

## Comparison of the Expression-Linked Extra Copy (ELC) and Basic Copy (BC) Genes of a Trypanosome Surface Antigen

John E. Donelson, William J. Murphy, Steven T. Brentano,  
Allison C. Rice-Ficht, and George D. Cain

*Departments of Biochemistry (J.E.D., W.J.M., S.T.B., A.C.R.-F.) and Zoology (G.D.C.),  
University of Iowa, Iowa City, Iowa 52242*

A recombinant clone of an expression-linked extra copy (ELC) gene of a trypanosome-variable surface glycoprotein was sequenced. In addition the sequences of the corresponding cDNA and portions of the two basic copy genes were determined. Comparison of these sequences reveals that the 5' boundary of the ELC-transposed segment (2.2 kb) occurs within a repetitive sequence about 700 bp upstream from the start codon of the coding sequence. This sequence does not contain internal symmetries and is not homologous with the repetitive sequence at the 3' boundary. The first 35 nucleotides of the cDNA are different than the corresponding ELC sequence and presumably were transcribed from another genomic location. A restriction fragment containing predominantly sequences outside of the 5' boundary hybridizes to a Pst I fragment whose length is variable in different trypanosome clones. This hybridization pattern is similar to that observed using probes for surface glycoprotein genes that are expressed via the nonduplication-associated (NDA) mechanism rather than the ELC mechanism. This indicates that there is a sequence correlation between these two DNA rearrangement mechanism.

**Key words:** trypanosome, genes, rearrangement, surface antigen

Recent evidence indicates that two types of DNA rearrangements are associated with the antigenic variation displayed by African trypanosomes [see 1 for a recent review]. The trypanosome genome contains several hundred, and perhaps several thousand, genes which code for immunologically distinct variants of the major trypanosome surface antigen, the variable surface glycoprotein (VSG). In an individual trypanosome only one of these genes is expressed at a given instant; the other genes appear to be transcriptionally silent. The crucial molecular event in the phenomenon of antigenic variation is the switch from the expression of one VSG gene to the expression of another. This switch appears to occur spontaneously at a very low rate. One type of DNA rearrangement that accompanies the switch is the formation of an expression-linked extra copy (ELC) of one of the silent basic copy (BC) genes. The

Received April 5, 1983; accepted August 1, 1983.

ELC is transcribed [2] and is usually present only in those trypanosomes synthesizing that particular VSG [1]. Other VSG genes are not expressed via the ELC mechanism. In these cases the transcribed gene is located on a restriction fragment whose size varies in different trypanosome clones. The size differences have been found to occur within the region 3' distal to the transcribed VSG gene and are due to DNA rearrangements whose exact nature is not clear. This type of rearrangement will be called the nonduplication-associated (NDA) mechanism, to distinguish it from an ELC where there is clearly a duplication and translocation of a previously silent BC gene. A final point is that all ELC and NDA genes appear to be located 7 to 30 kb from a chromosomal telomere, or at least a double-strand DNA break which mimics a telomere in Southern hybridization experiments [3,4]. Thus it is possible that the variability of the region 3' distal to NDA genes is simply related to the replication mechanism of the adjacent telomere rather than to expression of the VSG gene.

In the work to be described, we have obtained and characterized recombinant DNA clones of an ELC and its corresponding BC genes. Sequence comparisons have identified the boundaries of the transposed ELC segment. In addition we demonstrate that the sequences flanking the transposed ELC also flank a VSG that is expressed via the NDA mechanism. This provides the first evidence that the two types of DNA rearrangements may involve similar molecular mechanisms.

## **MATERIALS AND METHODS**

### **Trypanosomes**

The IaTat serodeme was derived from *Trypanosoma brucei* strain 427 [5] by Betsy Bricker. [Further details about the IaTat serodeme and the VSGs expressed by the individual clones are given in 6.] Trypanosome poly A<sup>+</sup> RNA and DNA were isolated from cloned trypanosome populations by a slight modification [of the procedure described in 7]. The DNA and RNA were resolved on a step gradient of 5.7 M CsCl and 3.0 M CsCl in 0.1 M disodium edetic acid (EDTA). The poly A<sup>+</sup> RNA was selected on oligo dT cellulose [8].

### **Construction and Screening of the IaTat 1.2 cDNA Library**

The cDNA library was constructed from poly A<sup>+</sup> RNA of IaTat 1.2 using the technique of Land et al [9] in which an oligo dC "tail" is added to the 3' end of the first strand cDNA and oligo dG is used to prime synthesis of the second strand. The double-stranded cDNAs were then inserted into the Pst I site of pBR322 via the oligo dC:oligo dG joining technique [10]. The library was screened [11] by the "plus/minus" hybridization procedure using [<sup>32</sup>P] first strand cDNAs from IaTats 1.2 and 1.10, respectively. Those colonies to which the IaTat 1.2 cDNA hybridized, but the IaTat 1.10 cDNA did not, were selected for further characterization.

### **Cloning Genomic Restriction Fragments**

Genomic DNA (100  $\mu$ g) was digested with Pst I or a combination of Hind III and Xba I and the resultant fragments were separated on a 0.7% preparative gel containing standard length restriction fragments in side lanes. DNA fragments of the appropriate size were eluted as described [26] and cloned into the Pst I site of plasmid pBR322 or the Hind III and Xba I sites of plasmid pXR-16 [Jim Hartley, personal communication]. Transformants were screened for the expected fragments by the

method of Grunstein and Hogness [27] using nick-translated cDNA fragments [13] as probes. Plasmids were prepared from positive clones.

### Genomic Southern Hybridizations

Genomic DNA (5  $\mu\text{g}$ ) from the different trypanosome clones was digested with the appropriate restriction enzymes, resolved on 0.7% agarose gels, and transferred to nitrocellulose paper as described [12]. Nick-translated restriction fragments [13] were used as the radioactive probes and hybridizations were conducted at 65°C for 12–13 hours in 4  $\times$  SET, 10 $\times$  Denhardt solution, 0.1% sodium pyrophosphate, 0.1% sodium dodecylsulfate (SDS), and 50  $\mu\text{g}/\text{ml}$  boiled sonicated salmon sperm DNA as described [14]. The filters were washed three times in 0.1  $\times$  SET, 0.1% SDS for one hour each at 65°C and exposed to Kodak XAR-5 film for two to 14 days at  $-70^\circ\text{C}$  with intensifier screens.

### DNA Sequence Analysis

The chemical reactions of Maxam and Gilbert [15] were used for all sequence determinations. Restriction fragments with recessed 3' termini were differentially end-labeled with [ $\alpha$ - $^{32}\text{P}$ ] dNTPs and Escherichia coli DNA polymerase I [16]. The restriction sites recognized by Hind III, Eco RI, Sal I, Sau 3A, and Ava II were used for end-labeling. The chemical degradation products were separated on 0.4-mm-thick, 80-cm-long gels of 5% or 6% polyacrylamide in 8 M urea and 50 mM TRIS-borate, 1 mM EDTA, pH 8.3.

## RESULTS

### The IaTat 1.2 Gene Family

The IaTat serodeme of trypanosome clones is derived from *Trypanosoma brucei* strain 427 [5,6]. A stabilate of the original clone was obtained by one of us (G.D.C.) in 1976 and passaged through a rabbit. Individual clones were isolated at increasing times of infection over a period of 84 days. The immunologically distinct clones are called IaTat 1.1 through IaTat 1.10 according to their times of isolation.

A cDNA library of the polyA<sup>+</sup> RNA from trypanosome clone IaTat 1.2 was constructed by inserting double-stranded cDNA into the Pst I site of pBR322 via the oligo dG:oligo dC technique. The VSG cDNAs were tentatively identified by screening the library with first-strand [ $^{32}\text{P}$ ]-cDNA prepared from the polyA<sup>+</sup> RNAs of IaTat 1.2 and IaTat 1.10. Colonies giving a positive signal in this "plus/minus" hybridization were selected for further characterization. One such colony was found to contain a cDNA insert of about 1,600 bp. Its complete sequence was determined and found to possess the features expected of a VSG cDNA [17]. An open translation reading frame of 493 codons is present. The last 100 codons specify an amino acid sequence very similar to the C-terminal sequences of members of one of the two VSG homology subsets [18]. Likewise, the 3' nontranslated region is very similar to the corresponding region in other VSG cDNAs [1].

When the IaTat 1.2 VSG cDNA is used as a radioactive probe in a Southern genomic blot, the result shown in Figure 1 is obtained. The genomic DNAs were digested with Hind III, an enzyme which does not cleave the cDNA. The probe hybridizes to two restriction fragments of 4.2 and 8.2 kb in genomes which are not expressing the IaTat 1.2 VSG. It hybridizes to these two fragments plus another

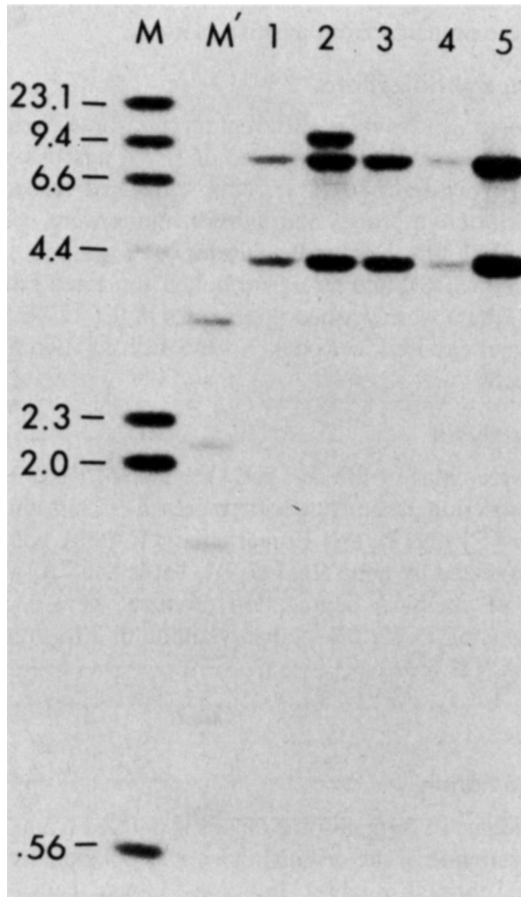


Fig. 1. Autoradiogram of genomic Southern blots using IaTat 1.2 cDNA as the radioactive probe. The genomic DNAs were digested with Hind III and were from IaTat 1.1 (lane 1), IaTat 1.2 (lane 2), IaTat 1.3 (lane 3), IaTat 1.9 (lane 4), and IaTat 1.10 (lane 5). Lanes marked M and M' contain known restriction fragments of  $\lambda$ DNA, pBR322 DNA, and SV40 DNA as length markers.

fragment of 10 kb in the genome expressing the IaTat 1.2 VSG. This information demonstrates that this VSG is expressed via the ELC mechanism. It also suggests that there are two corresponding BC genes to which the cDNA hybridizes.

A large number of Southern genomic blots were conducted using different restriction enzymes to map the locations of restriction sites within and surrounding the ELC and BC genes. This information is summarized in the form of restriction maps shown in Figure 2. The ELC is located about 8 kb from a telomere-like double-strand break. Sequences flanking the transposed segment appear to be devoid of restriction sites. These so-called "barren" regions without restriction sites are similar to those surrounding other ELC genes which have been observed by other laboratories [1]. The two BC genes are within more conventional genomic regions, surrounded by restriction sites. There are several restriction differences within the two BC genes.

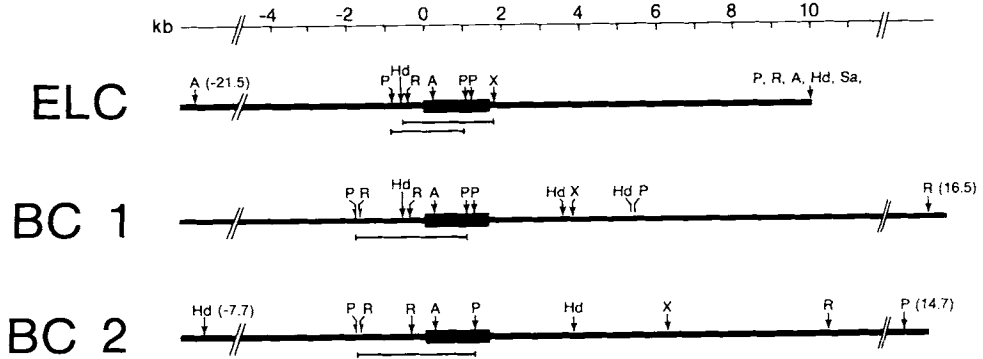


Fig. 2. Partial restriction maps of the three genomic regions to which Iatad 1.2 cDNA hybridizes. Rectangular boxes indicate the coding region. The restriction sites are for *Ava* II (A), *Pst* I (P), *Hind* III (Hd), *Eco* RI (R), *Xba* I (X), *Sal* I (Sa), and *Hinf* I (Hf). Note that a double-stranded DNA break occurs about 10 kb to the left of the start of the coding sequence in the ELC. This break mimics cleavage sites for the restriction enzymes. Short horizontal lines underneath the thick lines (representing the DNA) indicate genomic regions that were cloned directly into plasmid pBR322 or its derivative pXR-16 for sequence analysis.

For example, BC 2 is missing sites for *Pst* I, *Hind* III, and *Sau* 3A that are present in BC 1 and the ELC. This suggests that the ELC is a duplication of BC 1, rather than BC 2.

### The 5' Flanking Sequences of the ELC and Two BC Genes

Figure 2 also shows the locations of restriction fragments that were cloned directly from the IaTad 1.2 genome into a bacterial plasmid for further characterizations. These fragments were chosen because they would permit us to compare the sequences preceding the ELC and two BC genes and hopefully, to determine the 5' boundary of the transposed ELC segment. Sequence determination of the three corresponding regions were undertaken and a comparison of these sequences is shown in Figure 3. The comparison begins about 900 nucleotides before the ATG initiation codon of the VSG coding sequence. In addition a small portion of the 5' end of the IaTad 1.2 cDNA is shown near the bottom of the figure.

Several interesting features are apparent from the sequence comparison. First, the ELC and BC 1 sequences are very different during the initial 70 nucleotides and become similar, but nonidentical, for the next 140 nucleotides. They are then identical throughout the rest of the comparison which extends into the first few codons of the coding sequence. The comparison of BC 2 with the ELC and BC 1 begins at nucleotide position 403 and continues with one short interruption to the end of the figure. There are a variety of nucleotide replacements scattered throughout this region. Since the simplest explanation of these differences is the occurrence of random mutations within previously duplicated BC genes, the sequence determination of BC 2 was not continued. Furthermore, the differences confirm the Southern hybridization data which indicated that the ELC arose from duplication of BC 1 instead of BC 2.

A second feature of the comparisons in Figure 3 is that the first 35 nucleotides of the cDNA do not match the corresponding positions in the ELC. There are 23 differences in this region suggesting that the 35 nucleotides at the 5' end of the cDNA



were transcribed from another location. They are not homologous with any portion of the ELC or BC 1 which was sequenced and probably arose from a genomic DNA sequence to the left of the sequenced region of the ELC shown in Figure 3. An Rsa I site is present in the 35 nucleotides and preliminary Southern hybridizations of the Rsa I-digested genomic DNA seem to indicate that this sequence is located as far as 50 kb to the left of the ELC coding sequence [W.J.M., unpublished]. This implies, but does not prove, that a large intron is present in a precursor VSG RNA which is subsequently removed during processing to form the VSG mRNA. Similar observations have been made by Van der Ploeg et al [19] and Boothroyd and Cross [28] in their studies of three different VSG gene systems. And, in fact, the 5' termini of their three VSG cDNAs are identical to the 5' terminus of the Iatad 1.2 cDNA through the first 35 nucleotides, after which they are different. This suggests that all four VSG mRNAs are transcribed either from the same promoter or from separate promoters containing the same sequence at the start of transcription. At the moment there is no evidence to distinguish between these possibilities.

### The 5' Boundary of the ELC-Transposed Segment

From the sequence comparison of the ELC and BC shown in Figure 3, it would appear that the 5' end point of the duplication that generates the ELC is located to the right of nucleotide position 208. This nucleotide is on the right-hand side of a region of 140 nucleotides in which are 38 nucleotide differences. To the left of this 140 bp, the ELC and BC 1 sequences have virtually nothing in common; to its right they are identical. The nature of the sequences within and bordering this recombination boundary in the genomes of other trypanosome clones was examined by genomic DNA hybridizations using the Southern techniques [as described in 12 and 17].

Three restriction fragments from the cloned ELC were used as probes in these experiments. Using the numbering system of Figure 3, they are a Pst I/Dde I fragment (positions 1-154), a Pst I/Hind III fragment (positions 1-297) and a Hind III/Sal I fragment (positions 297-675).

The Hind III/Sal I fragment hybridized only to the expected ELC and two BC fragments in a genomic blot (not shown). Therefore this sequence within the ELC transposed segment is not repeated in the genome. In contrast the Pst I/Hind III fragment hybridizes to many regions of the genome (Fig. 4). This fragment includes the 140-bp region of nonperfect homology between the ELC and BC I. It suggests that this boundary region is a repetitive sequence in the genome. The third fragment used as a probe, the Pst I/Dde I fragment covering position 1-154 of the ELC sequence, yielded a quite different pattern. This fragment includes the 70-bp portion of the ELC sequence that is not homologous with BC 1 and another 84 bp of partial homology. Figure 5 shows the results of this probe hybridizing to the genomes of IaTats 1.1, 1.2, and 1.3. Under high-stringency hybridization conditions (see Materials and Methods), the probe hybridizes to (1) the ELC gene in IaTat 1.2, (2) two Pst I fragments containing the 5' portions of Iatad 1.2 BC genes, and (3) four other Pst I fragments. One of these additional Pst I fragments (indicated by the arrows pointing downward) is a different size in the three genomes; i.e., it is variable. When the genomes are double-digested with Pst I and Xba I and hybridized with the Pst I/Dde I probe, the variable Pst I fragment is converted to a constant-sized fragment of 2.4 kb in both genomes (arrows pointing upward). A similar result is obtained when the genomes of other trypanosome clones in the IaTat serodeme are used (not shown). In

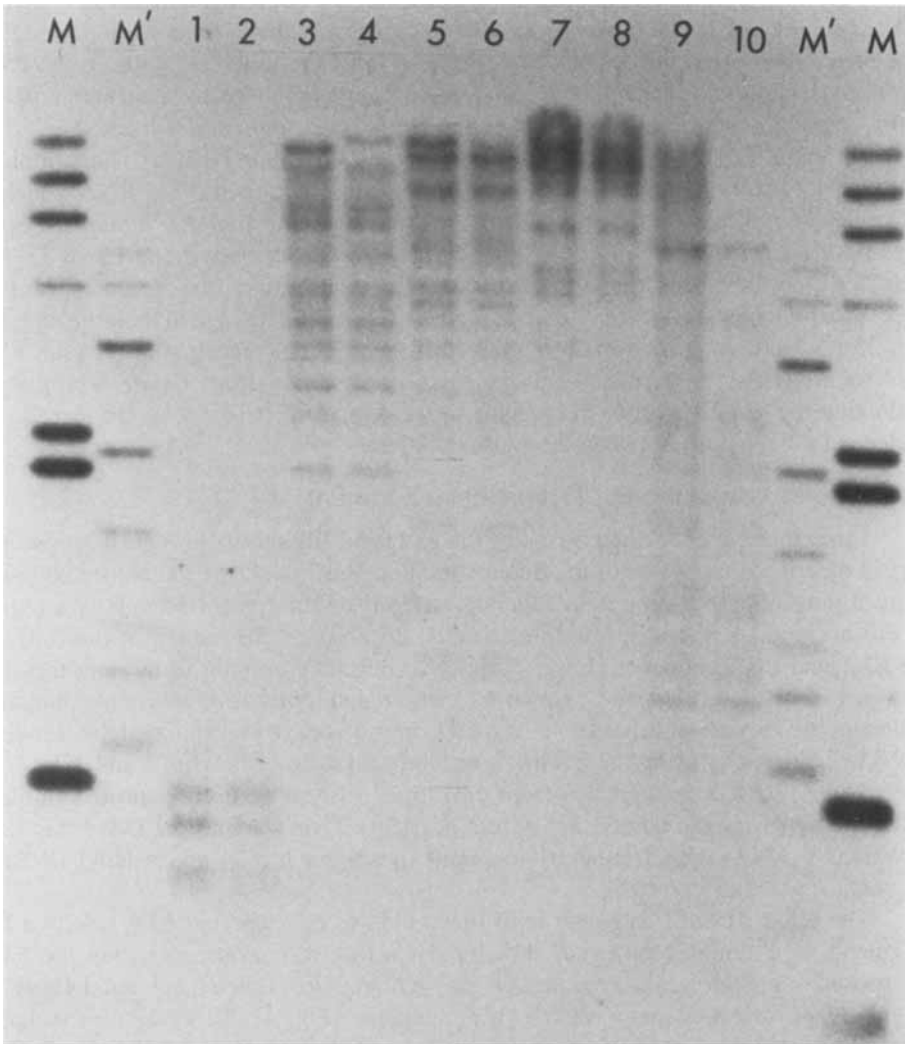


Fig. 4. Autoradiogram of a genomic Southern hybridization using as radioactive probe a restriction fragment of the ELC that extends from a Pst I site at position 1 to the Hind III site at position 297. Odd-numbered lanes contain IaTat 1.2 genomic DNA and even-numbered lanes contain IaTat 1.10 genomic DNA. The genomic DNAs were digested with Dde I (lanes 1 and 2), Hind III (lanes 3 and 4), Eco RI (lanes 5 and 6), Sal I (lanes 7 and 8), and Sau 3A (lanes 9 and 10). Lanes marked M and M' are the same as in Figure 1.

all cases the probe hybridizes to a sequence which occurs on a variable-sized Pst I fragment and a constant-sized Pst I/Xba I fragment of 2.4 kb. In other words, Xba I releases the ELC coding sequence from its 3' distal "telomere-like" end (see Fig. 2) and *also* separates this new hybridizable sequence from its adjacent variable regions. Therefore, the Southern hybridization pattern of this new sequence is very similar to that observed for VSG genes expressed via the NDA mechanism [20].



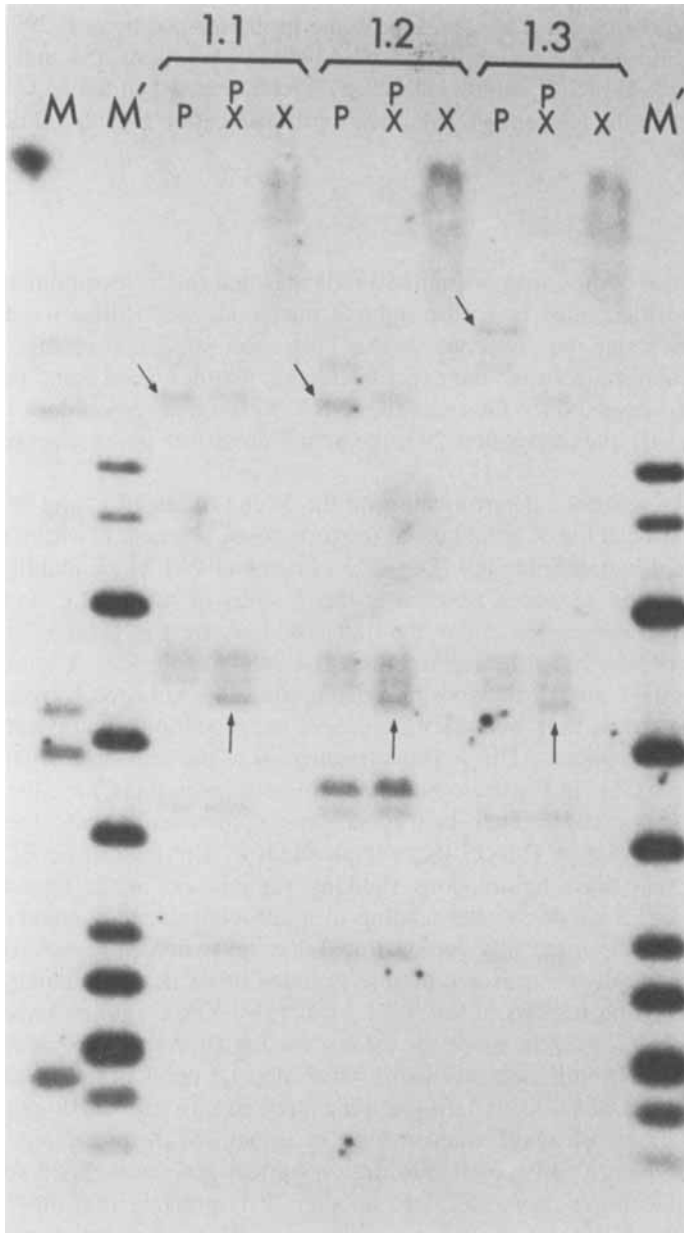


Fig. 5. Autoradiogram of a genomic Southern hybridization using as radioactive probe a restriction fragment of the ELC that extends from a Pst I site at position 1 to a Dde I site at position 154. Lanes grouped under the headings of 1.1, 1.2, and 1.3 contained genomic DNA from IaTats 1.1, 1.2, and 1.3, respectively. The genomic DNAs were digested with Pst I (lanes marked P), Pst I plus Xba I (lanes marked PX), or Xba I (lanes marked X). Arrows pointing downward indicate the variable-sized fragment detected in the Pst I digests. Vertical arrows pointing upward show the resultant constant-sized fragment of 2.4 kb in the Pst I/Xba I double digests.

Another point is that the probe from positions 1–154 does not contain the repetitive sequence element that is present in the probe for positions 1–297 (compare Figs. 4, 5). Therefore the repetition must be between positions 154 and 297. This region includes position 208, which is the last difference between the ELC and BC 1. Thus it appears that the border of the duplication/translocation is within this repetitive sequence.

## DISCUSSION

The sequence comparison in Figure 3 indicates that the 5' recombination site of the ELC transposition must be to the right of nucleotide 208. Since we don't have any information about the sequence at the "insertion site" before this ELC was generated, we cannot determine the exact nucleotide that defines the end point of the transposition. However, since the repetitive DNA within this region does not extend beyond the Hind III site at position 297, the actual crossover point is probably quite close to position 208.

Preliminary sequence determinations at the 3' end of the ELC and BC 1 coding sequences indicate that the 3' boundary of the transposed segment is within the codons for the C-terminal hydrophobic tail [David Dorfman and W.J.M., unpublished]. This is consistent with the sequence analysis at the 3' ends of other ELC and BC gene systems [21–23]. It demonstrates that the transposed segment in IaTat 1.2 is approximately 2.2 kb, of which 1.5 kb (493 codons) is coding sequence. A comparison of the sequences at the 5' and 3' transposition boundaries does not reveal any homologies or internal symmetries. Both boundary sequences are repetitious but do not share any common sequence elements. This is rather analogous to the sequence organization of yeast mating type [24]. In yeast, common sets of sequences flank both the  $\alpha$ - and a-cassette genes (basic copy genes in trypanosome nomenclature) and the gene expressed at the mating type (MAT) locus (equivalent to a trypanosome ELC). It has been proposed that these homologous flanking regions aid in the specific pairing between MAT and a silent cassette leading to a subsequent gene conversion event. Furthermore, it has been recently demonstrated that the switch from one yeast mating type to the other involves a transient double stranded break at the boundary of one of the conserved flanking regions of the MAT locus [25]. These authors have proposed two possible molecular mechanisms for the conversion from one gene to the other in the yeast MAT locus, both mechanisms of which depend upon strand invasion of the cleaved DNA at the MAT locus into the nonexpressed cassette. By hypothesizing a similar double-strand cleavage intermediate in either the trypanosome 5' or 3' recombination sequence, it is possible to draw identical gene conversion schemes for a switch from one trypanosome ELC to another. It is unlikely that the "telomere-like" break on the 3' side of all ELCs doubles as such an intermediate since, in the case of IaTat 1.2, the break is about 8 kb downstream from the termination codon, which seems unnecessarily far away. More likely a transient break would occur in one of the flanking repetitions close to the ELC gene as in yeast mating type. Certainly the similarities in the flanking sequence organization of the trypanosome VSG gene family and the yeast mating type suggest that there are common features in the mechanisms that the two organisms use to change the gene being transcribed.

The detection of a variable fragment with a hybridization probe that comes from predominantly outside the 5' recombination sequence was not anticipated (Fig. 5). As mentioned above, this variability is similar to that observed in Southern blots using cDNA probes for VSG isogenes that are expressed via the NDA mechanism rather than the ELC mechanism [20]. In this case the probe was a sequence adjacent to an ELC. In addition Xba I digestion converts the variable Pst I fragment to a constant-sized Pst I/Xba I fragment of 2.4 kb. It should be noted that the transposed ELC segment described here is also between Pst I and Xba I sites which are approximately 2.4 kb apart. This may imply that the sequences at the extremities of this new "variable-linked" 2.4-kb fragment are similar to those shown in Figure 3. This possibility is supported by preliminary sequence analysis of a recombinant clone of this 2.4 kb-fragment (data not shown). Clearly there are many questions about the relationship between these two genomic regions that cannot be answered by Southern hybridizations alone. For example, perhaps the "variable-linked" 2.4-kb fragment is the vestige of a previously expressed ELC which was not lost during the switch to expression of another VSG. This might suggest two or more potential "expression sites." Perhaps the variable-linked fragment merely contains a BC gene which is close to another telomere. To answer these questions this region must be examined more closely. These experiments are in progress. It is expected that they will provide more information about the first demonstration of a molecular connection between the two DNA rearrangement phenomena associated with VSG expression.

## ACKNOWLEDGMENTS

The authors thank Betsy Bricker for providing the clones of the IaTat serodeme and David Dorfman for performing some of the sequence analysis. This research was supported by public health service grants AI18954 and AM 21803.

## REFERENCES

1. Borst P, Cross GAM: *Cell* 29:291, 1982.
2. Pays E, Lheureux M, Steinert M: *Nature* 292:265, 1981.
3. Williams RO, Young JR, Majiwa PAO: *Nature* 299:417, 1982.
4. De Lange T, Borst P: *Nature* 299:451, 1982.
5. Cunningham MR, Vickerman K: *Trans R Soc Trop Med Hyg* 56:48, 1962.
6. Bricker B: "The Stability and Biochemical Characteristics of the Variant Surface Glycoproteins of *Trypanosoma brucei*." PhD thesis, Iowa City: University of Iowa.
7. Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ: *Biochemistry* 18:5294, 1979.
8. Aviv H, Leder P: *Proc Natl Acad Sci USA* 69:1408, 1972.
9. Land H, Grez M, Hauser H, Lindenmaier W, Schutz, G: *Nucleic Acids Res* 9:2251, 1981.
10. Rowekamp W, Firtel RA: *Dev Biol* 79:409, 1980.
11. St John T, Davis R: *Cell* 16:443, 1979.
12. Smith GE, Summers MD: *Anal Biochem* 109:123, 1980.
13. Rigby PWJ, Dieckmann M, Rhodes C, Berg P: *J Mol Biol* 113:237, 1977.
14. Maniatis T, Hardison RC, Lacy E, Lauer J, O'Connell C, Quon D, Sim GK, Efstratiadis A: *Cell* 15:687, 1978.
15. Maxam AM, Gilbert W: *Proc Natl Acad Sci USA* 74:560, 1977.
16. Hartley JL, Donelson JE: *Nature* 286:860, 1980.
17. Murphy WJ, Brentano ST, Rice-Ficht AC, Cain GD, Donelson JE: *J Mol Biol* (submitted).
18. Rice-Ficht AC, Chen KK, Donelson JE: *Nature* 294:53, 1981.

**12:JCB      Donelson et al**

19. Van der Ploeg LHT, Liu AYC, Michels PAM, De Lange T, Borst P, Majumder HK, Weber H, Veeneman GH, Van Boom J: *Nucleic Acids Res* 10:3591, 1982.
20. Williams RO, Young JR, Majiwa PAO: *Nature* 282:847, 1979.
21. Rice-Ficht AC, Chen KK, Donelson JE: *Nature* 298:676, 1982.
22. Bernards A, Van der Ploeg LHT, Frasch ACC, Borst P, Boothroyd JC, Coleman S, Cross GAM: *Cell* 27:497, 1981.
23. Pays E, Lheureux M, Steinert M: *Nucleic Acids Res* 9:4225, 1981.
24. Astel CR, Ahlstrom-Jonasson L, Smith M, Tatchell K, Nasmyth KA, Hall BD: *Cell* 27:15, 1981.
25. Strathern JN, Klar AJS, Hicks JB, Abraham JA, Ivy JM, Nasmyth KA, McGill C: *Cell* 31:183, 1982.
26. Smith HO: *Methods Enzymol* 65:371, 1980.
27. Grunstein M, Hogness D: *Proc Natl Acad Sci USA* 72:3961, 1975.
28. Boothroyd JC, Cross GAM: *Gene* 20:281, 1982.