tail and the 3' untranslated region. From the comparison presented in fig.5, it would appear that this latter segment is homologous but not identical in the VSG 117 and 221 genes. Experiments now in progress on genomic copies of these genes in their silent and expressed state should help in explaining the significance of this homology.

Acknowledgements

We wish to thank Mrs.K.Draper for careful typing of this paper, Miss.J. Van den Burg for technical assistance, J.H.J.Hoeijmakers and C.Weissmann for supplying the recombinant plasmids, and G.Allen, S.L.Coleman, A.A.Holder, H. Majumder, J.McConnell, M.Turner, H.Weber and C.Weissmann for communicating results prior to publication. This work was in part supported by a grant to P.B. from the Foundation for Fundamental Biological Research (BION) which is subsidised by the Netherlands Organisation for the Advancement of Pure Research (ZWO).

REFERENCES

1. Vickerman, K. (1978) Nature 273, 613-617.
2. Cross, G.A.M. (1978) Proc.R.Soc.B. 202, 55-72.
3. Cross, G.A.M., Holder, A.A., Allen, G. and Boothroyd, J.C. (1980) Amer.J. Trop.Med. 29 suppl., 1027-1032.
4. Cross, G.A.M. (1975) Parasitology 71, 393-417.
5. Bridgen, P.J., Cross, G.A.M. and Bridgen, J. (1976) Nature 263, 613-614.
6. Johnson, J.G. and Cross, G.A.M. (1979) Biochem.J. 178, 689-697.
7. Hoeijmakers, J.H.J., Borst, P., Van den Burg, J., Weissmann, C. and Cross, G.A.M. (1980a) Gene 8, 391-417.
8. Hoeijmakers, J.H.J., Frasch, A.C.C., Bernards, A., Borst, P. and Cross, G.A.M. (1980b) Nature 284, 78-80.
9. Frasch, A.C.C., Bernards, A., Van der Ploeg, L.H.T., Borst, P., Hoeijmakers, J.H.J., Van den Burg, J. and Cross, G.A.M. (1980) in The Biochemistry of Parasites and Host-Parasite Relationships : The Host-Invader Interplay, Van den Bossche, H. Ed., pp.235-239 Elsevier/North Holland Biomedical Press, Amsterdam.
10. Boothroyd, J.C., Cross, G.A.M., Hoeijmakers, J.H.J. and Borst,P. (1980) Nature 288, 624-626.
11. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldare, K. Eds., vol.65 pp.499-560, Academic Press.

12. Kozak, M. (1980) Cell 22, 7-8.
13. Staden, R. (1978) Nucleic Acids Res. 5, 1013-1015.
14. Davis, B.D. and Tai, P-C. (1980) Nature 283, 433-438.
15. Holder, A.A. and Cross, G.A.M. (1981) Molec.Biochem.Parasitol. 2, 135-150.
16. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
17. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142.
18. Majumder, H., Boothroyd, J.C. and Weber, H. (1981) Nucleic Acids Res. submitted.
19. McConnell, J., Gurnett, A.M., Cordingley, J.S., Walker, J.E. and Turner, M.J. (1981) Molec.Biochem.Parasitol., submitted.
20. Rodgers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L. and Wall, R. (1980) Cell 20, 303-312.
21. Verhoeyen, M., Fang, R., Min Jou, W., Devos, R., Huylebroek, D., Saman, E. and Fiers, W. (1980) Nature 286, 771-776.
22. Borst, P., Frasch, A.C.C., Bernards, A., Van der Ploeg, L.H.T., Hoeijmakers, J.H.J., Arnberg, A.C. and Cross, G.A.M. (1981) Cold Spring Harbor Symp. Quant.Biol. 45, in press.