MOLECULAR AND CELLULAR BIOLOGY, Dec. 1990, p. 6436–6444
0270-7306/90/126436-09$02.00/0

Vol. 10, No. 12

The Trypanosome Leucine Repeat Gene in the Variant Surface Glycoprotein Expression Site Encodes a Putative Metal-Binding Domain and a Region Resembling Protein-Binding Domains of Yeast, *Drosophila*, and Mammalian Proteins

B. L. SMILEY, A. W. STADNYK, P. J. MYLER, AND K. STUART*

Seattle Biomedical Research Institute, 4 Nickerson Street, Seattle, Washington 98109-1651

Received 7 June 1990/Accepted 18 September 1990

We have identified a new variant surface glycoprotein expression site-associated gene (ESAG) in *Trypanosoma brucei*, the trypanosome leucine repeat (T-LR) gene. Like most other ESAGs, it is expressed in a life cycle stage-specific manner. The N-terminal 20% of the predicted T-LR protein resembles the metal-binding domains of nucleic acid-binding proteins. The remainder is composed of leucine-rich repeats that are characteristic of protein-binding domains found in a variety of other eucaryote proteins. This is the first report of leucine-rich repeats and potential nucleic acid-binding domains on the same protein. The T-LR gene is adjacent to ESAG 4, which has homology to the catalytic domain of adenylate cyclase. This is intriguing, since yeast adenylate cyclase has a leucine-rich repeat regulatory domain. The leucine-rich repeat and putative metal-binding domains suggest a possible regulatory role that may involve adenylate cyclase activity or nucleic acid binding.

The genome of *Trypanosoma brucei* contains numerous variant surface glycoprotein (VSG) genes which are distributed among approximately 100 minichromosomes (50 to 150 kb) and about 10 larger chromosomes (45). The number of these chromosomes varies among stocks. Many VSG genes are present in tandem arrays at chromosomal internal sites (46), but VSG genes are also found adjacent to the telomeres of many, perhaps all, chromosomes (10, 25, 44, 49). Expression of the VSG genes is developmentally regulated, VSG being produced in life cycle stages occurring in the vertebrate host (bloodstream forms) but not in several stages occurring in the insect host except for the terminal (metacyclic) stage, which is infective to vertebrate hosts (3, 11). In addition, each trypanosome usually expresses only one of the numerous VSGs at a time. The VSG genes that are expressed are invariably at telomeric sites (4, 10, 24), apparently only on the larger chromosomes. The parasites also have the ability to change which VSG is produced (4). This results in antigenic variation, which permits evasion of the host immune defenses (3, 11). Antigenic variation occurs by switching the telomeric site that is expressed (25, 26, 50) or by recombination, usually a nonreciprocal gene conversion (25, 28), which changes the VSG-coding sequence in the expressed telomeric site.

The VSG gene is at the 3′ end of an approximately 50-kb region that contains at least seven other unique open reading frames (ORFs), referred to as expression site-associated genes (ESAGs) 1 to 7, whose expression is coordinated with that of the VSG gene (1, 9, 17, 30). Preliminary analyses of different expression sites indicates that the number of copies of these ESAGs can vary among expression sites, at least in different stocks, and that ESAGs 6 and 7 have substantial sequence homology. Most ESAG transcripts are present only in life cycle stages that produce VSG (30). It appears that ESAG expression is also VSG expression site specific, although this analysis is complicated by the presence of at least one copy from each ESAG family at each VSG expression site. Northern (RNA) blot analysis using an oligonucleotide probe specific for one ESAG 1 gene has shown that it is expressed in a variant antigen type (VAT)- and expression site-specific fashion (B. L. Smiley, J. K. Scholler, and K. Stuart, J. Cell Biol. **13E**:121, 1989). RNase protection studies using a probe that spans ESAGs 6 and 7 have also shown VAT- and expression site-specific expression (39), confirming their coordinate expression with the VSG gene. In addition, Northern blots using other ESAG probes usually give the most intense hybridization signal with RNA extracted from VATs expressing VSGs from the same expression site as that from which the probe is derived.

The subcellular locations and functions of the ESAG products are as yet unknown. Sequence analyses of ESAGs 1, 3, 4, 6, and 7 show N-terminal signal sequences (9, 30), suggesting export through the cellular membrane. A membrane location for ESAG 1 has been further substantiated, since antibodies prepared to recombinant ESAG 1 protein immunoprecipitate a glycoprotein that remains associated with the membrane fraction after osmotic lysis (8). The ESAG 4 predicted protein sequence contains a region that is homologous to the catalytic domain of adenylate cyclase of yeast cells (30).

We have identified another gene, the trypanosome leucine repeat (T-LR) gene, within the expression site of *T. brucei*, downstream of the ESAG 4 gene. Expression of this gene is developmentally regulated, possibly in a VAT-specific fashion. The gene encodes a predicted protein that is primarily composed of leucine-rich repeats, a putative metal-binding-like domain, and an associated basic region. The repeats have leucine residues at every second or third position plus a cysteine, and they have aliphatic and charged amino acids at specific positions. The characteristics of the repeat are similar to those of leucine-rich repeat domains that are present in the RAS-responsive regulatory domain of yeast adenylate cyclase (16), the lutropin-choriogonadotropin hor-

* Corresponding author.

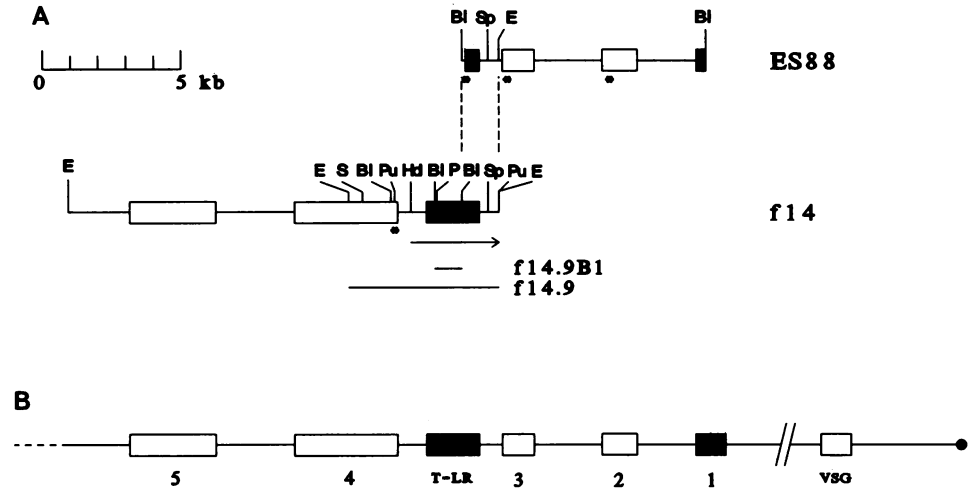


FIG. 1. (A) Alignment of two genomic clones containing the T-LR gene with an IsTAR 1 expression site. Symbols: ■, ESAG ORFs within regions that we have entirely sequenced; □, ESAGs identified by partial nucleotide sequence or hybridization analysis and comparison with other expression sites (17, 24, 30); * and →, regions that have been sequenced. (B) Composite map of an IsTaR 1 expression site generated from these and other clones and restriction mapping of genomic DNA. The f14 clone is from the T5 expression site (25, 39); the ES88 clone is from another expression site. Restriction sites are indicated only for the f14.9 subclone and corresponding region of ES88. Restriction enzyme abbreviations: Bl, *Bgl*II; E, *Eco*RI; Hd, *Hin*dIII; P, *Pst*I; Pu, *Pvu*II; S, *Sal*I; Sp, *Sph*I.

mone receptor that regulates adenylate cyclase activity via G-protein-mediated mechanisms (22), human placental RNase/angiogenin inhibitor (PRI) (18, 36), and several protein receptors that have been suggested to function in intercellular adhesion. The putative metal-binding/basic region domain near the N terminus resembles similar regions found in several nucleic acid-binding proteins.

MATERIALS AND METHODS

Organisms and DNA and RNA preparations. The derivation of VATs from the IsTaR 1 serodeme used in this study has been previously described (23–25, 40). A letter or number is usually used after the period to indicate the VAT; we use a superscript to indicate the previous VAT, if known (e.g., IsTat 1.A^7). For simplicity, we have omitted the IsTat 1. in this report. VAT A^{7}L was isolated from a relapse population of VAT 7 which has inactivated the telomere (T3) on which the expressed 7 VSG gene resides and now expresses the A VSG gene by telomere activation of the preexisting telomeric (T2) copy (25).

Bloodstream trypanosomes were grown in rats, and procyclic-stage trypanosomes were grown in supplemented SDM-79 culture medium. Trypanosome genomic DNA was prepared as previously described (23). RNA from either bloodstream or procyclic-stage trypanosomes was prepared by the guanidinium isothiocyanate method (6). Poly(A)$^+$-selected RNA was purified by oligo(dT)-cellulose chromatography (34) or by the Poly(A) Quik column (Stratagene), using the protocol and buffers supplied by the manufacturer.

Construction of clones. Chromosome-size DNA molecules from VAT 7 trypanosomes were separated by pulse-field gel electrophoresis (PFGE) as previously described (25). DNA from the M4 chromosome was electroeluted, sheared, methylated with *Eco*RI methylase, and ligated with *Eco*RI linkers. It was subsequently digested with *Eco*RI, size selected by agarose gel electrophoresis, and ligated into λ_{DASH} (Stratagene). Clone DTb1.7g-f14 was isolated by probing this genomic library with an expression site probe, pTg221.4 (17). A 5.5-kb *Eco*RI fragment from DTb1.7g-f14 was subcloned into pBluescript SK$^-$ (pTb1.7g-f14.9), and a 967-bp *Bgl*II fragment was further subcloned into pBS (pTb1.7g-f14.9B1). Clone pTb1.7g-ES88 was obtained by screening libraries prepared by ligation of 8- to 10-kb *Bgl*II fragments from VAT 7 genomic DNA into *Bam*HI-digested pBS$^-$ (Stratagene) with an oligonucleotide probe complementary to the 5′ coding region of the VAT 7 ESAG 1 gene (TTCACA AACAGCCTCCTCTCTCCCT). Abbreviated names f14, f14.9, f14.9B1, and ES88 will be used for simplicity. Restriction maps of these clones indicating the locations of ESAG coding regions are shown in Fig. 1.

Electrophoresis, hybridization, and sequencing. Genomic DNA (3 μg per lane) was digested with restriction enzymes according to the suppliers' instructions (Bethesda Research Laboratories, Inc., and New England BioLabs, Inc.), separated by agarose gel electrophoresis, and transferred to Nytran membranes as described previously (26) except that the DNA was cross-linked to the membranes by UV irradiation (7). The membranes were hybridized with riboprobes prepared from *Pst*I-digested f14.9B1, using T7 RNA polymerase (Stratagene) according to protocols provided by the manufacturer. Chromosome-size DNA molecules were separated by PFGE, transferred to Nytran membranes, and probed with nick-translated plasmid probes as described previously (37). Poly(A)$^+$ RNA (5 μg per lane) was separated on 2.2 M formaldehyde–1.2% agarose gels and transferred to Nytran membranes as previously described (12). Membranes were probed at 42°C as described previously (12, 26), using 10^6 dpm of ^{32}P-labeled nick-translated probe per ml, and washed twice for 30 min each time at 50°C in 1× SSPE–0.1% Sarkosyl, followed by two washes with 0.2× SSPE–0.05% Sarkosyl at the same temperature. Double-stranded DNA from plasmid clones was sequenced with the Sequenase kit (U.S. Biochemical Corp.), using the dideoxy-chain termination method (5, 41). Oligonucleotides complementary to the AnTat 1.3A ESAG 2 (CCACACGTCGCCAT ACACAG), ESAG 3 (CCATAACCCATCTAGGCCTC) (1),

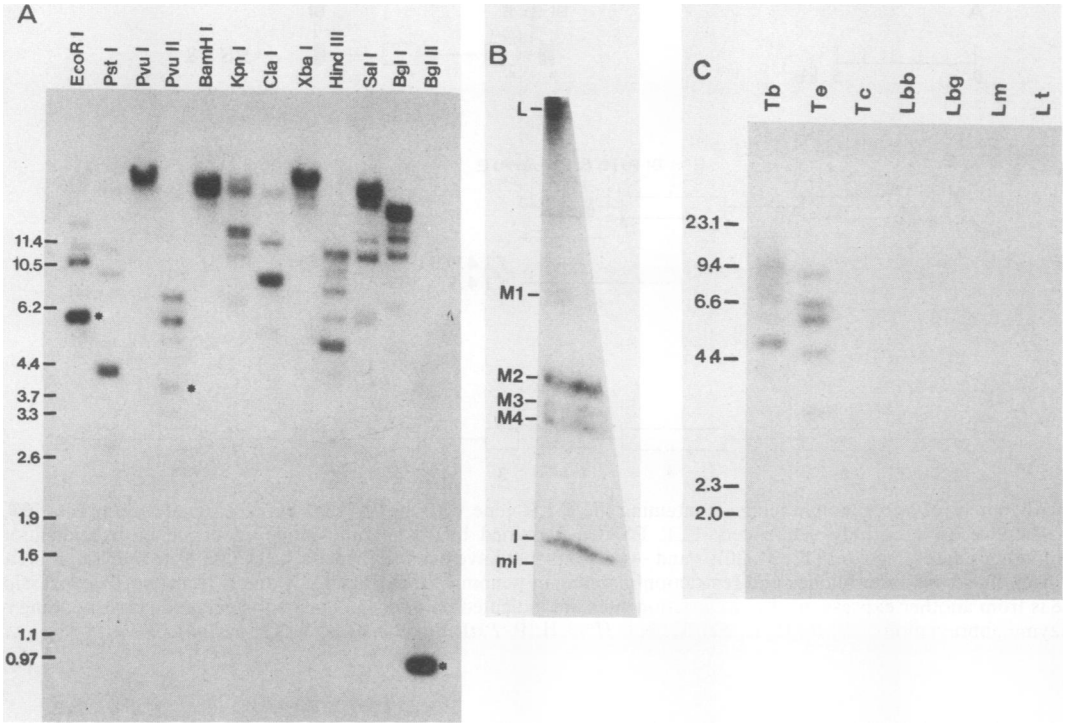


FIG. 2. Analysis of genomic copies of the T-LR sequence. (A) Genomic DNA was digested with the indicated restriction enzymes, electrophoresed, and transferred to Nytran. The filter was probed with ^{32}P-labeled riboprobe from f14.9B1. *, Fragment of the size expected from the genomic copy of f14. Size markers on the left are in kilobases. (B) Chromosome-size DNA from VAT 7 was separated by PFGE, transferred to Nytran, and probed with ^{32}P-labeled nick-translated f14.9B1. L, Large chromosome(s) (>3 Mb); M1 to M4, megabase-size chromosomes (>2, 1.7, 1.5, and 1.4 Mb, respectively); I, intermediate-size chromosomes (375 kb); mi, minichromosomes (50 to 150 kb). (C) *Hin*dIII-digested genomic DNA from each species was transferred to Nytran, and the filter was probed with ^{32}P-labeled nick-translated f14.9B1. Species: Tb, *T. brucei*; Te, *T. equiperdum*; Tc, *T. cruzi*; Lbb, *Leishmania braziliensis braziliensis* 675; Lbg, *Leishmania braziliensis guyanensis* CUMC1-1A; Lm, *Leishmania major*; Lt, *Leishmania tropica*.

ESAG 4 (GTCACCGAGGGGTCCATAC), and ESAG 5 (GC CTGTGCCGACCCACTGC) (30) were used as hybridization probes or for internal sequencing of ES88 and f14.9.

Computer analysis. Analysis of DNA and protein sequences was carried out by using DNASTAR (DNASTAR Inc., Madison, Wis.) and ESEE (E. Cabot, Simon Fraser University) software. A search of the Swiss-Prot data base, release 13.0, was performed by using the FASTA program (31) provided by GenBank.

Nucleotide sequence accession number. The GenBank accession number for the sequence of the T-LR open reading frame and flanking sequences in f14 is M38528.

RESULTS

Sequence analysis of genomic clones from VSG gene expression sites led to the identification of a previously unrecognized gene, the T-LR gene. This gene is located between ESAGs 3 and 4 in the IsTaR 1 T5 expression site, whose map (Fig. 1) was determined by restriction enzyme and hybridization analyses of M4 genomic DNA and expression site clones (25, 39; unpublished data). The locations of ESAGs 1, 2, and 3 in ES88 and of ESAGs 4 and 5 in f14 were determined by hybridization with specific oligonucleotides and nucleotide sequence analysis as diagrammed. T-LR sequence occurs in a region of overlap between clones f14 and ES88.

The T-LR gene is a member of a multicopy family of genomic sequences. An internal *Bgl*II fragment (subclone f14.9B1) hybridized to up to nine fragments in Southern blots of total genomic DNA, depending on the restriction enzyme used (Fig. 2A, *Hin*dIII lane). Considering the higher relative intensities of some bands, there are more than nine different copies of the T-LR gene within the *T. brucei* IsTaR 1 genome. A single intense 967-bp fragment was seen in *Bgl*II digests, indicating that the *Bgl*II sites are conserved in all copies of the T-LR gene. The restriction fragment polymorphisms seen with other enzyme digests may reflect greater sequence diversity in intergenic sequences outside the T-LR coding region. PFGE analysis showed that T-LR sequences occur in all megabase (L and M1 to M4) and intermediate-size chromosomes but not in minichromosomes (Fig. 2B). These results are consistent with the presence of a T-LR gene in each VSG gene expression site, but it is absent from minichromosomes, which contain only some VSG gene expression site-associated sequences (unpublished data). Hybridization of f14.9B1 to genomic DNA from a variety of trypanosomatid species (Fig. 2C) indicated that the T-LR gene was present in *T. brucei* and the closely related *Trypanosoma equiperdum* but not in *Trypanosoma cruzi* nor in *Leishmania* species. Thus, T-LR may be present only in salivarian trypanosomes.

Northern blot analysis showed the presence of a 2.2- to 2.4-kb T-LR transcript in bloodstream- but not procyclic-form RNA (Fig. 3). The f14.9B1 probe detected the tran-

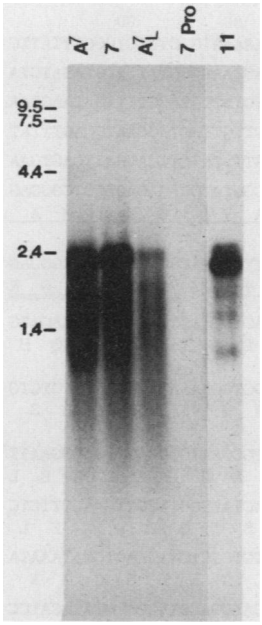


FIG. 3. Major steady-state transcripts of T-LR. Poly(A)$^+$ RNA from the VATs shown was separated on agarose gels, transferred to Nytran, and probed with nick-translated f14.9B1. Sizes determined from RNA molecular weight markers (Bethesda Research Laboratories) are shown at the left in kilobases.

script in poly(A)$^+$ RNA from VATs A^7, 7, A^7L, and 11 but not in RNA from procyclic forms derived from VAT 7 (7 Pro). T-LR gene transcripts were also detected in RNA from all other bloodstream VATs (3, 5, and 5^{A3}) tested (data not shown). The hybridization had a lower intensity in some VATs, multiple minor transcripts were seen in all VATs, especially VAT 11, and the size of the major transcript varied slightly between VATs. These results may reflect differences in RNA processing between VATs, possibly the utilization of different 5′ splice leader acceptor sites among different VATs. The variation in mRNA length between VATs is consistent with a VAT-specific expression of T-LR genes in different expression sites.

The 3,146-bp nucleotide sequence of the T-LR gene and its flanking sequences contains a 1,890-bp ORF (Fig. 4). The transcript observed in the Northern blots (Fig. 3) is of the size expected from this ORF, taking into account the splice leader, 5′ and 3′ untranslated sequences, and poly(A) tail. The ORF codes for a 630-amino-acid predicted protein with a molecular mass of 70 kDa and pI of 6.9. However, the protein shows an unequal charge distribution, with three domains of widely different pIs. The N-terminal 46 amino acids have a pI of 8.0, the next 68 have a pI of 11.8, while the remainder of the molecule has a pI of 5.1. The N-terminal amino acid sequence lacks the hydrophobic signal peptide sequence characteristic of proteins exported to the cell membrane, and the entire protein sequence lacks a membrane-spanning hydrophobic domain. Seven potential N-glycosylation sites are located in the protein sequence, although these are unlikely to be glycosylated if the protein remains internal. Clones f14 and ES88 must be derived from two different expression sites since T-LR DNA sequences from these clones have only 89% identity, and there is no evidence for duplication of the T-LR gene in chromosome M4 expression sites. Furthermore, hybridization and sequence analyses of the ESAG 1 gene from ES88 indicates that this clone is not from an M4 expression site (unpublished data). The predicted amino acid sequences have 84% identity and 11% conservative or neutral replacements.

The N-terminal region of the predicted T-LR protein contains a putative metal-binding domain with the form Cys x_2 Cys x_{13} Cys x His x_2 Cys x_2 Cys x_6 Cys x_2 Cys (Fig. 4). Two basic regions (containing five of eight and five of seven basic amino acids) are found C terminal to this metal-binding-like domain (double underlined in Fig. 4). This closely resembles the zinc finger or binuclear metal-binding and nucleic acid specificity domains of other proteins (2, 27).

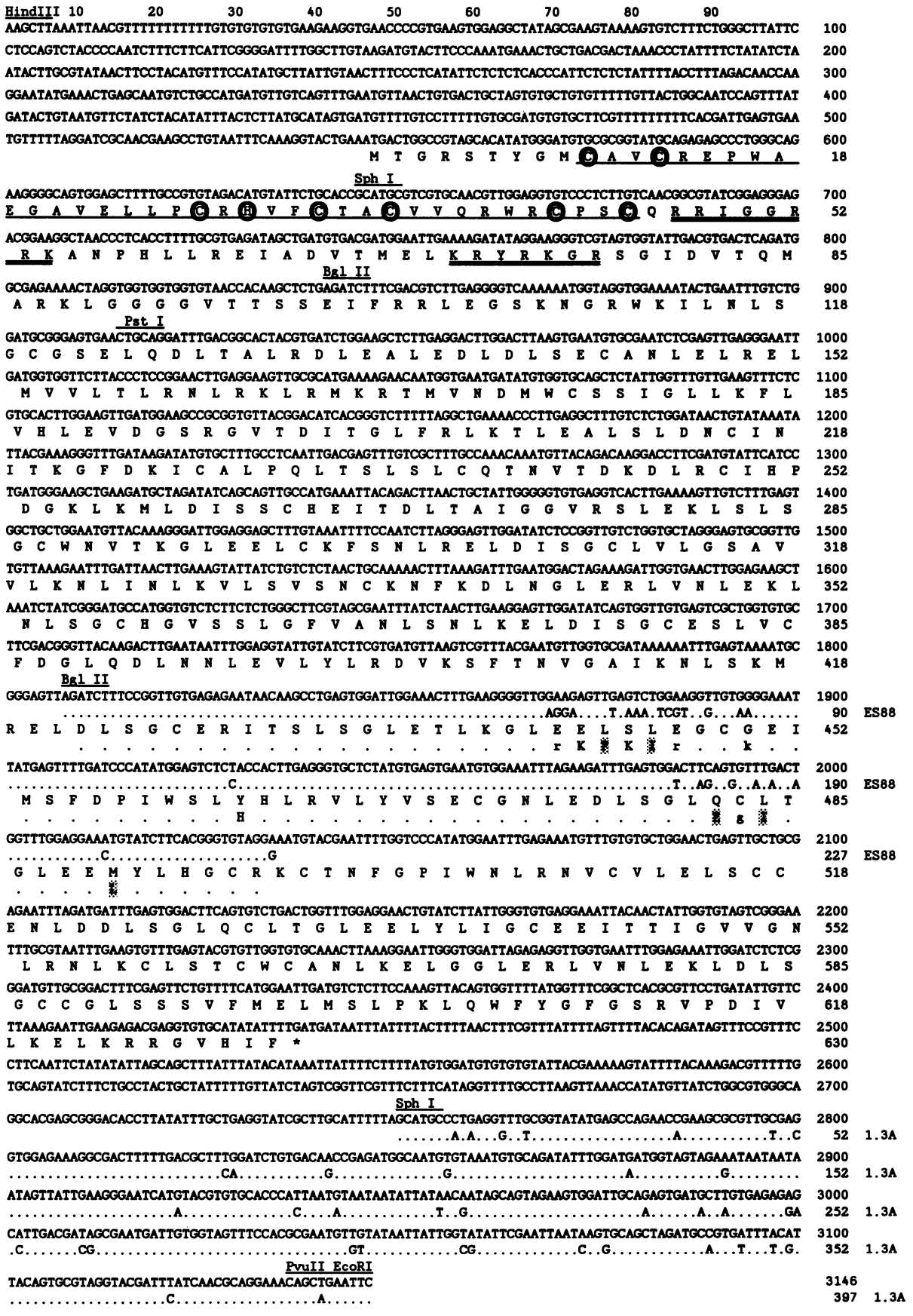
The remainder of the predicted protein sequence consists of 22 repeats of a leucine-rich motif with an average size of 23 amino acids, starting 115 amino acids from the N terminus and continuing to the C terminus of the molecule (Fig. 5A). The consensus sequence of the repeat (Fig. 5B) has leucine residues every second or third amino acid, a cysteine at position 21, a serine at position 19, and a glycine at position 20. An aliphatic amino acid is conserved at position 1, a charged amino acid is usually present at position 14, and an asparagine residue is often conserved at positions 9, 12, and 23.

DISCUSSION

We have identified a family of T-LR genes that are located between ESAGs 3 and 4 in *T. brucei* expression sites. There are more than nine copies of the T-LR genes, distributed among the larger and intermediate-size chromosomes. This is similar to the number and distribution of VSG gene expression sites, implying that each expression site has a T-LR gene. T-LR gene transcripts occur in all bloodstream-stage VATs examined but not in procyclic-stage VATs (which do not produce VSG). Differences in transcript number and size between VATs implies that different T-LR family members may be expressed in the different VATs, perhaps in coordination with the VSG gene, and may be the result of differences in splice leader acceptor or polyadenylation sites among the family members. All transcript sizes are consistent with the presence of the entire T-LR ORF along with limited 5′ and 3′ untranslated sequences and a poly(A) tail.

The C-terminal 82% of the protein predicted from the T-LR ORF is composed of 22 repeats of a 23-amino-acid sequence with a characteristic periodic occurrence of several amino acids, particularly leucine (Fig. 5A). While the predicted amino acid sequences differ between two copies of the T-LR gene (Fig. 4), the leucine-rich repeat character is retained. The consensus amino acid sequence of the repeat is very similar to that of several leucine-rich repeat domains that occur in different proteins in a variety of species (Fig. 5B and Table 1). Several features of the *T. brucei* T-LR consensus repeat distinguish it from the leucine-rich repeat domains of other proteins (Fig. 5B). The repeat units of other proteins are usually one to three amino acids longer and often have a conserved proline residue rather than the aliphatic amino acid at position 1 of the T-LR repeat. Other proteins (with the exception of PRI[b]) contain a conserved asparagine residue at position 21 rather than the cysteine found in T-LR. In contrast to the T-LR repeat, the other repeats generally do not contain conserved asparagine residues at positions 9, 12, and 23, and they lack the charged amino acids at position 14.

Conservation of the leucine-rich motif in such a diverse group of species implies that this domain has an important


HindIII 10 20 30 40 50 60 70 80 90
Sph I
Bgl II
Pst I
Bgl II
Sph I
PvuII EcoRI
ES88
ES88
1.3A
1.3A
1.3A
1.3A

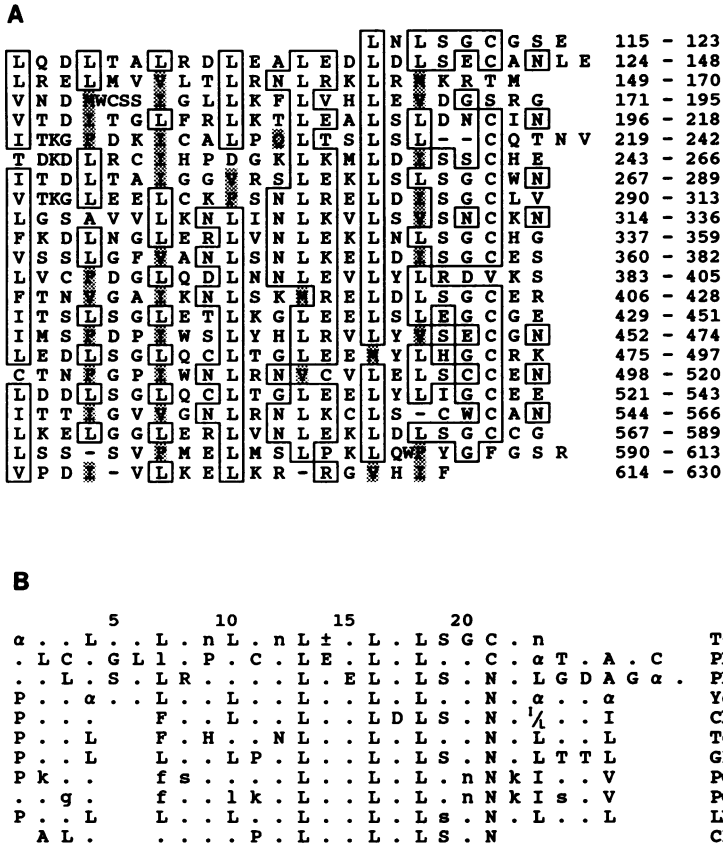


FIG. 5. Leucine-rich repeats of T-LR. (A) The repeat units are numbered with the first and last amino acids. Boxes surround identical matches to the T-LR consensus repeat; shading identifies conservative replacements based on the PAM 250 matrix used by the FASTA program (31). (B) The consensus leucine-rich repeats of other proteins are aligned according to the T-LR consensus. A dot indicates variation of amino acids at that position, a lowercase letter is used if less than 50% of the residues in the repeats match that sequence, ± indicates charged amino acids, and α indicates aliphatic amino acids. The sequence of PRI[a] refers to the first 28 amino acid module in each repeat, and PRI[b] is the second 29-amino-acid module of the PRI repeat. Protein abbreviations: PRI, placental RNase/aniogenin inhibitor; AdCy, adenylate cyclase; GPIbα, platelet membrane glycoprotein Ibα; PG I and PG II, bone proteoglycans I and II; LRG, serum leucine-rich α_2-glycoprotein.

function which is conserved in *T. brucei*. The chaoptin (33), Toll (15), bone proteoglycan I and II (14), and platelet membrane glycoprotein Ib (19, 20, 43) leucine-rich repeat domains appear to be arrayed on the external surface of the cells and have been suggested to function in intercellular adhesion. The serum leucine-rich α_2-glycoprotein (42) is a minor serum component whose function is not known. CD14 is found on the surface of myeloid cells and appears to be a receptor for the lipopolysaccharide-binding protein complex (38). The lutropin-choriogonadotropin receptor repeat domain has been suggested to function as a hormone receptor (22), and hormone binding may transduce a signal to the intracellular domain which regulates adenylate cyclase activity in a G-protein-mediated fashion. The *Saccharomyces cerevisiae* adenylate cyclase repeat domain binds RAS, which regulates adenylate cyclase catalytic activity (16). Human PRI (18, 36) binds to and inhibits RNases, possibly playing a role in regulation of cytoplasmic RNA levels. The emergent theme appears to be specific protein binding, with the effect dependent on the nature of the interacting proteins. Thus, it seems likely that the T-LR leucine-rich repeat is a binding domain that is specific for an intracellular protein, since it lacks a signal peptide, transmembrane domain, or hydrophobic C-terminal glycolipid anchor addition site.

The homology of T-LR to the leucine-rich repeat of *S. cerevisiae* adenylate cyclase is particularly intriguing since the *T. brucei* T-LR gene is immediately downstream of the ESAG 4 gene, which shows homology to the catalytic domain of adenylate cyclase (30). In *S. cerevisiae* adenylate cyclase, the leucine-rich repeat domain is involved in binding RAS protein and regulation of the catalytic activity of the enzyme (13). Thus, in *S. cerevisiae*, the leucine-rich repeat region and the adenylate cyclase catalytic domain are on the

FIG. 4. Nucleotide and amino acid sequences of T-LR. The DNA sequence from the *Hin*dIII site of f14.9 is shown along with the translation of the T-LR ORF. The comparison with 227 bp of ES88 and its ORF starts at the ES88 *Bgl*II site; a dot indicates an exact match, and the single-letter code indicates a difference. For the amino acid comparison, a conserved change is shaded, a neutral change is in uppercase, and other changes are in lowercase. The putative metal-binding region is underlined, the cysteine and histidine residues in this region are circled, and the basic regions are double underlined. Restriction sites are overlined.

TABLE 1. Comparison of various leucine repeat domains

Protein (source)[a]	N-terminal signal peptide	Transmembrane-spanning domain	Copies	Reference(s)	Function
Chaoptin (*Drosophila* sp.)	+	+	41	33	Intercellular adhesion, photoreceptor cell alignment
Toll (*Drosophila* sp.)	+	+	15	15	Intercellular adhesion, establishes dorsal-ventral axis
GPIbα (human)	+	+	7	20, 43	Intercellular adhesion, to endothelial cells binds thrombin and von Willebrand factor
GPIbβ (human)	+	+	1	19	Same as GPIbα
PG I and II (human)	+	?	12, 10	14	Binds collagen and fibronectin
CD14 (mouse)	+	−	13	38	Receptor for lipopolysaccharide-bound lipopolysaccharide-binding protein[b]
LH-CG-R (human)	+	+	14	22	Binds hormones lutropin and choriogonadotropin, G-protein-mediated regulation of adenylate cyclase activity
AdCy (*Saccharomyces cerevisiae*)	−	−	26	16	Binds RAS, regulating adenylate cyclase activity
AdCy (*Schizosaccharomyces pombe*)	−	−	17	51	Unknown
PRI (human)	−	−	7[c]	18, 36	Binds and inhibits RNase and angiogenin
LRG (human)	+[d]	?	9	42	Unknown
T-LR (*Trypanosoma brucei*)	−	−	22		Unknown

[a] LH-CG-R, Lutropin-choriogonadotropin receptor. For other abbreviations, see legend to Fig. 5B.
[b] S. D. Wright, personal communication.
[c] Seven modules each containing two copies of differing forms of the repeat.
[d] Assumed from extracellular location.

same molecule, while in *T. brucei*, the leucine-rich repeat and adenylate cyclase catalytic domain analog are on separate polypeptides. A requirement for different patterns of production of a protein-binding domain and catalytic domain during the life cycle may be the basis for the existence of T-LR and ESAG 4 in two genes. While the T-LR gene is expressed in bloodstream but not procyclic forms, an ESAG 4 homologous gene is expressed in both stages, suggesting that adenylate cyclase activity is required in both stages of the life cycle whereas T-LR function is not. The apparent involvement of cyclic AMP in the slender-to-stumpy (nondividing) differentiation in bloodstream forms during the life cycle of *T. brucei* (21, 32) may reflect a role of the T-LR and ESAG 4 gene products in this process, possibly involving a RAS cognate. The T-LR repeat also contains clusters of aspartate and glutamate residues interspersed with hydrophobic residues that have been implicated in calcium binding (35), which is intriguing since adenylate cyclase activity is regulated by calcium in *T. brucei* (48).

The T-LR protein contains a cysteine-rich region near its N terminus, adjacent to a basic domain. This region closely resembles zinc finger or binuclear cluster metal-binding domains of other proteins that bind nucleic acids in a sequence-specific, metal-dependent fashion and regulate gene expression (27; for a review see reference 2). The presence of a putative metal-binding domain implies a role for the T-LR gene product that entails nucleic acid binding, suggesting that it may be involved in regulating gene expression at either the DNA or RNA level.

The cysteine at position 21 and charged residue at position 14 of the T-LR and PRI[b] leucine-rich repeats differ from the other leucine-rich repeat sequences (Fig. 5B). This is particularly interesting since the leucine-rich repeat of PRI inhibits RNase activity, and VSG and ESAG transcript abundance is regulated in a stage-specific fashion, perhaps posttranscriptionally (29). This raises the possibility that T-LR is involved in the posttranscriptional regulation of genes in VSG expression sites. The restriction of the T-LR gene to those trypanosomes which undergo antigenic variation (Fig. 2C) and the presence of T-LR mRNA in bloodstream but not procyclic forms are consistent with such a role. A potential nucleic acid-binding regulatory function and a possible adenylate cyclase-dependent regulatory role are not necessarily mutually exclusive possibilities. However, an association with adenylate cyclase implies a location near the plasma membrane, since adenylate activity has been found in *T. brucei* membrane fractions (47), while a role entailing interaction with nucleic acids implies a nuclear location. The role of T-LR, and why it is associated with other ESAGs and VSG genes as part of a single transcription unit (30), is a topic for future experimentation.

ACKNOWLEDGMENTS

We thank Andrea Perrollaz, Nicole Nelson, and Lin Bennett for technical assistance and Harry Charboneau (University of Washington) for helpful discussions.

This work was supported by Public Health Service grant AI 17375 from the National Institutes of Health. K.S. is a Burroughs Wellcome Scholar in Molecular Parasitology.

LITERATURE CITED

1. **Alexandre, S., M. Guyaux, N. B. Murphy, H. Coquelet, A. Pays, M. Steinert, and E. Pays.** 1988. Putative genes of a variant-specific antigen gene transcription unit in *Trypanosoma brucei*. Mol. Cell. Biol. **8**:2367–2378.
2. **Berg, J. M.** 1990. Zinc fingers and other metal-binding domains. J. Biol. Chem. **265**:6513–6516.
3. **Boothroyd, J. C.** 1985. Antigenic variation in african trypanosomes. Annu. Rev. Microbiol. **39**:475–502.
4. **Borst, P., and G. A. M. Cross.** 1982. Molecular basis for trypanosome antigenic variation. Cell **29**:291–303.
5. **Chen, E. J., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**:165–170.
6. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**:156–159.
7. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.
8. **Cully, D. F., C. P. Gibbs, and G. A. M. Cross.** 1986. Identification of proteins encoded by variant surface glycoprotein expression site-associated genes in *Trypanosoma brucei*. Mol. Biochem. Parasitol. **21**:189–197.

9. **Cully, D. F., H. S. Ip, and G. A. Cross.** 1985. Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene family in *Trypanosoma brucei*. Cell **42**:173–182.
10. **De Lange, T., and P. Borst.** 1982. Genomic environment of the expression-linked extra copies of genes for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. Nature (London) **299**:451–453.
11. **Donelson, J. E., and A. C. Rice-Ficht.** 1985. Molecular biology of trypanosome antigenic variation. Microbiol. Rev. **49**:107–125.
12. **Feagin, J. E., and K. Stuart.** 1985. Differential expression of mitochondrial genes between life cycle stages of *Trypanosoma brucei*. Proc. Natl. Acad. Sci. USA **82**:3380–3384.
13. **Field, J., H. Xu, T. Michaeli, R. Ballester, P. Sass, M. Wigler, and J. Colicelli.** 1990. Mutations of the adenylyl cyclase gene that block RAS function in *Saccharomyces cerevisiae*. Science **247**:464–467.
14. **Fisher, L. W., J. D. Termine, and M. F. Young.** 1989. Deduced protein sequence of bone small proteoglycan I (biglycan) shows homology with proteoglycan II (decorin) and several nonconnective tissue proteins in a variety of species. J. Biol. Chem. **264**:4571–4576.
15. **Hashimoto, C., K. L. Hudson, and K. V. Anderson.** 1988. The *Toll* gene of Drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. Cell **52**:269–279.
16. **Kataoka, T., D. Broek, and M. Wigler.** 1985. DNA sequence and characterization of the S. cerevisiae gene encoding adenylate cyclase. Cell **43**:493–505.
17. **Kooter, J. M., H. J. van der Spek, R. Wagter, C. E. d'Oliveira, F. van der Hoeven, P. J. Johnson, and P. Borst.** 1987. The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in T. brucei. Cell **51**:261–272.
18. **Lee, F. S., E. A. Fox, H.-M. Zhou, D. J. Strydom, and B. L. Vallee.** 1988. Primary structure of human placental ribonuclease inhibitor. Biochem. **27**:8545–8553.
19. **Lopez, J. A., D. W. Chung, K. Fujikawa, F. S. Hagen, E. W. Davie, and G. J. Roth.** 1988. The α and β chains of human platelet glycoprotein Ib are both transmembrane proteins containing a leucine-rich amino acid sequence. Proc. Natl. Acad. Sci. USA **85**:2135–2139.
20. **Lopez, J. A., D. W. Chung, K. Fujikawa, F. S. Hagen, T. Papayannopoulou, and G. J. Roth.** 1987. Cloning of the α chain of human platelet glycoprotein Ib: a transmembrane protein with homology to leucine-rich α_2-glycoprotein. Proc. Natl. Acad. Sci. USA **84**:5615–5619.
21. **Mancini, P. E., and C. L. Patton.** 1981. Cyclic 3′,5′-adenosine monophosphate levels during the development cycle of *Trypanosoma brucei*. Mol. Biochem. Parasitol. **3**:19–31.
22. **McFarland, K. C., R. Sprengel, H. S. Phillips, M. Kohler, N. Rosemblit, K. Nikolics, D. L. Segaloff, and P. H. Seeburg.** 1989. Lutropin-choriogonadotropin receptor: an unusual member of the G protein-coupled receptor family. Science **245**:494–499.
23. **Milhausen, M., R. G. Nelson, M. Parsons, G. Newport, K. Stuart, and N. Agabian.** 1983. Molecular characterization of initial variants from the IsTat I serodeme of *Trypanosoma brucei*. Mol. Biochem. Parasitol. **9**:241–254.
24. **Myler, P., R. Nelson, N. Agabian, and K. Stuart.** 1984. Two mechanisms of expression of a predominant variant antigen gene of *Trypanosoma brucei*. Nature (London) **309**:282–284.
25. **Myler, P. J., R. F. Aline, Jr., J. K. Scholler, and K. D. Stuart.** 1988. Multiple events associated with antigenic switching in *Trypanosoma brucei*. Mol. Biochem. Parasitol. **29**:227–241.
26. **Myler, P. J., J. Allison, N. Agabian, and K. Stuart.** 1984. Antigenic variation in African trypanosomes by gene replacement or expression of alternate telomeres. Cell **39**:203–211.
27. **Pan, T., and J. E. Coleman.** 1990. GAL4 transcription factor is not a "zinc finger" but forms a $Zn(II)_2Cys_6$ binuclear cluster. Proc. Natl. Acad. Sci. USA **87**:2077–2081.
28. **Pays, E.** 1985. Gene conversion in trypanosome antigenic variation. Prog. Nucleic Acid Res. Mol. Biol. **32**:1–26.
29. **Pays, E., H. Coquelet, A. Pays, P. Tebabi, and M. Steinert.** 1989. *Trypanosoma brucei*: posttranscriptional control of the variable surface glycoprotein gene expression site. Mol. Cell. Biol. **9**:4018–4021.
30. **Pays, E., P. Tebabi, A. Pays, H. Coquelet, P. Revelard, D. Salmon, and M. Steinert.** 1989. The genes and transcripts of an antigen gene expression site from T. brucei. Cell **57**:835–845.
31. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
32. **Reed, S. L., A. S. Fierer, D. R. Goddard, M. E. M. Colmerauer, and C. E. Davis.** 1985. Effect of theophylline on differentiation of *Trypanosoma brucei*. Infect. Immun. **49**:844–847.
33. **Reinke, R., D. E. Krantz, D. Yen, and S. L. Zipursky.** 1988. Chaoptin, a cell surface glycoprotein required for Drosophila photoreceptor cell morphogenesis, contains a repeat motif found in yeast and human. Cell **52**:291–301.
34. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **Sambrook, J. F.** 1990. The involvement of calcium in transport of secretory proteins from the endoplasmic reticulum. Cell **61**:197–199.
36. **Schneider, R., E. Schneider-Scherzer, M. Thurnher, B. Auer, and M. Schweiger.** 1988. The primary structure of human ribonuclease/angiogenin inhibitor (RAI) discloses a novel highly diversified protein superfamily with a common repetitive module. EMBO J. **7**:4151–4156.
37. **Scholler, J. K., S. G. Reed, and K. Stuart.** 1986. Molecular karyotype of species and subspecies of *Leishmania*. Mol. Biochem. Parasitol. **20**:279–293.
38. **Setoguchi, M., N. Nasu, S. Yoshida, Y. Higuchi, S. Akizuki, and S. Yamamoto.** 1989. Mouse and human CD14 (myeloid cell-specific leucine-rich glycoprotein) primary structure deduced from cDNA clones. Biochim. Biophys. Acta **1008**:213–222.
39. **Stadnyk, A. W., J. K. Scholler, P. J. Myler, and K. D. Stuart.** 1990. Ribonuclease protection determines sequences specific to a single variable surface glycoprotein gene expression site, p. 99–109. *In* N. Agabian and A. Cerami (ed.), Parasites: molecular biology, drug and vaccine design. Alan R. Liss, Inc., New York.
40. **Stuart, K., E. Gobright, L. Jenni, M. Milhausen, L. Thomashow, and N. Agabian.** 1984. The IsTaR 1 serodeme of *Trypanosoma brucei*: development of a new serodeme. J. Parasitol. **70**:747–754.
41. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
42. **Takahashi, N., Y. Takahashi, and F. W. Putnam.** 1985. Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich α_2-glycoprotein of human serum. Proc. Natl. Acad. Sci. USA **82**:1906–1910.
43. **Titani, K., K. Takio, M. Handa, and Z. M. Ruggeri.** 1987. Amino acid sequence of the von Willebrand factor-binding domain of platelet membrane glycoprotein Ib. Proc. Natl. Acad. Sci. USA **84**:5610–5614.
44. **Van der Ploeg, L. H. T., A. Bernards, F. A. Rijsewijk, and P. Borst.** 1982. Characterization of the DNA duplication-transposition that controls the expression of two genes for variant surface glycoproteins in *Trypanosoma brucei*. Nucleic Acids Res. **10**:593–609.
45. **Van der Ploeg, L. H. T., C. L. Smith, R. I. Polvere, and K. M. Gottesdiener.** 1989. Improved separation of chromosome-sized DNA from *Trypanosoma brucei*, stock 427-60. Nucleic Acids Res. **17**:3217–3228.
46. **Van der Ploeg, L. H. T., D. Valerio, T. De Lange, A. Bernards, P. Borst, and F. G. Grosveld.** 1982. An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. Nucleic Acids Res. **10**:5905–5923.
47. **Voorheis, H. P., J. S. Gale, M. J. Owen, and W. Edwards.** 1979. The isolation and partial characterization of the plasma membrane from *Trypanosoma brucei*. Biochem. J. **180**:11–24.
48. **Voorheis, H. P., and B. R. Martin.** 1980. 'Swell dialysis' demonstrates that adenylate cyclase in *Trypanosoma brucei* is

regulated by calcium ions. Eur. J. Biochem. **113**:223–227.
49. **Williams, R. O., J. R. Young, and P. A. Majiwa.** 1982. Genomic environment of *T. brucei* VSG genes: presence of a minichromosome. Nature (London) **299**:417–421.
50. **Williams, R. O., J. R. Young, and P. A. O. Majiwa.** 1979. Genomic rearrangements correlated with antigenic variation in *Trypanosoma brucei*. Nature (London) **282**:847–849.
51. **Yamawaki-Kataoka, Y., T. Tamaoki, C. Hye-Ryun, H. Tanaka, and T. Kataoka.** 1989. Adenylate cyclases in yeast: a comparison of the genes from *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **86**:5693–5697.