MOLECULAR AND CELLULAR BIOLOGY, Mar. 1989, p. 1332–1335
0270-7306/89/031332-04$02.00/0

Vol. 9, No. 3

Variation of Tandem Repeats in the Developmentally Regulated Procyclic Acidic Repetitive Proteins of *Trypanosoma brucei*

MICHAEL R. MOWATT,†* GREGORY S. WISDOM, AND CHRISTINE E. CLAYTON
The Rockefeller University, 1230 York Avenue, New York, New York 10021

Received 26 September 1988/Accepted 28 November 1988

The procyclic acidic repetitive proteins (PARPs) of *Trypanosoma brucei* are developmentally regulated surface proteins encoded by a family of polymorphic genes. We have determined the complete nucleotide sequence of a novel member of the PARP gene family and investigated its expression. The amino acid sequence deduced from the parpAα gene showed a marked conservation of both the amino- and carboxy-terminal regions compared with other PARPs but revealed the substitution of a pentapeptide for the dipeptide repeating unit that is characteristic of all other PARPs. Northern hybridization analysis indicated that expression of the parpAα gene, like that of other members of this gene family, is confined to the procyclic stage of the *T. brucei* life cycle. This result implies coordinate regulation of the unlinked genetic loci that encode PARPs.

African trypanosomes are unicellular parasitic protozoa that undergo several developmental transitions in the course of transmission between mammalian hosts by an insect vector, the tsetse fly (genus *Glossina*). One such transition is the transformation of infectious bloodstream forms to noninfectious procyclic forms in the midgut of the fly after uptake in a blood meal (reviewed in reference 18). *Trypanosoma brucei* bloodstream-form trypomastigotes readily differentiate to procyclic culture trypomastigotes (PCTs) in vitro (3). We previously identified a family of genes that are strongly induced during procyclic differentiation and encode a set of polymorphic proteins (11, 12) situated on the parasite surface (J. P. Richardson, R. P. Beecroft, D. L. Tolson, M. K. Liu, and T. W. Pearson, Mol. Biochem. Parasitol., in press). The proteins are characterized by a central domain composed of a variable number of tandemly repeated glutamic acid-proline (EP) dipeptides; on the basis of this unusual composition, we adopted the acronym PARP for these procyclic acidic repetitive proteins (11). This procyclic surface protein, first characterized biochemically by Richardson et al. (in press), is also referred to as procyclin (Richardson et al., in press; I. Roditi, H. Schwarz, T. W. Pearson, R. P. Beecroft, M. K. Liu, J. P. Richardson, H.-J. Buhring, J. Pleiss, R. Bulow, R. O. Williams, and P. Overath, submitted for publication).

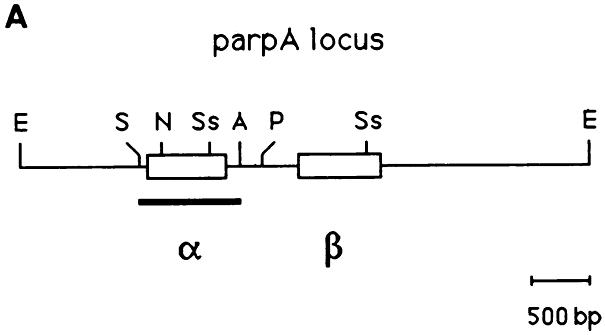
In *T. brucei*, PARPs are encoded at two distinct genetic loci, which we have termed parpA and parpB (11). Each locus consists of a pair of genes, α and β (Fig. 1A and 2). Hybridization analysis of cloned genomic PARP sequences revealed that the genes of two allelic parpB loci, parpB-1 and parpB-2, were characterized by EP repeat regions of different sizes (12). Similar analysis of the cloned parpA locus showed that, although the Aβ gene had an EP repeat region equal in size to that of the B-2β gene, the Aα gene was not detected by a nucleotide probe specific for the EP repeat region. This observation, in conjunction with other anomalous hybridization results, prompted an examination of the Aα gene structure in its finest detail.

Structure of the parpAα gene. We used the full-length PARP cDNA B-1α[667] (a cDNA clone derived from a parpB-1α gene transcript previously designated as α-667 [12]) as a probe to localize the Aα gene to an 850-base-pair (bp) *Sca*I-*Ase*I fragment of the genomic parpA clone (Fig. 1A). The complete nucleotide sequence of this fragment was determined by directed subcloning of overlapping restriction fragments in M13mp18 (19) followed by chain termination sequence analysis (1, 16) by using Sequenase (United States Biochemical Corp., Cleveland, Ohio) as described by the supplier. On average, the identity of a given nucleotide was determined four times, i.e., twice on each strand. The sequence is shown in Fig. 1B. By comparison with the B-1α[667] cDNA sequence, the Aα coding region extends from nucleotide (nt) 1 through 345 and constitutes the longest open reading frame on either strand of the 850-bp fragment. The Aα gene encodes a protein of 114 amino acids with a molecular weight of 11,610. The Aα polypeptide shares several structural features with other PARPs (discussed below) but is most notably distinguished from them by the substitution of the pentapeptide glycine-proline-glutamic acid-glutamic acid-threonine (GPEET) for the EP dipeptide as the repeating unit. By analogy to the B-1α[667] cDNA, the addition of the mini-exon, a 39-nt leader RNA spliced to all *T. brucei* mRNAs (9; reviewed in references 2, 4, and 17), is likely to occur after the dinucleotide AG (nt −32). Mini-exon addition at this site would leave a 31-nt 5′ untranslated region that is identical in sequence to that of the B-1α[667] cDNA as well as the parpAβ, B-2α, and B-2β genomic sequences determined previously (11). Finally, and again by analogy to the B-1α[667] cDNA, polyadenylation is likely to occur after nt 644 to yield a 3′ untranslated region that is 86% identical to the 3′ untranslated region of the B-1α[667] cDNA but only 40% identical to that of a β-427 cDNA (14). Mini-exon addition and polyadenylation at the sites specified above would yield Aα mRNA approximately 900 nt long.

GPEET repeat is unique to the parpAα gene. The substitution of GPEET for EP as the repeating unit in the Aα protein explains the failure of previous attempts to detect the Aα gene by hybridization with an EP repeat nucleotide probe (12). To assess the suitability of a GPEET nucleotide probe for use in the study of Aα gene expression, we tested an 85-bp *Hae*III fragment derived from the Aα repeat region in Southern hybridization analyses of the cloned PARP

* Corresponding author.
† Present address: Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, Building 4, Room B1-31, Bethesda, MD 20892.

A



B

```
AGTACTGAGTTTAACATGTTCTCGTGATCGCTGCACGCGCCTTCGAGTTTTTTTCCTTTTCCCCATTTTT
             *               *               *
           -100             -80             -60

                                         Met Ala Pro Arg Ser Leu
TTCAACTTGAAGACTTCAATTACACCAAAAAGTAAAATTCACA    ATG GCA CCT CGT TCC CTT
   *               *                       *
 -40             -20                       1

Tyr Leu Leu Ala Ile Leu Leu Phe Ser Ala Asn Leu Phe Ala Gly Val Gly Phe
TAT CTG CTC GCT ATT CTT CTG TTC AGC GCG AAC CTC TTC GCT GGC GTG GGA TTT
 *                          *                     *
20                         40                    60

Ala Ala Ala Ala Asp Glu Ser Ala Ser Asn Val Ile Val Lys Gly Gly Lys Gly
GCC GCA GCC GCT GAT GAG TCG GCT AGC AAC GTT ATC GTG AAG GGA GGC AAA GGA
         *                          *                         *
        80                         100                       120

Lys Glu Arg Glu Asp Gly Pro Glu Glu Pro Glu Glu Thr Gly Pro Glu Glu Thr
AAG GAG AGG GAG GAC GGC CCT GAG GAG CCG GAA GAG ACC GGA CCA GAA GAG ACC
                 *                          *                         *
                140                        160                       180

Gly Pro Glu Glu Thr Gly Pro Glu Glu Thr Gly Pro Glu Glu Thr Gly Pro Glu
GGA CCA GAA GAA ACC GGA CCA GAA GAA ACC GGA CCA GAA GAG ACG GGG CCG GAA
                         *                          *
                        200                        220

Glu Thr Gly Pro Glu Glu Thr Glu Pro Glu Pro Glu Pro Gly Ala Ala Thr Leu
GAG ACG GGA CCA GAG GAA ACT GAA CCT GAA CCT GAA CCT GGT GCT GCA ACG CTG
     *                          *                          *
    240                        260                        280

Lys Ser Val Ala Leu Pro Phe Ala Val Ala Ala Ala Ala Leu Val Ala Ala Phe
AAA TCT GTT GCA CTT CCG TTT GCA GTC GCG GCT GCT GCT CTC GTT GCC GCA TTC
             *                          *                         *
            300                        320                       340

---
TAA GCGGATGCAAGCGTGTAAAGCGCCTCGGAGGAACGAAACCCCTTTGAAAAGGTGGCTTTCGTTTATA
                 *                  *                  *
                360                380                400

TATTCTCCATCTGGTGCATCGTGTTTATTTCGTATTATTCCCTGTCAAAAGCTGCGGATATTCATTTAATA
      *                  *                  *                  *
     420                440                460                480

TTCTTTTCGTTATATATTTTTAGTTCTATCCTTTATTTTATCCGTTTTAATCGTCTTCTGAGACCCACAGC
              *                  *                  *
             500                520                540

CCTGTAGATTTCTGTGATGTTTCGGTTGCGTATTCCATAATTTTAAGCGTTTCACTTCTATTTTTTTTCAT
     *                  *                  *                  *
    560                580                600                620

TCCTTTGAATTTGGATCTTAAAATTATCATTGGTGCCTTGTGTTATTGTGCGTGCTGCGTGTGAATGTGGC
             *                  *                  *
            640                660                680

GCTCTGCCTTTTAATTTGTTGGATGAGCTATTTCATTAAT
   *                  *
  700                720
```

FIG. 1. Structure of the parpAα gene . (A) Restriction map of the parpA locus, cloned from *T. brucei* 427 (11), with the approximate locations of the α and β genes indicated by open boxes. The bar indicates the 850-bp *ScaI-AseI* fragment chosen for sequence analysis. Restriction sites: A, *AseI*; E, *EcoRI*; N, *NheI*; P, *PvuII*; S, *ScaI*; Ss, *SspI*. (B) Nucleotide sequence of the *ScaI-AseI* fragment and the deduced amino acid sequence of the Aα polypeptide. Nucleotides are numbered consecutively, with position 1 corresponding to the first nucleotide of the initiation codon.

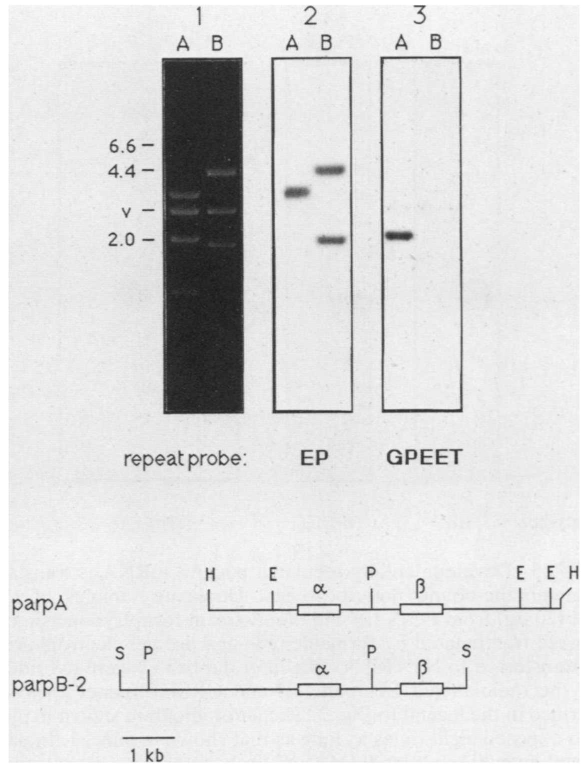


FIG. 2. GPEET nucleotide probe uniquely detects the parpAα gene. The two recombinant plasmids used in this analysis, parpA, derived from *T. brucei* strain 427, and parpB-2, derived from strain TREU 667 BUT55, have been described previously (12). Two micrograms of plasmids parpA (lanes A) and parpB-2 (lanes B) was digested to completion with *EcoRI-PvuII* and *SalI-PvuII*, respectively. Samples (400 ng) of each digestion were loaded pairwise and in duplicate on a 1% agarose gel, fractionated electrophoretically, and transferred to NYTRAN (Schleicher & Schuell, Inc., Keene, N.H.). The filters were trimmed appropriately and hybridized with radiolabeled DNA fragments corresponding to the PARP EP repeat (panel 2) or the Aα GPEET repeat (panel 3). The 171-bp EP repeat probe was described previously (12) and was labeled by the method of Feinberg and Vogelstein (7). The GPEET repeat probe corresponds to Aα nt 144 through 228 (Fig. 1B) cloned in M13mp18; a radiolabeled antisense probe was synthesized by extension of the universal sequencing primer in the presence of [^{32}P]dCTP. Panel 1 shows an ethidium bromide-stained gel section before transfer. The autoradiogram shown in panel 3 was exposed twice as long as that shown in panel 2. Lambda *HindIII* size markers are indicated (in kilobases), as is the linear plasmid vector (v) at 2.8 kilobases. The restriction maps of the parpA and parpB-2 plasmid inserts are shown below with the approximate locations of the α and β genes indicated by open boxes.

genes (Fig. 2). In contrast to the EP repeat probe, which detects the parpAβ, B-2α, and B-2β genes (in addition to the parpB-1α and B-1β genes; data not shown), the GPEET repeat probe uniquely detects the Aα gene. Genomic Southern hybridization analysis confirmed this result, proved that all GPEET sequences in the trypanosome genome map to the Aα gene, and demonstrated that the GPEET probe recognizes the Aα genes of the two independent *T. brucei* isolates 427 and TREU 667 BUT55 (data not shown).

Developmental regulation of parpAα gene expression. To assess the pattern of expression of parpAα in the trypanosome life cycle, we compared EP and GPEET probe hybridizations in analyses of RNA prepared from *T. brucei* 427 (variant 117) bloodstream-form trypomastigotes and PCTs

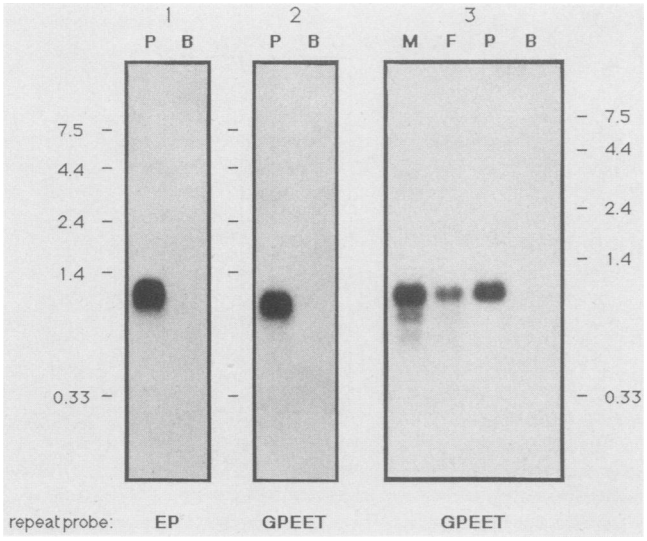


FIG. 3. Developmentally regulated parpAα mRNA is translated on membrane-bound polyribosomes. Duplicate samples of total RNA (20 μg) from PCTs (P) and bloodstream-form trypomastigotes (B) were fractionated by formaldehyde-agarose gel electrophoresis and transferred to NYTRAN. The filter duplicates were hybridized with the radiolabeled EP (panel 1) and GPEET (panel 2) probes described in the legend to Fig. 2. The autoradiogram shown in panel 2 was exposed eight times as long as that shown in panel 1. In panel 3, total membrane-bound (M) and free cytosolic (F) polysomal RNAs from procyclic trypanosomes are compared with total procyclic (P) and bloodstream-form (B) RNAs (5 μg of each) as described previously (11) but by using the radiolabeled GPEET probe. The sizes of RNA ladder markers (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) are indicated in kilobases.

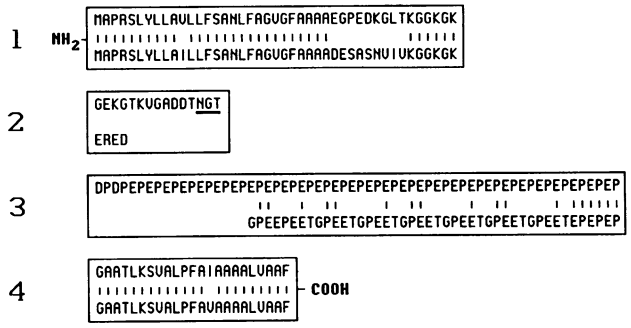


FIG. 4. Comparison of PARP polypeptides B-1α and Aα. The deduced amino acid sequences of the B-1α (11) and Aα (this report) polypeptides are aligned, with the B-1α sequence positioned above. For comparison, the proteins are divided into four domains: (1) amino-terminal domain; (2) transition domain with consensus N-linked glycosylation signal underlined; (3) repeat domain; and (4) carboxy-terminal hydrophobic domain. Vertical bars indicate identity between the two sequences.

(Fig. 3). Both probes detect RNAs present in PCTs but absent in bloodstream-form trypomastigotes. The average size of the Aα mRNA is approximately 100 nt less than that of the 1,000-nt mRNAs recognized by the EP probe and is in good agreement with the size computed from the Aα genomic sequence. Although total RNA preparations were tested in this analysis, identical hybridization patterns were observed in the analysis of poly(A)$^+$ RNA (data not shown). Hybridization of the GPEET probe to blots of PCT polysomal RNAs established that, like EP repeat-containing PARP mRNAs (11), Aα mRNA is enriched in membrane-bound polyribosomal RNA preparations (Fig. 3, panel 3). This result suggests that the Aα polypeptide, like other PARPs, is membrane associated.

From the perspective of transcriptional regulation, it is significant that both EP- and GPEET-encoding PARP genes are simultaneously expressed during procyclic differentiation. We previously used in vitro translation of total PCT RNA to show that *T. brucei* 427 transcribes at least three different EP-encoding PARP genes (12). Since the parpA locus contains only one EP-encoding PARP gene, parpAβ, we conclude that one or both of the parpB allelic loci are transcribed in this strain, i.e., the parpA locus does not encode all PARP mRNAs in this strain. In addition, since we have obtained a cDNA clone derived from a parpB-1α gene transcript (11, 12) and can identify parpAα transcripts in the steady-state RNA of TREU 667 procyclic trypanosomes (data not shown), it is clear that both loci are transcribed in this strain. These data strongly support the notion that the two unlinked loci, parpA and parpB, are coordinately regulated during development. We have not determined the relative abundance of GPEET- and EP-encoding PARP mRNAs in procyclic trypanosomes; however, nuclear transcription analyses, currently in progress, will address this issue and may help to identify regions that coordinately regulate transcription of the parpA and parpB loci.

Comparison of PARP primary structures. The substitution of repeating units is the most striking feature that distinguishes the Aα polypeptide from other PARPs. Nevertheless, a direct comparison of the primary structures of the Aα and B-1α polypeptides revealed additional noteworthy distinctions and similarities (Fig. 4). The amino-terminal regions (domain 1) of the two proteins are well conserved with the exception of divergent nonapeptides that separate the basic hexapeptide lysine-glycine-glycine-lysine-glycine-lysine from a 27-residue amino-terminal signal peptide that is removed, at least in the case of glycosylated B-1α-like PARPs (14; C. Clayton, manuscript in preparation). Likewise, the carboxy-terminal regions (domain 4), which are rich in hydrophobic residues, are nearly identical. In direct contrast to the conserved amino- and carboxy-terminal domains are the divergent transition and repeat domains, designated domains 2 and 3, respectively (Fig. 4). Besides the 11-residue size difference between the transition domains of the B-1α and Aα polypeptides, B-1α is distinguished by the presence of the consensus N-linked glycosylation signal Asn-Gly-Thr (10). This feature predicts the absence of N-linked carbohydrate on the Aα polypeptide, a biochemical property with potential applicability to the chromatographic separation of the different PARPs. In fact, the validity of this hypothesis is supported by the observation that concanavalin A affinity-purified PARP yields the amino-terminal sequence of the B-1α, and not the Aα, polypeptide shown in Fig. 4 (Clayton, in preparation). Finally, in Aα the substitution of repeats is not complete relative to B-1α, nor is the Aα repeat region homogeneous with respect to the pentapeptide repeating unit (Fig. 4, domain 3). In Aα $(GPEET)_6$ is preceded by the degenerate octapeptide GPEEPEET and followed by $(EP)_3$. This conservation of dipeptide repeats abutting a hydrophobic carboxy-terminal domain may reflect constraints imposed by the retention of carboxy-terminal processing signals, e.g., for glycosyl-phosphatidylinositol anchor addition, termed glypiation by Cross (6), as proposed for this procyclic protein by Ferguson and Williams (8).

The functional significance of tandem repeat substitution in PARP is as unclear as the function of PARP itself. From

the conservation of amino- and carboxy-terminal regions, however, we predict that the Aα protein is processed in the same manner and targeted to the same cellular location as other PARPs, i.e., the parasite surface. *T. brucei* and other parasites of humans and animals are known to evade immune destruction in their mammalian hosts by various means of altering their antigenic character. Since PARPs are expressed by *T. brucei* during residence in its insect vector, however, this variability in tandem repeats cannot be involved in evasion of the mammalian immune system. Instead, we believe it likely that PARPs play a role in parasite development in the tsetse fly. Roditi et al. (submitted) used molecular modeling of the EP repeat to predict that the protein is a rigid rod 14 to 18 nm long and elegantly showed that the repeat region extends beyond the dense coat of variant surface glycoprotein that is shed in the course of procyclic differentiation. It would be interesting to determine whether the GPEET repeat region adopts a similar conformation or whether the processed Aα polypeptide with its divergent amino-terminal primary structure is masked by EP-containing PARPs. This determination, as well as an assessment of function, awaits the generation of serological reagents that distinguish the two types of PARPs.

The organization of conserved and variable elements in an externally disposed membrane protein is strongly reminiscent of the S antigens of *Plasmodium falciparum* (5). These proteins exhibit changes in the number and type of tandem repeats and are encoded by variant genes at a single locus (13, 15). As Cowman et al. have suggested for the S antigen genes (5) and as we have suggested for the PARP genes (12), diversity among these genetic elements was probably generated by a combination of point mutations, unequal crossing over, and gene conversion. Unlike the PARPs of *T. brucei*, however, S antigens are expressed by *P. falciparum* during growth in its mammalian host, and therefore different selective pressures must have effected this conserved organization of structural motifs.

We thank P. Overath (Max Planck Institut für Biologie, Tübingen, Federal Republic of Germany) for communicating results before publication, as well as G. A. M. Cross, S. Mayor, and S. Vijayasarathy for critically evaluating the manuscript. We are especially grateful to G. A. M. Cross for his encouragement and support.

This work was supported by Public Health Service grant AI22229 from the National Institutes of Health and by a Hirschl/Weill-Caulier award to C.E.C.

LITERATURE CITED

1. **Biggin, M., T. Gibson, and G. Hong.** 1983. Buffer gradient gels and [^{35}S]label as an aid to rapid DNA sequence determinations. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
2. **Borst, P.** 1986. Discontinuous transcription and antigenic variation in trypanosomes. Annu. Rev. Biochem. **55**:701–732.
3. **Brown, R. C., D. A. Evans, and K. Vickerman.** 1973. Changes in oxidative metabolism and ultrastructure accompanying differentiation of the mitochondrion in *Trypanosoma brucei*. Int. J. Parasitol. **3**:691–704.
4. **Clayton, C.** 1988. The molecular biology of the Kinetoplastidae. Genet. Eng. **7**:1–56.
5. **Cowman, A. F., R. B. Saint, R. L. Coppel, G. V. Brown, R. F. Anders, and D. J. Kemp.** 1985. Conserved sequences flank variable tandem repeats in two S-antigen genes of *Plasmodium falciparum*. Cell **40**:775–783.
6. **Cross, G. A. M.** 1987. Eukaryotic protein modification and membrane attachment via phosphatidylinositol. Cell **48**:179–181.
7. **Feinberg, A., and B. Vogelstein.** 1982. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–12.
8. **Ferguson, M. A. J., and A. F. Williams.** 1988. Cell surface anchoring of proteins via glycosyl-phosphatidylinositol structures. Annu. Rev. Biochem. **57**:285–320.
9. **Freistadt, M., H. T. Robertson, A. Branch, and G. A. M. Cross.** 1987. Direct analysis of the mini-exon-donor RNA of *Trypanosoma brucei*: detection of a novel cap structure also present in messenger RNA. Nucleic Acids Res. **15**:9861–9880.
10. **Hart, G. W., K. Brew, G. A. Grant, R. A. Bradshaw, and W. J. Lennarz.** 1979. Primary structural requirements for the enzymatic formation of the N-glycosidic bond in glycoproteins. J. Biol. Chem. **254**:9747–9753.
11. **Mowatt, M. R., and C. E. Clayton.** 1987. Developmental regulation of a novel repetitive protein of *Trypanosoma brucei*. Mol. Cell. Biol. **7**:2838–2844.
12. **Mowatt, M. R., and C. E. Clayton.** 1988. Polymorphism in the procyclic acidic repetitive protein gene family of *Trypanosoma brucei*. Mol. Cell. Biol. **8**:4055–4062.
13. **Nicholls, S. C., Y. Hillman, M. J. Lockyer, K. G. Odink, and A. A. Holder.** 1988. An S antigen gene from *Plasmodium falciparum* contains a novel repetitive sequence. Mol. Biochem. Parasitol. **28**:11–20.
14. **Roditi, I., M. Carrington, and M. Turner.** 1987. Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*. Nature (London) **325**: 272–274.
15. **Saint, R. B., R. L. Coppel, A. F. Cowman, G. V. Brown, P.-T. Shi, N. Barzaga, D. J. Kemp, and R. F. Anders.** 1987. Changes in repeat number, sequence, and reading frame in S-antigen genes of *Plasmodium falciparum*. Mol. Cell. Biol. **7**:2968–2973.
16. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
17. **Van der Ploeg, L. H. T.** 1986. Discontinuous transcription and splicing in trypanosomes. Cell **47**:479–480.
18. **Vickerman, K.** 1985. Developmental cycles and biology of pathogenic trypanosomes. Br. Med. Bull. **41**:105–114.
19. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.