Volume 10 Number 19 1982 Nucleic Acids Research

An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome

L.H.T.Van Der Ploeg, D.Valerio*, T.De Lange, A.Bernards, P.Borst and F.G.Grosveld[1]

Section for Medical Enzymology and Molecular Biology, Laboratory of Biochemistry, University of Amsterdam, Jan Swammerdam Institute, P.O. Box 60.000, 1005 GA Amsterdam, The Netherlands, and [1]Laboratory of Gene Structure and Expression, NIMR, Mill Hill, London NW7 1AA, UK

Received 12 July 1982; Accepted 14 August 1982

ABSTRACT

Trypanosoma brucei contains more than a hundred genes coding for the different variant surface glycoproteins (VSGs). Activation of some of these genes involves the duplication of the gene (the basic copy or BC) and transposition of the duplicate to an expression site (yielding the expression-linked copy or ELC). We have cloned large fragments of genomic DNA in cosmid vectors in Escherichia coli. Cosmids containing the BCs of genes 117, 118 and 121 were readily obtained, but DNA containing the ELCs was strongly selected against in the cosmid and plasmid cloning systems used. We have analysed the distribution of VSG genes in the genome using probes for the sequences at the edges of the transposed segment which are partially homologous among these genes. In genomic cosmid clone banks, about 9% of all colonies hybridize with probes from the 5'- and 3'-edges of the transposed segment, showing that these sequences are linked in the genome. Moreover, the 117 and 118 BC cosmids contain several additional putative VSG genes in tandem, as deduced from hybridization and sequence analyses. We conclude that the VSG genes are highly clustered and share common sequences at the borders of the transposed segment.

INTRODUCTION

A large repertoire of genes coding for the VSGs exists in the genome of T. brucei [1-3]. Through the sequential expression of different VSG genes the trypanosome evades the host immune response and hence survives in the mammalian bloodstream [4-6]. The activation of some VSG genes involves the duplication of a silent or BC gene and the transposition of the duplicate to an expression site elsewhere in the genome [1,3]. The extra VSG gene copy or ELC is preferentially digested by DNase I in isolated trypanosome nuclei [7] and the one used for messenger RNA (mRNA) synthesis [8].

Our studies on the duplication and transposition event have mainly dealt with the expression of the VSG genes 117 and 118 of T. brucei stock 427. We have shown that these genes are

0305-1048/82/1019-5905$ 2.00/0

transposed to similar or identical expression sites [9]. The transposed segment starts 1-2 kb in front of the gene [9] and ends within the 3'-end of the gene [8]. The VSG gene transcripts most likely originate from a promoter added onto the transposed segment in the expression site. Transcription thus yields large precursor RNAs which are spliced to yield the mature VSG mRNA [10].

Many unsolved problems in antigenic variation, like the mode of evolution of the large repertoire of VSG genes and the basis for ordered expression, require knowledge of the total size of the VSG gene repertoire and the distribution of these genes through the genome. Since heterogeneity is the hallmark of VSG genes, a definitive picture can only be drawn when a vast number of VSG genes has been cloned and checked for function. Our analysis of a limited number of VSG genes has suggested, however, three methods to tentatively identify VSG genes that have not yet been cloned:

1. We have shown that VSG gene probes recognize families of related VSG genes in nuclear blots. Homology between members of a family increases from 5' to 3'. A single VSG gene probe can, therefore, be used to detect multiple VSG genes [11].

2. The 3'-edge of the transposed segment of VSG genes contains a sequence that is found with minor variations in all VSG genes analysed thusfar [6 and Liu, A.Y.C. and Michels, P.A.M., unpublished].

3. The 5'-edge of the transposed segment contains another repetitive sequence (denoted A) which is also present before other VSG genes analysed [9].

It seems likely that sequence A and the 3' conserved sequences of the transposed segments are required for the duplication-transposition process that activates VSG genes and that they can be used to mark their presence in the genome. We have used these sequences to study the number and distribution of VSG genes in the trypanosome genome. We find that there are 10^3 potential VSG genes and that they are highly clustered.

MATERIALS AND METHODS

Isolation of trypanosome nuclear DNA: *T. brucei* stock 427

was grown in rats and purified free from blood elements as described [12]. Trypanosomal DNA was prepared from T. brucei strains MITat 1.4 (117a), 1.5 (118a) and 1.6 (121) as described in ref. 9.

Isolation of plasmid and cosmid DNA: The isolation method of Birnboim and Doly [13] was used.

Restriction endonuclease digestion, electrophoresis and transfer of DNA to nitrocellulose filters: Incubation conditions for restriction endonucleases were as specified previously [14]. DNA digests were size-fractionated by electrophoresis through horizontal agarose slab gels as described [15]. After electrophoresis the DNA was transferred to nitrocellulose filters by the acid blotting procedure as described in refs 16 and 17.

Two-dimensional blotting hybridization: The procedure followed was that of Sato et al. [18] with the minor modifications introduced by J.Groffen (personal communication). In short, cosmid DNA was digested with restriction enzymes, size fractionated in a 0.7% agarose gel and transferred to a nitrocellulose filter. A second sample of digested cosmid DNA was labelled by nick-translation, size fractionated in a 0.7% agarose gel and pre-treated for hybridization. The filter was laid on the gel with the slots perpendicular to each other. After the blotting hybridization step, the filter was washed three times at 65°C in 3 x SSC for 15 min, dried and exposed to Kodak XR-I film at -70°C or further washed three times for 15 min at 65°C in 1 x and 0.3 x SSC and then exposed at -70°C for 12-48 h.

Isolation of DNA probes and filter hybridization: Probes of cloned DNA were isolated by preparative agarose gel electrophoresis of the appropriate restriction endonuclease digests, followed by isolation of the specific DNA fragments from the agarose [20], which were then labelled by nick-translation [21]. Hybridization of filters at 65°C in 3 x SSC, 10% dextran sulphate and subsequent post-hybridizational washes at 3 x, 1 x, 0.3 x and 0.1 x SSC, respectively, at 65°C were done as described [10,16,22].

Construction of cosmid libraries: The cosmid cloning procedure used was mainly that described by Grosveld et al. [23]. In short, clone banks were constructed from the nuclear DNA isolated of variants 117a, 118a and 121. The DNA was partially digested with MboI and size-fractionated on sucrose gradients as described [23]. Fractions containing DNA with an average size of 30-80 kb were used in the cloning procedure. Two pJB8-derived cosmid vectors - pRTI and POPFI - were used in the cloning procedure ([24]; Lund, T. and Grosveld, F.G., unpublished). The pRTI vector DNA was digested with BamHI and treated with alkaline phosphatase (BAPF, Worthington) as described [23]. The POPFI vector was first digested with either ClaI or EcaI and then de-phosphorylated. The left and right vector arms of the ClaI or EcaI-digested vector samples, both containing the cos site, were then created by digestion with BamHI. The ligation was performed as described [23], with the modification that both arms of the POPFI vector were added in a molecular ratio vector:insert of 5:1. The ligated DNA was packaged and transducted on E. coli 1400 or ED 8767 [25] bacterial strains as described [23].

Bacteria were plated on ampicillin-containing agar plates at a density of 10,000 colonies per 14-cm diameter petridish. Millipore HATF filters (pore size, 0.45 µm) were used in all cases. Peplica plating of the recombinant bacteria was performed as described by Hanahan and Meselson [26] with the modifications described by Grosveld et al. [23]. Positive colonies were detected with the 5'-half gene probes of the appropriate cDNA clones: PstI-SalI (900 bp) from TcV 117-5 (probe N117 in Fig. 4), PstI-EcoRI (500 bp) from TcV 118-2 (probe N118), and the PstI fragment (850 bp) of the TcV 121-7 cDNA (probe N121 in Fig. 4).

The use of these fragments rather than the complete cDNA avoids additional hybridization due to sequences in the 3' part of the cDNA which are present in many copies in the genome. Colonies containing the VSG inserts were picked and further purified by repeating the screening procedures once. Analysis of purified cosmid DNA from the positive colonies was done by Southern blotting and subsequent hybridization of the clones with the appropriate cDNA probes.

Two experiments were done to verify that all colonies in the clone bank contain a trypanosome DNA insert. First, forty randomly-picked colonies were analysed and all contained a DNA insert of about 35 kb. Second, a mock transduction was carried out with ligated vector arms only. The number of colonies was 0.1% of the ligation with trypanosome DNA.

Identification and characterization of cosmid clones containing the 118 BC VSG gene: The complexity of the trypanosomal nuclear DNA is 35×10^6 bp [27]. The 118 cosmid clone bank contained 80,000 colonies. It should, therefore, be equal to about 80 times the size of the genome. In 60,000 colonies of this clone bank 33 positive recombinants were detected with the 118-2 PstI-EcoRI 5' cDNA sub-probe (probe N118 in Fig. 4). These colonies were isolated and analysed in detail. Fig. 1 shows the EcoRI restriction enzyme digestion pattern of 21 cosmid clones. The co-migrating fragments in the restriction enzyme pattern indicate that overlapping clones have been isolated. The presence of the 118 VSG gene was visualized using Southern blotting analysis at stringent hybridization conditions. Fig. 1 shows eight positive clones hybridizing with the 118-2 3' cDNA sub-probe. That the cosmid clones contain the 118 BC VSG gene and not a related gene was verified in protection experiments of 118 cosmid DNA by 118 mRNA to S_1 nuclease digestion [28] and comparison of the nuclear DNA physical maps with those of the cloned cosmid DNA.

The insert sizes of the 33 118 BC cosmid clones varied from approx. 30 to 40 kb. Fig. 2 shows the 118 BC physical map. The overlapping clones comprise an area of 60 kb of genomic DNA.

Cloning of the 117 BC VSG gene: The 117a nuclear DNA cosmid clone bank measured ten times the genome size. Three 117 BC cosmid clones were isolated. The identity of the clones was verified in S_1 nuclease analysis of the gene using the 117 cosmid clones and 117 mRNA, as well as by comparison with the previously isolated BC clone in lambda.gt.WES [8]. The physical maps of the overlapping cosmid clones are presented in Fig. 2.

Cloning of the 121 VSG gene: There are multiple isogenes of the 121 VSG gene in the genome (Bernards, A. and Van der

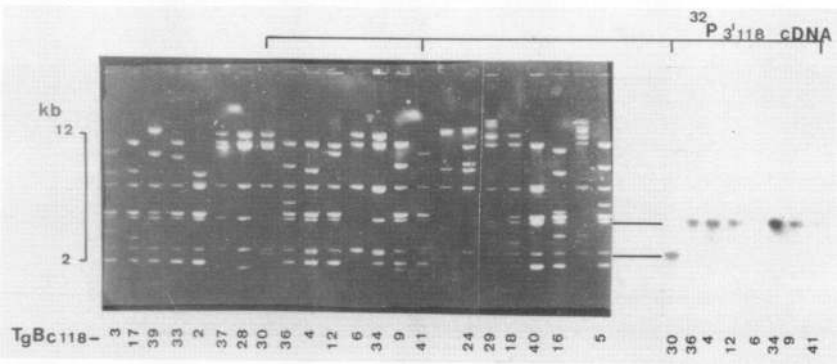

Fig. 1. Characterization of cosmid clones containing the 118 BC VSG gene. Cosmid DNA was isolated from colonies hybridizing with a 5' half-gene cDNA sub-probe from TcV 118-2 (N118; Fig. 4), digested with EcoRI and size-fractionated in a 0.7% agarose gel. The left-hand part of the figure shows the gel after staining with ethidium bromide. After transfer of the DNA to nitrocellulose filters, DNA fragments containing the 118 BC VSG gene (TgBc-118) were detected using a 3' half-gene cDNA sub-probe from TcV 118-2 (C118; Fig. 4) at 0.1 x SSC (65°C) stringency of hybridization. The right-hand panel shows the autoradiogram of part of the hybridized filter. Lanes without number contain cosmid DNA without 118 BC VSG gene.

Ploeg, L.H.T., unpublished). Restriction enzyme digests of the nuclear DNA from variants 118a and 121 suggest the presence of three 121 BC isogenes that are so closely related that they cannot be distinguished even at stringent hybridization conditions. A fourth gene, probably the 121 ELC, is present in 121 nuclear DNA and - in altered form - in 221 nuclear DNA.

From the 121 nuclear DNA cosmid clone bank, measuring 10 times the genome size, only one cosmid was isolated containing a 121 VSG gene (Fig. 2; TgBc 121-10). A second isogene of the 121 VSG genes was isolated from the 118 cosmid clone bank. Fig. 2 shows the physical maps of the isogenes. The 121 BC VSG genes on the cosmid clones were analysed in S_1 nuclease protection experiments with 121 mRNA for the presence of the 121 BC gene and co-migration of nuclear DNA and cosmid clone fragments for the absence of cloning artefacts. The analysis showed that indeed the 121 VSG gene is isolated and that all fragments in the cosmid clones are co-linear with the sequences in the genome (not shown).

Cloning of the 221 VSG gene: All attempts to obtain 221 genes from the cosmid clone bank have failed. Because the 221 VSG genes are located next to a DNA segment without MboI sites that resembles the end of a chromosome (Bernards, A. and De Lange, T., unpublished), MboI partial digestion must have selected against cloning of these genes. Another clone bank was, therefore, constructed with 118 nuclear DNA partially digested with HindIII, pBR322 as vector and E. coli HB101 [29] as host (see ref. 30). From this bank one plasmid was obtained with a 10-kb insert containing part of a 221 gene and the adjacent upstream genomic sequences (see Fig. 3).

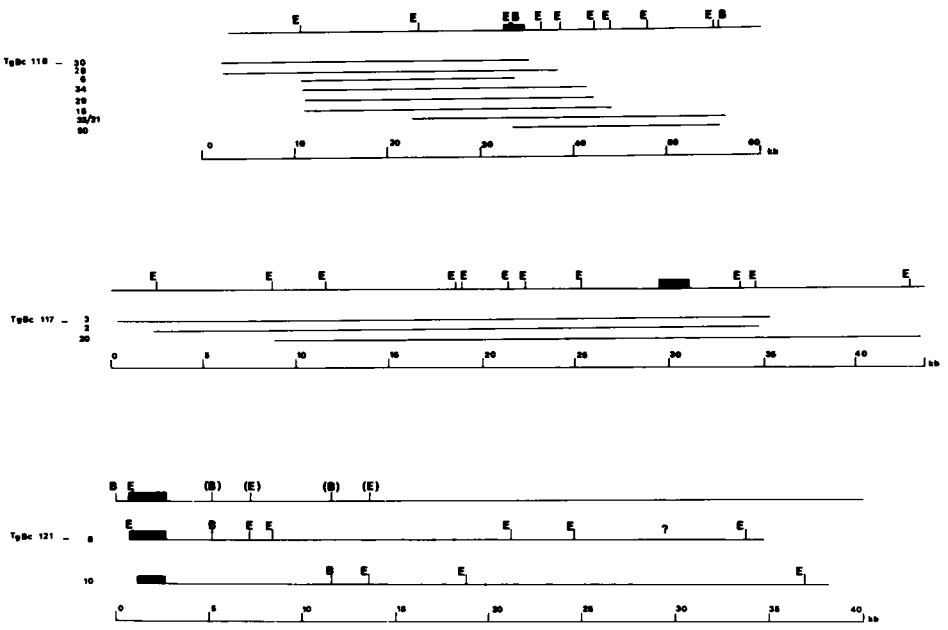

Fig. 2. Physical maps of the 117, 118 and 121 BC cosmid clones. The maps were prepared using the two-dimensional blotting hybridization technique [18] on EcoRI and PvuII-digested DNA of the VSG 118 BC clones (TgBc-118) and EcoRI and PstI digestions of the VSG 117 BC clones (TgBc-117). Maps of the 121 VSG cosmid clones (TgBc-121) were constructed using EcoRI and EcoRI+ BamHI double digests. The alignment of the overlapping BC clones is indicated below the maps. The black boxes correspond to the VSG gene coding sequence. The direction of transcription is from left to right. Abbreviations: E, EcoRI; B, BamHI.

RESULTS

Distribution of VSG genes containing recombinants in the cosmid clone banks

To determine whether VSG genes have the same chance of ending up in cosmid clones as an average stretch of trypanosome DNA, a large bank containing DNA from variant 118a was screened for the presence of the 118 BC and ELC genes. The 118 genes are most suitable for this purpose because the 118 BC gene belongs to a relatively small family of rather distantly-related genes (see ref. 6). Hence, a probe corresponding to the 5'-half of the gene will only detect a single gene in Southern blots of nuclear DNA from a variant that does not express the 118 gene, even at lowstringency hybridization conditions. This allows unambiguous

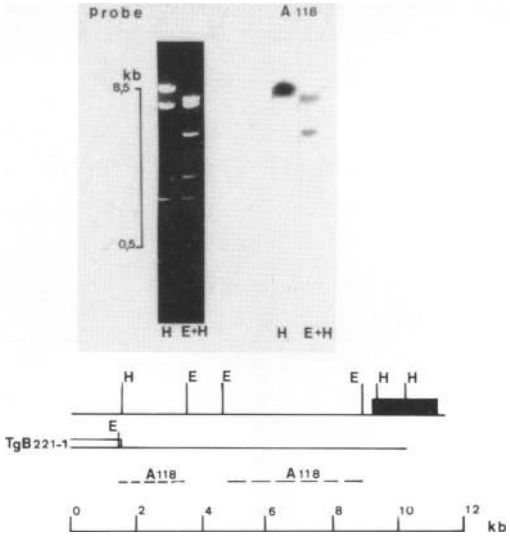

Fig. 3. Hybridization of the 5' flanking region of the 221 VSG gene with the 5' border probe. DNA of plasmid TgB-221-1 was digested with HindIII or EcoRI+ HindIII. The DNA was size-fractionated in a 0.7% agarose gel and transferred to nitrocellulose filters, which were hybridized with the 5' border probe of the VSG 118 gene transposed segment (A118) at 3 x SSC (65°C) stringency of hybridization. Below the physical map of the 221 VSG gene (black box), the genomic clone is shown (open box = vector). The dotted line indicates the location of the hybridizing fragments. Abbreviations: E, EcoRI; H, HindIII.

and quantitative identification of the gene in clone banks. Moreover, gene titrations indicate that only a single copy of the 118 BC is present per diploid nucleus [31]. Since the complexity of T. brucei DNA is known (35 x 10^3 kb [27]), the number of cosmids containing the 118 BC or ELC gene can be predicted if no bias occurs in the cloning procedure.

Our 118 cosmid clone bank contained 80,000 colonies with an average insert of circa 35 kb and, therefore, contained about 80 times the genome size. In 60,000 colonies 30 BC and 30 ELC gene cosmids would be expected; 33 were found, all of which turned out to contain the BC gene as discussed in Methods. This result shows that there is a strong selection against the ELC genes, but no bias towards cloning of the 118 BC gene. Since the 118 BC gene is apparently surrounded by other, distantly-related VSG genes (see below), it seems likely that there is no bias against BC genes in general in our cosmid cloning procedure.

Selection against cloning of the 117 and 118 ELCs

A 117 or 118 ELC-containing cosmid clone was not found in the clone bank. The presence of both the BC and ELC sequences in the nuclear DNA fractions from the sucrose gradient used for cloning was verified by Southern blotting an EcoRI digest of 117a or 118a DNA and hybridization with the respective cDNA probes. Both DNA preparations contained the known BC and ELC bands (not shown). The lack of ELC-containing cosmids could have been due to the lack of MboI sites in the vicinity of the ELC gene. The 117 and 118 VSG genes in the expression site are flanked at their 3'-side by a stretch of 8 kb DNA, not cut by any restriction enzymes including MboI [9]. This stretch ends in a discontinuity in the DNA, preferentially degraded by treatment with BAL31 and it is, therefore, thought to be the end of a chromosome [32]. Partial digestion with MboI will, therefore, strongly select against cloning of the ELC fragments, because one end of the MboI digestion product cannot be ligated into the cosmid vectors. It should be possible, however, to isolate cosmid clones extending upstream from a MboI site in the gene.

To enhance the probability for isolation of such cosmids, another approach was used. Nuclear DNA of variant 118a was digested to completion with BamHI, generating an ELC-containing fragment extending from the BamHI site in the 118 VSG gene towards the BamHI site 25 kb upstream, which has two sticky ends. A BamHI fragment of 25 kb is also obtained containing the 3'-end of the BC gene. These two 25-kb fragments were separated from the smaller BamHI 118 VSG gene digestion products on sucrose gradients as described [23]. The 25-kb fraction containing the BC and ELC BamHI fragments, as shown by Southern blotting and subsequent hybridization, was used to construct a cosmid clone bank in the BamHI-digested POPFI vector as described in Methods. After hybridization of the recombinant clone bank with the total 118-2 cDNA sub-probe (Fig. 4, probe N118, C118) eight cosmid clones were obtained, all containing the 25-kb 3' BC fragment (represented in Fig. 2 by cosmid TgBc 118-50). No ELC cosmid clone was obtained. We conclude that the 5' BamHI ELC fragment contains sequences which interfere with its propagation in the bacterial host (E. coli st 1400 or ED 8767). The likely presence

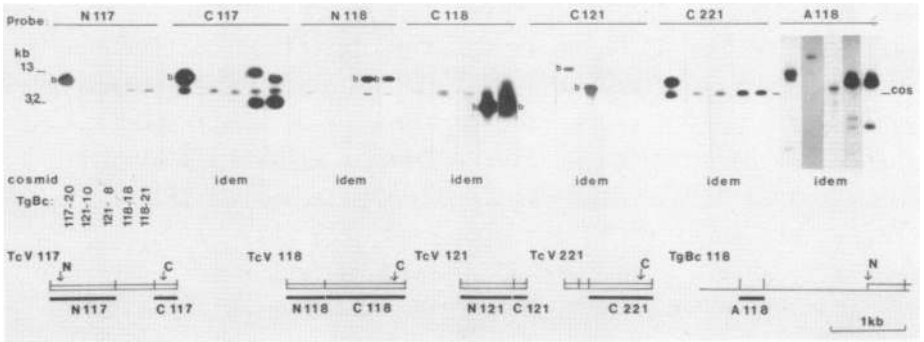

Fig. 4. Characterization of putative VSG genes on the 117, 118 and 121 BC cosmid clones. EcoRI digested cosmid DNA was size-fractionated in a 0.7% agarose gel and transferred to nitrocellulose filters, which were hybridized with the probes indicated. Post-hybridizational washes were performed at 3 x SSC (65°C). The open boxes indicate the coding sequence; b, bands containing the homologous BC gene.

of repetitive sequences in the 'uncuttable' regions surrounding the ELCs could be the cause of selection against ELC clones in the cosmid libraries. To lower the amount of 'uncuttable' region sequences per transformant, we constructed a library of small randomly-sheared fragments of 118a nuclear DNA inserted in pBR328 (unpublished). This library represents 10-20 times the genome and yielded the expected amount of BC clones, but no ELC clones. Thus, also short (smaller than 4 kb) segments of the 5' and 3' 'uncuttable' regions interfere with propagation in the host (E.coli HB101).

Identification of potential VSG genes on the cosmid clones by hybridization with cDNA probes

The hybridization of our cDNA probes with EcoRI restriction enzyme digests of some of the isolated cosmid DNAs is illustrated in Fig. 4. Probes that correspond to the 5'-half of the cDNA (N probes) only hybridize with a single band in cosmids containing the homologous BC gene. (Note that there is an additional band, indicated by 'cos', in many lanes; this band contains the cosmid vector DNA and it hybridizes to the pBR322 DNA still present in variable amounts in the different probes.) The 3'-half-gene probes hybridize to additional bands, besides the one that contains the homologous BC gene (marked b). This extra hybridization is only seen under non-stringent conditions and we attri-

bute it to the presence of additional VSG genes that are in part homologous to the probe [6,11,33-37].

The results in Fig. 4 define the sequence homology that is required for our probes to detect a VSG gene. The three 3'-half-gene probes C117, C118 and C221 are known to share limited blocks (up to 16 nucleotides) of sequence homology in the 3' untranslated part of the VSG mRNA [36], but this is not sufficient for cross-hybridization, as shown by the fact that the C117 and C221 probes do not hybridize to the EcoRI fragment that contains the 118 BC gene. The cross-hybridization is, therefore, caused by protein-coding sequences in the 3' terminus of the gene.

We conclude that the three different C probes - C117, C118 and C221 - recognize three different populations of VSG genes. The distribution of the members of these three VSG gene families on the cosmid clones will be discussed below.

Identification of VSG genes by the presence of the 5'-end sequence A of the transposed segment

We have shown previously that the 5'-end of the transposed segments of VSG genes 117 and 118 contains a homologous segment. This segment denoted 'A' is present in many copies in the genome and it differs from segment C, which is present at the 3'-edge of the transposed segment [9]. Sequence analysis of DNA upstream of the 118 BC gene has shown that segment A consists of a series of similar repeats (with an average length of 74 bp), five of which are located in a 375-bp MboII-HinfI fragment (Rijsewijk, F.A.M. and Van der Ploeg, L.H.T., unpublished). This fragment was used to demonstrate that segment A is also present in front of the 221 BC gene, cloned in plasmid TgB-221-1 (see Fig. 3).

A complication is that segment A is also found in a satellite-like DNA that is only infrequently cut by restriction endonucleases (see Fig. 5). Whereas digestion of nuclear DNA with a combination of the tetranucleotide-cleaving enzymes MspI, RsaI, HhaI and MboI reduces the overall length of nuclear DNA and of fragments hybridizing with probe C to around 0.5 kb, probe A mainly detects large DNA under these conditions.

VSG genes are clustered in trypanosome DNA

The results presented in the preceding section indicate

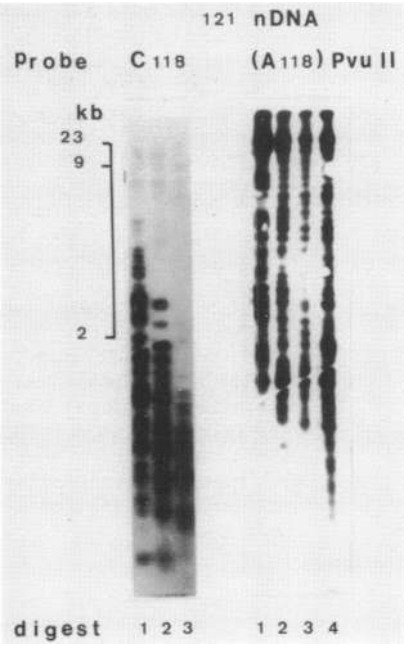

Fig. 5. Hybridization analysis of total nuclear DNA with the border fragments of the transposed segment of the VSG 118 gene. Variant 121 nuclear DNA was digested with: lane 1, MspI; 2, MspI+RsaI; 3, MspI+ RsaI+HhaI; 4, MspI+RsaI+HhaI+ MboI. The DNA was size-fractionated in a 0.7% agarose gel and transferred to nitrocellulose filters. Hybridization of the filters was performed with the 1-kb PvuII fragment isolated from the cosmid TgBc 118-33 (containing probe A118) and the 3' cDNA fragment from the cDNA clone TcV 118-3 (C118; Fig. 4). Post-hybridizational washed were performed at 3 x SSC (65°C).

that each of the three BC genes analysed in detail is associated with repeat A. Evidence that this association may be a characteristic of VSG genes in general, came from a comparison of the hybridization of the edge probes A and C with the colonies of the cosmid clone bank.

Fig. 6 shows the hybridization of the clone bank with pBR322, segment A118 and the 3' cDNA sub-probes C117 and C221. Only a section of the clone bank is shown. To aid the comparison of the different panels, 30 colonies hybridizing with probe A have been numbered. 22 of these hybridize with probe C117 and 19 with C221. All colonies that hybridize with C probes also hybridize with the A probe. Only 13 colonies hybridize with all three probes - A118, C117 and C221. Hybridization with C117 and C221 seems, therefore, randomly distributed in the population of colonies hybridizing with probe A118. Two of the 30 colonies hybridize with the A118 probe only. These could represent colonies containing a cloned A-satellite segment, clones with VSG genes not recognized by probes C117 and C221, or clones containing a single A segment. The results for the 1500 colonies were the same as for the 30 shown in Fig. 6. Under non-stringent hybridi-

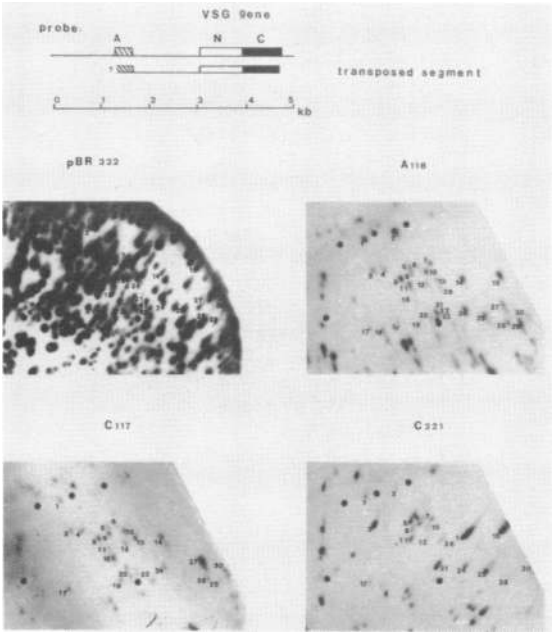

Fig. 6. Analysis of the distribution of putative VSG genes in the cosmid clone bank. Replica filters of a 118 cosmid clone bank containing only 1500 colonies were hybridized with various probes: plasmid pBR322, the 375-bp MboII-HinfI fragment containing the border of the 118 VSG gene transposed segment (probe A118), and the 3'-oriented sub-probes from the cDNA clones TcV 117-8 (probe C117) and TcV 221-12 (probe C221; Fig. 4). Post-hybridizational washes were performed at 3 x SSC (65°C). 30 colonies have been numbered in identical positions to facilitate comparison.

zation conditions, probe A118 recognized 135 of 1500 colonies. This repetitive element is, therefore, likely to be distributed in 9% of the genome.

It is clear from the figure that the hybridization intensity of colonies showed large variations. We attribute this to differences in sequence homology with the probes, colony size and cosmid copy number per cell. All colonies showing reproducible hybridization above the (low) background were, therefore, counted as positive.

Detailed analysis of the existence of putative VSG genes on individual cosmid clones

The presence of multiple VSG genes in individual cosmids was further analysed by two-dimensional blotting hybridization and hybridization with probes A and C. Fig. 7 shows a two-dimensional blotting hybridization of the TgBc 118-33 VSG cosmid clone after digestion with PvuII. The lower half of this figure summarizes the information obtained from this and other two-dimensional blotting hybridization experiments on the position of repetitive elements in this 118 cosmid. In experiments in which both dimensions contain the same digest of one clone, a

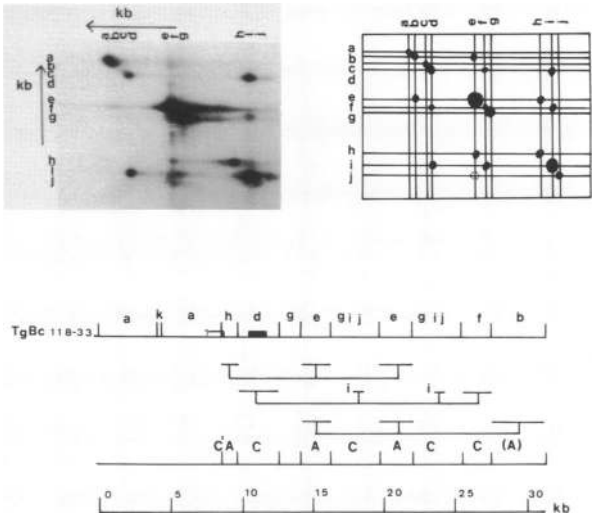

Fig. 7. Two-dimensional electrophoresis blotting hybridization analysis of a PvuII digest of the VSG 118 BC cosmid TgBc-33. Post-hybridizational washes were performed at 65°C in 3 x SSC. The left-hand upper panel shows the autoradiogram; the right-hand upper panel a schematic outline of the hybridization pattern. The spot in the horizontal dimension between bands D and E is derived from a partial digestion product. The lower half of the figure shows the PvuII map of TgBc 118-33 together with a summary of the hybridizing bands. The open box indicates the position of a VSG gene detected by sequence analysis. The cross-hybridizing bands are connected by horizontal lines. A and C refer to the 5' and 3' border fragments of the 118 transposed segment, respectively. See legend to Fig. 4 for abbreviations.

The repetitive elements on the cosmid clone TgBc 118-33 are located on the PvuII fragments h (containing repeat A) and e (which hybridizes strongly to PvuII fragment h). Because the PvuII fragment e is part of a duplicated area of the genome it migrates as a double band. Different cosmid clones, containing a normal and a shortened PvuII fragment e, were therefore used to show that both PvuII bands e contain the 74-bp repeat A. This is shown by the hybridization of probe A118 to an EcoRI digestion of cosmid TgBc 118-18 in Fig. 4. Three hybridizing bands are visible: one of 10 kb (containing the 5'-end of the 118 transposed segment),one of 2.3 kb (containing PvuII fragment e on a border fragment of the clone), and a 1.9-kb EcoRI fragment (containing the second PvuII fragment e). In the scheme, fragment h can therefore be inferred to hybridize with both PvuII fragments e. In addition to the hybridization with the 5' repeat (A118), fragments are found to hybridize with the 118 BC VSG gene coding sequence as well. This hybridization is located at the 3' side of each PvuII fragment e. In the autoradiogram this is visible as the strong hybridization of PvuII fragment d (containing the 118 BC gene) with the 800-bp double fragment PvuII i. Because a double band is observed, different cosmids were again used to prove that the 3'-end of the 118 VSG gene does indeed hybridize with both segments (not shown).

diagonal of the hybridizing spots is observed originating from identical digestion products in both dimensions. Off-diagonal spots are derived from (repeated) sequences present in fragments of different size and, therefore, from different areas of the cosmid clone. The repetitive elements on the cosmid clone TgBc 118-33 are located on the PvuII fragments h and e, containing sequences that hybridize with probe A118 and the fragments d, i and f which contain sequences that hybridize with 3' VSG gene coding sequences (for a more detailed discussion see legend to Fig. 7).

In the lower half of Fig. 7 a general picture of the location of the repeats A and C on the 118 BC cosmid clone is shown. At least three sets of linked repeats A and C are found in a tandem array. A fourth set is present at the very 3'-end of the cosmid clone, as the 2-kb PvuII fragment f hybridizes with fragment PvuII d. It is, however, difficult to identify its bordering 5' repeat because only the PvuII fragments e and not the 118 5' border repeat A (on fragment h) hybridizes with the 3.6-kb PvuII fragment b. In the other 118 BC cosmid clones extending 3' (118-40; not shown) and 5' (118-18 and 118-28; see Figs 2 and 3), additional hybridization with the repeats A and C is found, but this has not been precisely mapped.

A similar pattern of multiple regions hybridizing with probes A and C was observed for the 40-kb area surrounding the 117 BC VSG gene (Fig. 8). The repeats A and C are found at regular intervals on the cloned DNA. However, a 5' border repeat A is not always found in tandem with a 3' repeat C, even though it was tested with the 3'-ends of the 117, 118 and 221 cDNA sub-probes.

Only two fragments of the 121 cosmids show hybridization with probe A118 (Fig. 4). This hybridization is located at the 3'-side of the 121 VSG gene. The clones are further devoid of repeats A and C. In this case, tandem linkage of A and C hybridization is less obvious than with the 117 and 118 cosmids and no hybridization with the C117, C118 or C221 probes was observed at all (see Fig. 4). We return to this point in the Discussion.

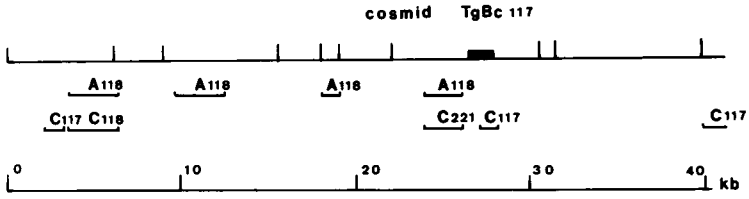

Fig. 8. Positions of putative VSG genes on the VSG 117 BC cosmid clones. In the EcoRI restriction enzyme map, the fragments hybridizing with the 5' border probe of the VSG 118 transposed segment and with four different cDNAs are shown. See legend to Fig. 4 for abbreviations.

DISCUSSION

We have used two types of hybridization probes to identify potential VSG genes in trypanosome DNA. Probe A is derived from the 5'-edge of the transposed segment of the 118 BC gene. It hybridizes with sequences in front of the 117, 118 and 221 BC genes and it may, therefore, recognize a large fraction of the potential VSG genes present. Probes C117, C118 and C221 are derived from regions corresponding to the C-terminus of three VSG mRNAs. Each of these probes hybridizes to more than 50 bands in nuclear DNA blots, but C117 and C118 recognize a different set of sequences than C221. How many other sets are present that are not detected by our C probes is not known. It is certain, however, that there are potential VSG genes that are recognized neither by probe A nor by the available probes C, because DNA sequence analysis has recently uncovered an additional VSG gene in front of the 118 BC gene that had not been detected by hybridization (see Fig. 7, fragment C'). On the other hand, it is possible that our probes also detect sequences not linked to VSG genes and this is even certain for the A probe which hybridizes to a satellite-like DNA. Since the 74-bp repeats which make up segment A are not cut by MboI, the restriction endonuclease used in making the cosmids,it is likely that the satellite-like DNA containing A is not present in our clone bank.

Notwithstanding these uncertainties, our data provide strong evidence for the clustering of VSG genes in the trypanosome genome. The fact that the A probe and the C probes hybridize to the same sub-set of approximately 9% of the cosmid clones, shows that these sequences are linked in the genome. The C117

and C221 probes hybridize in part to the same cosmids, showing linkage of potential VSG genes. Further evidence for linkage comes from the analysis of the 118 cosmids which show the presence of at least five potential VSG genes in tandem. Multiple hybridizing fragments were also observed with the 117 cosmids, but in this case, no clear tandem linkage of A and C repeats is obvious. We attribute this to our inability to detect the 3' edges of all VSG gene transposition units present. With the 121 cosmids the situation is less clear. The two isogenes isolated differ by multiple inversions, deletions and insertions in their 3' flanking regions. The 121 gene is expressed very early in infection and such genes are known to have unusual 3' flanking sequences (see refs 5,6). The presence of only two areas in this region that hybridize with probe A may be related to this.

Can one assess the number of VSG genes in the trypanosome nucleus from these data? If we assume that the VSG gene packing present in the 118 cosmid (five genes in about 30 kb) is representative for the 9% cosmids found to contain potential VSG genes, the total number of genes in a diploid nucleus with a genomic complexity of 35×10^3 kb would be $0.09 \times 2 \times 35 \times 10^3 \times (5/30) = 1050$. A further assumption in this calculation is that there is no bias in cloning VSG genes relative to the remainder of the nuclear DNA. This is the case for the 118 VSG gene (see Results), but the number of 117 and 121 cosmids isolated was below that expected. The special clan of BC genes that do not give rise to a detectable ELC when activated and to which our 221 gene belongs [6], is also relatively refractory to cloning. Furthermore, the actual transposed segment in the case of the 118 BC gene is maximally 3.5 kb and if the packing of genes is tighter than suggested in Fig. 7, a further source of underestimation might have been introduced. On the other hand, the assumption that all cosmids hybridizing with probe A are packed with VSG genes may lead to an overestimate of the gene number as suggested by the results with the 121 cosmids. We consider 10^3 genes, therefore, a reasonable estimate.

An interesting point is that the 118 cosmids do not contain genes that cross-hybridize with 5'-half-gene 118 probes, even though such genes are present elsewhere in the genome. This is

even more remarkable for the 117 cosmid since the 117 VSG gene is a member of a large gene family and even 5'-half-gene probes hybridize to multiple bands in Southern blots. Related genes are thus not highly clustered in the genome. Dispersal could be related to the duplication-transposition mechanism for VSG gene activation, i.e. genes could be mis-transposed or the gene that has been discarded from the expression site might occasionally re-insert in the genome. Whether the 74-bp repeat has a function in the recombinational event at transposition is not clear since it is not used in that sense in the 221 VSG gene. However, the repeat could fullfil a function in maturation of precursor mRNAs since two variant-specific RNAs map at similar positions, approximately 74 bp apart in this repeat, in front of the 118 VSG gene [10]. Why VSG genes are not randomly spread throughout the genome is not clear. An obvious possibility is that genes are most efficiently transposed into an expression site if they are located on the same chromosome. Trypanosomes do not have condensed chromosomes in any phase of their cell cycle and the chromosomal location of VSG genes can, therefore, not be studied by cytological hybridization. Attempts to fractionate intact trypanosome DNA into chromosome-like molecules are under way.

ACKNOWLEDGEMENTS

We thank Dr. A.Y.C.Liu and Mr. F.A.M.Rijsewijk for sharing unpublished results; Dr. R.A.Flavell and Dr. T.Lund (Lab. of Gene Structure and Expression, National Institute for Medical Research, Mill Hill, London, U.K.) for the gift of the pRTI and POPFI vectors, helpful discussions and hospitality to one of us; Dr. J.Groffen (idem) for advice on the two-dimensional blotting hybridization technique; Miss F.Tausk and Miss J.Van den Burg for assistance; Dr. R.Benne for critical reading of the manuscript; and Mrs G.Siebericchs for typing the first draft of the manuscript. This work was supported in part by grants to PB from the Foundation for Fundamental Biological Research (BION), which is subsidized by The Netherlands Organization for the Advancement of Pure Research (ZWO) and The World Health Organization (T16/181/T7/34). LHTVDP received an EMBO short-term fellowship in 1980.

Abbreviations: BC, basic copy; bp, base pair(s); cDNA, complementary DNA; ELC, expression-linked copy; kb, kilo-base pair(s); mRNA, messenger RNA; SDS, sodium dodecylsulphate; SSC, 0.15 M NaCl, 0.015 M sodium citrate (pH 7.0); VSG, variant surface glycoprotein.

* Present address: Laboratory of Physiological Chemistry, Sylvius Laboratories, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands.

REFERENCES

1 Hoeijmakers, J.H.J., Frasch, A.C.C., Bernards, A., Borst, P. and Cross, G.A.M. (1980) Nature 284, 78-80.
2 Capbern, A., Giroud, C., Baltz, T. and Mattern, P. (1977) Exptl.Parasitol. 42, 6-13.
3 Pays, E., Van Meirvenne, N., Le Ray, D. and Steinert, M. (1981) Proc.Natl.Acad.Sci.U.S. 78, 2673-2677.
4 Vickerman, K. (1978) Nature 273, 613-617.
5 Marcu, K.B. and Williams, R.O. (1981) in: Genetic Engineering (Setlow, S.K. and Hollaender, A., Eds), Vol. 3, Plenum Press, New York, pp. 129-155.
6 Borst, P. and Cross, G.A.M. (1982) Cell 29, 291-303.
7 Pays, E., Lheureux, M. and Steinert, M. (1981) Nature 292, 365-367.
8 Bernards, A., Van der Ploeg, L.H.T., Frasch, A.C.C., Borst, P., Boothroyd, J.C., Coleman, S. and Cross, G.A.M. (1981) Cell 27, 497-505.
9 Van der Ploeg, L.H.T., Bernards, A., Rijsewijk, F.A.M. and Borst, P. (1982) Nucl.Acids Res. 10, 593-609.
10 Van der Ploeg, L.H.T., Liu, A.Y.C., Michels, P.A.M., De Lange, T., Borst, P., Majumder, H.K., Weber, H., Veeneman, G.H. and Van Boom, J. (1982) Nucl.Acids Res. 10, 3591-3604.
11 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, 935-943.
12 Fairlamb, A.H., Weislogel, P.O., Hoeijmakers, J.H.J. and Borst, P. (1978) J.Cell Biol. 76, 293-309.
13 Birnboim, H.C. and Doly, J. (1979) Nucl.Acids Res. 7, 1513-1523.
14 Borst, P. and Fase-Fowler, F. (1979) Biochim.Biophys. Acta 565, 1-12.
15 Borst, P., Fase-Fowler, F. and Gibson, W.C. (1981) Mol.Biochem.Parasitol. 3, 117-131.
16 Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc.Natl.Acad.Sci.U.S. 76, 3683-3687.
17 Southern, E.M. (1975) J.Mol.Biol. 98, 503-517.
18 Sato, S., Hutchison, C.A. and Harris, J.I. (1977) Proc.Natl.Acad.Sci.U.S. 74, 542-546.
19 Denhardt, D.T. (1966) Biochem.Biophys.Res.Commun. 23, 641-646.
20 Vogelstein, B. and Gillespie, D. (1979) Proc.Natl.Acad.Sci.U.S. 76, 615-619.
21 Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J.Mol. Biol. 113, 237-251.
22 Jeffreys, A.J. and Flavell, R.A. (1977) Cell 12, 429-439.
23 Grosveld, F.G., Dahl, H.H.M., De Boer, E. and Flavell, R.A. (1981) Gene 13, 227-237.
24 Ish-Horowicz, D. and Burke, J. (1981) Nucl.Acids Res. 9, 2989-2998.
25 Cami, B. and Kourilskey, P. (1978) Nucl.Acids Res. 5, 2381-2390.
26 Hanahan, D. and Meselson, M. (1980) Gene 10, 63-67.
27 Borst, P., Fase-Fowler, F., Frasch, A.C.C., Hoeijmakers, J.H.J. and Weijers,P.J. (1980) Mol.Biochem.Parasitol. 1, 221-246.
28 Michels, P.A.M., Bernards, A., Van der Ploeg, L.H.T. and Borst, P. (1982) Nucl.Acids Res. 10, 2353-2366.
29 Boyer, H.W. and Rouland-Dussoix, D. (1969) J.Mol.Biol. 41, 459-472.
30 Morrison, D.A. (1979) in: Methods in Enzymology (Wu, R., Ed.), Vol. 68, Academic Press, New York, pp. 326-331.
31 Borst, P., Van der Ploeg, M., Van Hoek, J.F.M., Tas, J. and James, J. (1982) Mol.Biochem.Parasitol., in press.
32 De Lange, T. and Borst, P. (1982) Nature, submitted.
33 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.), North-Holland, Amsterdam, pp. 235-239.
34 Rice-Ficht, A.C., Chen, K.K. and Donelson, J.E. (1981) Nature 294, 53-57.
35 Matthyssens, G., Michiels, F., Hamers, R., Pays, E. and Steinert, M. (1981) Nature 293, 230-233.
36 Majumder, H.K., Boothroyd, J.C. and Weber, H. (1981) Nucl.Acids Res. 9, 4745-4753.
37 Boothroyd, J.C., Cross, G.A.M., Hoeijmakers, J.H.J. and Borst, P. (1980) Nature 288, 624-626.