The Procyclin Repertoire of *Trypanosoma brucei*
IDENTIFICATION AND STRUCTURAL CHARACTERIZATION OF THE GLU-PRO-RICH POLYPEPTIDES*

(Received for publication, June 30, 1999, and in revised form, August 2, 1999)

Alvaro Acosta-Serrano‡§, Robert N. Cole¶, Angela Mehlert¶, Mary G.-S. Lee**‡‡, Michael A. J. Ferguson†, and Paul T. Englund‡

The surface of the insect stages of the protozoan parasite *Trypanosoma brucei* is covered by abundant glycosyl phosphatidylinositol (GPI)-anchored glycoproteins known as procyclins. One type of procyclin, the EP isoform, is predicted to have 22–30 Glu-Pro (EP) repeats in its C-terminal domain and is encoded by multiple genes. Because of the similarity of the EP isoform sequences and the heterogeneity of their GPI anchors, it has been impossible to separate and characterize these polypeptides by standard protein fractionation techniques. To facilitate their structural and functional characterization, we used a combination of matrix-assisted laser desorption ionization and electrospray mass spectrometry to analyze the entire procyclin repertoire expressed on the trypanosome cell. This analysis, which required removal of the GPI anchors by aqueous hydrofluoric acid treatment and cleavage at aspartate-proline bonds by mild acid hydrolysis, provided precise information about the glycosylation state and the number of Glu-Pro repeats in these proteins. Using this methodology we detected in a *T. brucei* clone the glycosylated products of the EP3 gene and two different products of the EP1 gene (EP1-1 and EP1-2). Furthermore, only low amounts of the nonglycosylated products of the GPEET and EP2 genes were detected. Because all procyclin genes are transcribed polycistronically, the latter finding indicates that the expression of the GPEET and EP2 genes is post-transcriptionaly regulated. This is the first time that the whole procyclin repertoire from procyclic trypanosomes has been characterized at the protein level.

African trypanosomes are protozoan parasites responsible for sleeping sickness in humans and the disease nagana in livestock. Trypanosomes alternate between the invertebrate vector (Glossina, or tsetse fly) and the mammalian host, and the different life cycle stages are uniquely adapted to survive in each host. In the mammalian bloodstream form of the parasite, 10^7 identical variant surface glycoprotein (VSG) molecules are expressed on the plasma membrane, forming a dense coat. The parasite can survive the immune attack of its host because it undergoes antigenic variation, a process in which its surface coat is replaced by another composed of an antigenically different VSG molecule (1, 2). When the trypanosome is ingested by the tsetse fly and differentiates into a procyclic form, the VSG coat is totally replaced by one composed of an array of different proteins, known as procyclins (3–5).^2^ The *Trypanosoma brucei* procyclins, present in about 2.2 × 10^6^ copies/cell, have very unusual structures, with the C-terminal domain consisting of amino acid repeats. One set of proteins, the EP isoforms, is predicted to contain between 22–30 Glu-Pro (EP) repeats, whereas the GPEET form, in contrast, has six Gly-Pro-Glu-Glu-Thr (GPEET) repeats followed by three EP repeats (see Fig. 1A for a schematic diagram of procyclin structures and Fig. 1B for their amino acid sequences). Unlike VSG, procyclin is encoded by a small number of different genes, and therefore it has a limited potential for variation. In the *T. brucei* 427 strain, the EP isoforms are encoded by the EP1 (4, 6–8), EP2 (10), and EP3 (5, 9) genes, whereas GPEET-procyclin is the product of the GPEET gene (11). All procyclin genes are contained in four expression sites (per diploid genome), and each site contains two procyclin genes in a tandem array (see map in Fig. 1C). Transcription of these genes is polycistronic and can occur simultaneously from two or more expression sites (reviewed in Ref. 12).

There are other important structural features of the procyclin proteins (Fig. 1A). Some EP forms can be N-glycosylated with the products of the EP1 and EP3 genes, but not that of the EP2 gene, containing a site for N-glycan addition (*i.e.* Asn29; Fig. 1B). GPEET-procyclin also has no N-glycosylation site. The N-glycan is Man3GlcNAc2, and it is unusual in that it does not exhibit microheterogeneity (13, 14). In addition, all of the procyclin proteins have GPI anchors of unusual structure, characterized by a very large (average of 30 sugar residues), heterogeneous branched poly-N-acetyllactosamine side chain (13, 15). This side chain serves as the sialic acid acceptor for the cell surface trans-sialidase (16, 17). The anchors on EP- and

---

* This work was supported by National Institutes of Health Grants AI21334 (to P. T. E.) and AI28953 (to M. S.-G. L.) and by Wellcome Trust Program Grant 054491 (to M. A. J. F.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported in part by a postdoctoral Fellowship from Consejo Nacional de Investigaciones Cientı´ficas y Tecnolo´gicas (Venezuela). To whom correspondence should be addressed. Tel.: 410-955-3458; Fax: 410-955-7810; E-mail: aacostas@welchlink.welch.jhu.edu.

¶ Supported by the American Health Assistance Foundation.

‡‡ Burroughs Wellcome Fund New Investigator in Molecular Parasitology.

---

1 The abbreviations used are: VSG, variant surface glycoprotein; EP-procyclin, a form of procyclin rich in Glu-Pro repeats; GPEET-procyclin, a form of procyclin rich in Gly-Pro-Glu-Glu-Thr repeats; GPI, glycosyl phosphatidylinositol; aq HF, aqueous hydrofluoric acid; MS, mass spectrometry; MALDI-TOF-MS, matrix-assisted laser desorption ionization time-of-flight MS; TFA, trifluoroacetic acid; ESI-MS, electrospray ionization-mass spectrometry; EtN, ethanolamine; PNGase F, peptide N-acetyl-b-glucosaminidase amidase F; CID, collision-induced dissociation.

2 These proteins have also been called procyclin acidic repetitive proteins (PARPs), but a new standard nomenclature for genes and proteins now uses the term procyclin (37).
In elucidating the function of procyclin molecules, it is of great interest to determine whether all of the different EP-procyclin genes are expressed as proteins and to determine the relative levels of expression. It is also important to establish whether this expression varies with time or under different experimental conditions. Such variation could have considerable biological relevance. For example, a programmed variation of procyclin expression could control the behavior of the parasite in its insect vector. Although the EP and GPEET forms can be detected by monoclonal antibodies (20), to date it has not been possible to analyze or resolve the protein products of the different EP-procyclin genes, because of their very similar amino acid sequences (Fig. 1B) in addition to the extreme heterogeneity in their GPI anchor.

In this paper, we report a complete characterization of the EP-procyclin repertoire by mass spectrometry, revealing most of the species predicted from gene sequences. To apply this technique, it was necessary to remove the GPI anchor by treatment with aqueous hydrofluoric acid (aq.HF), a well-characterized method that has a minimal effect on the polypeptide chain and glycosidic bonds. This method was previously used in a mass spectrometric analysis of the phosphorylation of purified GPEET-procyclin (19). However, to fully characterize these molecules it was also essential to develop a new method involving mild acid hydrolysis that selectively cleaves the EP1 and EP3 (but not EP2) gene products at Asp-Pro sequences. Using this methodology, we were able to identify by mass spectrometry all of the procyclin polypeptides present in a T. brucei 427 strain as well as to obtain new information on procyclin expression. Furthermore, for each of the EP isoforms, we determined an accurate molecular mass, the extent of glycosylation, and the exact number of EP repeats.

**EXPERIMENTAL PROCEDURES**

**Parasites**—The wild type procyclic T. brucei brucei 427 clone had been stored at New York University in the laboratory of M. G.-S. Lee. This cell line was originally obtained in 1988 by *in vitro* differentiation of T. b. brucei 118 bloodstream clone 1 (6). The parasites had been cultured as procyclic cells for a total of about 6 weeks (since differentiation) before procyclin was isolated for the experiments described in Figs. 2–5. Ten cloned lines from the original 1988 clone were obtained at New York University by stably transforming the 1988 clone with the same DNA construct (H23H-B7) and using hygromycin as a selectable marker (22). The H23-H-B7 construct contains the *hsp70* intergenic region and the *dhp* gene, the β-tubulin intergenic region, and a targeting sequence derived from the VSG 118 expression site (22). The transformed parasites were cloned by limiting dilution in the presence of trypanosomes not resistant to the drug. Another cloned line (clone 6) was obtained at Johns Hopkins University from the same T. brucei 1988 strain by limiting dilution in conditioned medium. Parasites were grown at 27 °C in SDM-79 medium (23), supplemented with 10% heat-inactivated fetal bovine serum (Life Technologies, Inc.).

**Purification of Procyclins**—For the experiments shown in Figs. 2–5, the procyclins (including GPEET- and EP forms) were purified from 1011 freeze dried trypanosomes by organic solvent extraction followed by octyl-Sepharose chromatography (Amersham Pharmacia Biotech) (13, 15). As judged by SDS-polyacrylamide gel electrophoresis and silver staining, the procyclin-containing fractions from the octyl-Sepharose column showed a major component with a highly polydisperse migration and an apparent molecular mass of ~45 kDa (not shown). Sugar analysis of a pool of these fractions by gas chromatography mass-spectrometry (24) yielded a composition of Man, Gal, GlcNAc, sialic acid, and myo-inositol of 5.2:14.8:27:4.2:1, similar to values reported previously (13, 15). Some preparations also contained a minor component with Mr near 11 kDa on SDS-polyacrylamide gel electrophoresis. This component was identified as the kinetoplastid membrane protein-11, a protein known to co-purify with procyclin preparations (25). MALDI-TOF-MS analysis revealed this component as a broad peak with an average molecular mass of 11,070 Da (not shown).

The mass spectrometry analysis of procyclins from cloned cell lines was performed directly from n-butanol extracts without further purification by octyl-Sepharose chromatography. The mass spectra of these polypeptides contain a variable number of Glu-Pro repeats in the C-terminal domain. EP-procyclin products of the EP1 and EP3 genes are glycosylated at Asn29 and contain an Asp-Pro-Asp-Pro sequence that is sensitive to cleavage by mild acid (28). The EP2 product contains neither a glycosylation site nor Asp-Pro bonds (10). GPEET-procyclin, the product of the GPEET gene, has no site for N-glycosylation and contains six GPEET repeats (11). This molecule is extensively phosphorylated at threonines in the repeats (18, 19). B, predicted amino acid sequences of the products of the EP-procyclin genes (EP1-1, EP1-2, EP3, and EP2) (5–10) and the GPEET-procyclin gene (GPEET) (11). Another EP1 gene product predicted to contain 29 EP repeats and Gly at position 1011 was partially rescued after overexpression of either nonglycosylated at threonines in the repeat (18, 19). As judged by SDS-polyacrylamide gel electrophoresis and silver staining, the procyclin-containing fractions from the octyl-Sepharose column showed a major component with a highly polydisperse migration and an apparent molecular mass of ~45 kDa (not shown). Sugar analysis of a pool of these fractions by gas chromatography mass-spectrometry (24) yielded a composition of Man, Gal, GlcNAc, sialic acid, and myo-inositol of 5.2:14.8:27:4.2:1, similar to values reported previously (13, 15). Some preparations also contained a minor component with Mr near 11 kDa on SDS-polyacrylamide gel electrophoresis. This component was identified as the kinetoplastid membrane protein-11, a protein known to co-purify with procyclin preparations (25). MALDI-TOF-MS analysis revealed this component as a broad peak with an average molecular mass of 11,070 Da (not shown).

The mass spectrometry analysis of procyclins from cloned cell lines was performed directly from n-butanol extracts without further purification by octyl-Sepharose chromatography. The mass spectra of these polypeptides contain a variable number of Glu-Pro repeats in the C-terminal domain. EP-procyclin products of the EP1 and EP3 genes are glycosylated at Asn29 and contain an Asp-Pro-Asp-Pro sequence that is sensitive to cleavage by mild acid (28). The EP2 product contains neither a glycosylation site nor Asp-Pro bonds (10). GPEET-procyclin, the product of the GPEET gene, has no site for N-glycosylation and contains six GPEET repeats (11). This molecule is extensively phosphorylated at threonines in the repeats (18, 19). B, predicted amino acid sequences of the products of the EP-procyclin genes (EP1-1, EP1-2, EP3, and EP2) (5–10) and the GPEET-procyclin gene (GPEET) (11). Another EP1 gene product predicted to contain 29 EP repeats and Gly at position 1011 was partially rescued after overexpression of either nonglycosylated at threonines in the repeat (18, 19). As judged by SDS-polyacrylamide gel electrophoresis and silver staining, the procyclin-containing fractions from the octyl-Sepharose column showed a major component with a highly polydisperse migration and an apparent molecular mass of ~45 kDa (not shown). Sugar analysis of a pool of these fractions by gas chromatography mass-spectrometry (24) yielded a composition of Man, Gal, GlcNAc, sialic acid, and myo-inositol of 5.2:14.8:27:4.2:1, similar to values reported previously (13, 15). Some preparations also contained a minor component with Mr near 11 kDa on SDS-polyacrylamide gel electrophoresis. This component was identified as the kinetoplastid membrane protein-11, a protein known to co-purify with procyclin preparations (25). MALDI-TOF-MS analysis revealed this component as a broad peak with an average molecular mass of 11,070 Da (not shown).

The mass spectrometry analysis of procyclins from cloned cell lines was performed directly from n-butanol extracts without further purification by octyl-Sepharose chromatography. The mass spectra of these polypeptides contain a variable number of Glu-Pro repeats in the C-terminal domain. EP-procyclin products of the EP1 and EP3 genes are glycosylated at Asn29 and contain an Asp-Pro-Asp-Pro sequence that is sensitive to cleavage by mild acid (28). The EP2 product contains neither a glycosylation site nor Asp-Pro bonds (10). GPEET-procyclin, the product of the GPEET gene, has no site for N-glycosylation and contains six GPEET repeats (11). This molecule is extensively phosphorylated at threonines in the repeats (18, 19). B, predicted amino acid sequences of the products of the EP-procyclin genes (EP1-1, EP1-2, EP3, and EP2) (5–10) and the GPEET-procyclin gene (GPEET) (11). Another EP1 gene product predicted to contain 29 EP repeats and Gly at position 24 has been reported from another strain (4), but it was not detected in this study. For simplicity, only the sequence of mature proteins (without signal peptides and GPI anchor signals) are shown; sequence numbering starts at the N terminus of the mature protein (i.e. Ala1). Underlined letters represent sites for N-glycosylation (Asn29). ~10 and ~4 with brackets indicate the N-terminal sequences missing in some of the species identified by MALDI-TOF-MS (Figs. 2A and 6). C, simplified diagram of procyclin loci in T. brucei 427 strain. Loci names are at left. All genes are allelic, and the products of the EP1-1 and EP1-2 genes differ in sequence (see text). This model is not drawn to scale and shows only the pair of procyclin genes present in each locus. The EP1-1, EP1-2, EP2, EP3, and GPEET genes were formerly called R2a, R1a, Rb, Ab, and As respectively (see new nomenclature in Ref. 37). A more detailed map of procyclin expression sites is shown elsewhere (12, 37).

GPEET-procyclin are similar or identical in structure (13). Additionally, GPEET-procyclin is modified by phosphorylation on six out of seven threonine residues in the repeat sequence (18–20), but this modification has not been reported for the EP forms.

The function of any of the procyclins is still unclear, although recently it was demonstrated that mutant parasites that express no EP isoforms cannot establish heavy infections in tsetse fly midguts (21). Furthermore, the wild type phenotype was partially rescued after overexpression of either nonglycosylated or glycosylated EP isoforms (21). Interestingly, both EP- and GPEET-procyclin are co-expressed on the parasite surface, although their ratio varies in different clones. When some of these clones are maintained in culture for several months, there is a shift in expression, from EP- to GPEET-procyclin, ending with cells containing low levels of EP proteins (13, 18).
preparations were as clean as those obtained using procyclin purified by the standard chromatographic method (see method in Fig. 6).

**Removal of the Procyclin GPI Anchors by aq.HF Dephosphorylation**—Octyl-Sepharose purified procyclin (1–2 nmol, based on *m*yo-inositol content) or *n*-butanol extracts from cloned cells were dephosphorylated with 50 µl (or 25 µl in the case of butanol extracts) of cold 48% aq.HF (Aldrich) for 16 h at 0 °C (19). After hydrolysis, samples were quickly frozen in dry ice/ethanol and dried in the Speed-Vac. Samples were resuspended in 5–10 µl of water and stored at −20 °C.

**MALDI-TOF-MS**—Mass spectra were acquired in a PerSeptive Biosystems Voyager-DE mass spectrometer calibrated with insulin, thioridoxin, and apomyoglobin. For the analysis of native procyclin (500 pmol; non-aq.HF-treated), samples were co-crystallized with sinapinic acid. Polypeptides dephosphorylated with aq.HF (50 pmol) provided better spectra using α-cyano-4-hydroxycinnamic acid as the matrix. All spectra were collected in the negative ion mode.

**Analysis of N-terminal Sequences by Electrospray Ionization-Mass Spectrometry (ESI-MS) and Tandem-Mass Spectrometry—**ESI-MS was performed using a Finnigan LCQ atmospheric pressure ionization quadrupole ion trap mass spectrometer (ThermoQuest Corp.). The mass spectrometer was equipped with an X-Y-Z-positioner carriage source (Michrom BioResources). Samples were resuspended in 50 mM formic acid before analysis. For analysis of the EP-procyclin N-terminal sequences by ESI-MS, 4.5 nmol of native procyclin was deglycosylated with 250 units of PNGase F (Biolabs) for 2 h at 37 °C. After deglycosylation, polypeptides were desalted using a ZipTip (Millipore Corp.) containing C18 silica. Samples were eluted as recommended by the manufacturer, dried in a Speed-Vac, and resuspended in 50 mM formic acid before analysis.

**Enzymatic Deglycosylation of Procyclin Polypeptides—**aq.HF-treated procyclin (200 pmol) was deglycosylated with 500 units of peptide *N*4([a]-acetyl-β-glucosaminyl) asparagine amidase F (PNGase F) (New England Biolabs) in 10 µl of 25 mM sodium phosphate at 37 °C for 2 h. After digestion, 1 µl of the sample was directly mixed with 1 µl of α-cyano-4-hydroxycinnamic acid and analyzed by MALDI-TOF-MS. For analysis of procyclin N-terminal polypeptides by ESI-MS, the polypeptides were incubated after mild acid hydrolysis and octyl-Sepharose chromatography (see above) were incubated with PNGase F as described above except that the buffer was 50 mM sodium phosphate and the digestion conditions were 37 °C for 5 h. Prior to mass spectral analysis, the deglycosylated polypeptides were desalted using a ZipTip (Millipore Corp.) containing C18 silica. Samples were eluted as recommended by the manufacturer, dried in a Speed-Vac, and resuspended in 50 mM formic acid for analysis.

**Characterization of Native Procyclin by Mass Spectrometry—**For MALDI-TOF-MS analysis, we first used native procyclin purified by organic solvent extraction and octyl-Sepharose chromatography. The negative ion spectrum revealed a highly heterogeneous group of [M-H]− pseudomolecular ions in the range of m/z 14,000–20,000 (average m/z, 17,400), consistent with a range of 17.4 kDa (not shown). A comparable spectrum was reported previously for purified GPEET-procyclin, and the extensive heterogeneity observed is mainly due to the polydisperse side chain of the GPI anchor (13, 19). To distinguish the different EP-procyclin species, it was first essential to remove the GPI anchors.

**Characterization of the EP-procyclin Polypeptides by Mass Spectrometry after Removal of GPI Anchors—**We removed the anchors by treatment with cold 48% aq.HF, a reagent that cleaves the phosphodiester bond between the anchor ethanolamine and the GPI glycan core (26, 27). This treatment preserves most polypeptide chains with N-glycan moieties intact and leaves the GPI ethanolamine amide-linked to the C-terminal α-carboxyl. In the case of EP-procyclin, HF also produced minor cleavages in the protein backbone, mainly at the mild acid-sensitive Asp-Pro bonds (see below).

The negative ion MALDI mass spectrum of HF-treated procyclins revealed many [M-H]− pseudomolecular ions in the range of m/z 5,000–12,000 (Fig. 2A). Although the spectrum is complex, nearly all of these species were interpretable in terms of the well-characterized procyclin gene sequences and N-glycosylation patterns. Table I lists the masses and assignments of the various species detected. Two ions at m/z 11,531 (ion 1)
and 10,430 (ion 3) had molecular masses for two different glycosylated products in the EP1 gene, each also containing one ethanolamine. According to their cDNA sequences and results shown later in this paper, these proteins differ in both the number of EP repeats (30 for ion 1, 25 for ion 3) and in one amino acid at position 24 (ions 4, 6–8) (Fig. 1B). We have designated these ions as the products of the EP1-1 and EP1-2 genes, respectively. Another major [M-H]⁻ ion at m/z 9,723 (ion 4) matches the calculated mass of the glycosylated product of the EP1-1 gene, a protein with 22 EP repeats (5, 9) (Fig. 1B). Interestingly, no ion corresponding to the product of the EP1-2 gene was observed. However, a tiny peak at m/z 8,344 (ion 10) was tentatively assigned as a product of this gene, but this polypeptide contained only 24 EP repeats instead of the 25 predicted from the published cDNA sequence (10). To distinguish both EP1-procycalin species, we have designated them as EP2-1 (containing 25 EP repeats) and EP2-2 (containing 24 EP repeats). As expected, the mass of EP2-2–procyclin (ion 10) predicted that this polypeptide was not N-glycosylated (10).

For each of the full-length EP-procyclin gene products observed in the negative ion spectrum, we also detected a corresponding ion that lacked ten amino acids from the N terminus (AEGPEDKGKL; Fig. 1B). Thus, the [M-H]⁻ ions at m/z 10,533 (ion 2), 9,432 (ion 5), and 8,725 (ion 8) (marked by asterisks in Fig. 2A) represent the truncated forms of the EP1-1, EP1-2, and EP3 gene products, respectively. These fragments varied in abundance in different preparations, and control experiments strongly indicated that they originate during the aq.HF treatment (see “Discussion”).

Another group of lower intensity [M-H]⁻ ions in the range of m/z 7,500–9,500 (ions 6 and 11–15; Table I), mainly represent minor products of nonspecific cleavages from the N termini of the products of the EP1-1, EP1-2, and EP3 genes. Some of them are worth discussing because they support the assignment of two types of EP1 gene products. For instance, those ions at m/z 8,175 (ion 12) and 8,276 (ion 11) are consistent with a polypeptide containing Ser at position 24 as well as 30 EP repeats, which we have designated the EP1-2 polypeptide (Table I). Likewise, the ion at m/z 9,277 (ion 6) is consistent with a polypeptide containing a Gly at position 24 as well as 30 EP repeats, expected for the EP1-1 protein. As discussed below, we confirmed these proposed structures of the EP1-1 and EP1-2 proteins by tandem mass spectrometry. Finally, another group of ions (ions 17, 18, 22, 23, 27, 28, and 29) derive from cleavage at Asp-Pro sequences that mediate the reaction of the HP de-phosphorylation. We will discuss the significance of these ions below, in the section on mild acid hydrolysis.

**Characterization of GPEET-procyclin Polypeptides**—We also detected polypeptides derived from GPEET-procyclin, the product of the GPEET gene. The ion at m/z 6,142 (ion 21) corresponded to the intact GPEET-procyclin polypeptide and that at m/z 6,222 (ion 20) to the same polypeptide containing one phosphate group. Presumably the other phosphates had been...
removed during the aq.HF dephosphorylation. These values are consistent with those recently reported for the same molecule (19). The ion at m/z 5,703 (ion 25) corresponds to nonphosphorylated GPEET-procyclin that has lost its N-terminal sequence VIVK (Fig. 1B).

Glycosylation of the EP-procyclin Polypeptides—The major EP-procyclin polypeptides identified in Fig. 2A have molecular masses consistent with its single Asn being modified by a Hex$_3$HexNAc$_2$ oligosaccharide. Previous structural analyses have proven that this glycan is Man$_n$GlcNAc$_2$ (13, 14). To further confirm our polypeptide assignments and to determine the occupancy of the N-glycosylation sites, we deglycosylated a sample like that in Fig. 2A with PNGase F and then analyzed the products by MALDI-TOF-MS (Fig. 2B). The major ions corresponding to the intact polypeptide products of the EP1-1, EP1-2, and EP3 genes are missing (ions 1, 3, and 4) as are the corresponding ions that have lost the N-terminal decapeptide (ions 2, 5, and 8). In their place are new ions at m/z 10,314, 9,213, and 8,506, which have values corresponding to the deglycosylated products of the EP1-1, EP1-2, and EP3 genes, respectively, as well as ions at 9,316 m/z (EP1-1), 8,215 (EP1-2), and 7,508 (EP3), which are the deglycosylated species with truncated N termini. In all cases, the mass of the deglycosylated species is 1217 Da less than that of the native species, a mass corresponding to that of a Hex$_3$HexNAc$_2$ glycan. The corresponding deglycosylated species detected in Fig. 2B were present only in very low levels in the MALDI spectrum shown in Fig. 2A, indicating that most of the EP-procyclin polypeptides were occupied by a single N-glycan, as predicted in a previous report (13). Furthermore, we found no evidence of any heterogeneity in this glycan, also in agreement with previous reports (13, 14).

Interestingly, after PNGase F incubation (Fig. 2B and inset), two new very intense ions at m/z 3,119 and 3,331 were detected. These ions were identified as the deglycosylated N-terminal fragments Ala$^1$–Asp$^{34}$ (m/z 3,331) and Ala$^1$–Asp$^{32}$ (m/z 3,119) of the EP1-1 and EP3$^3$ proteins, produced by cleavage at Asp-Pro bonds. An amplification of the region between m/z 3,000–3,500 (inset, Fig. 2B) showed that there were two additional ions. These were identified as the deglycosylated N-terminal fragments Ala$^1$–Asp$^{34}$ (m/z 3,360) and Ala$^1$–Asp$^{32}$ (m/z 3,147) of the EP1-2 protein. As shown in Fig. 1B, the EP1-2 and EP1-1 gene products differ in this region only by the presence of a Gly or a Ser at position 24. The corresponding ions in their glycosylated form (e.g. m/z 4,577 and 4,365) were never observed in the sample not treated with PNGase F (Fig. 2A), even when analyzed in the positive mode (not shown).

Characterization of the C-terminal Fragments of the EP-procyclins after Selective Mild Acid Hydrolysis of the Asp-Pro Bonds—To confirm the number of EP repeats, we partially hydrolyzed the procyclin polypeptides (already aq.HF-treated) with 40 mM TFA at 100 °C for 15 min. Under these relatively mild conditions, Asp-Pro peptide bonds (found at residues 32–35 in EP3 and EP1 procyclins) are preferentially cleaved (28). The negative ion MALDI-TOF-MS spectrum after mild acid hydrolysis (Fig. 3) confirmed that cleavage occurs at these Asp-Pro sequences, generating distinct C-terminal fragments. Every EP-procyclin molecule was cleaved at one of these sites, because no ions in the range m/z 8,000 to 12,000 were detected (not shown). In the range of m/z 5,000 and 7,000, three major ions were observed. These ions were identified as the C-termi-
Ala¹–Asp³². However, the more informative series were the triply charged b ions (b³⁺), because they defined the region where the three polypeptides have different amino acid sequences (i.e., positions 18, 24 and 25). The results are summarized in Fig. 4B, in which the observed masses (from the spectrum in Fig. 4A) can be compared with the predicted masses of the b³⁺ series for the three EP-procyclin gene products (only N-terminal fragments). These results, together with those obtained from MALDI-TOF-MS analyses (Fig. 2A and Table I), confirm the presence of the products of the EP1-1, EP1-2, and EP3 genes.

Is EP-procyclin Phosphorylated?—It is well established that the GPEET-procyclin is highly phosphorylated on the amino acid repeats (18–20), but the presence of this modification has not been determined for the EP isoforms. The EP-procyclins contain several Thr or Ser residues that could be phosphorylation sites. Consistent with phosphorylation, we found that MS analysis of procyclins treated for 2–8 h with aq.HF showed, in addition to the phosphorylated GPEET forms, new peaks (not detected in the MS in Fig. 2A) that were larger by 80 Da than each EP isoform (not shown). This finding suggested the presence of singly phosphorylated EP-procyclin species. We then studied whether this phosphate is present in the N-terminal domain or linked to the C-terminal ethanolamine, deriving from the GPI anchor. We incubated native procyclin for 16 h with calf intestine alkaline phosphatase and then treated it for 6 h with aq.HF, and finally submitted it to mild acid hydrolysis. Analysis by MALDI-TOF-MS showed that each of the major C-terminal polypeptides (m/z 5,191, 5,870, and 7,001) of the EP-procyclins is partly phosphorylated, resulting in ions of m/z 5,011, 5,690, and 6,811, respectively (Fig. 5A). We also detected phosphorylated species of the less abundant C-terminal fragments that were cleaved at the upstream Pro-Asp bond (i.e., m/z 5,403, 6,082, and 7,213), as well as the fragment P(EP)²¹G-EtN (m/z 4,965). Because these phosphate groups had resisted hydrolysis by alkaline phosphatase, these results indicate that they were originally present in the native molecules as phosphodiesters. Furthermore, subsequent deglycosylation of the same sample showed that none of the N-terminal fragments was phosphorylated (Fig. 5B). Taken together, these results provide strong evidence that in contrast to GPEET-procyclin, the EP-procyclins are not phosphorylated. The residual phosphate groups detected by mass spectrometry after short HF treatment must derive from the phosphodiester bond originally present between the C-terminal ethanolamine and the GPI glycan core.

Stability of Expression of the Procyclin Repertoire—The parasites used for this study had been stored at New York University since 1988, soon after they had been transformed from a cloned bloodstream form (strain 118). They had been cultured only about 6 weeks before the isolation of procyclins used in the experiments shown in Figs. 2–5. To determine whether the
EP-procyclin repertoire was identical in all cells in the population, we examined 10 clones from this population obtained at New York University and 1 additional clone obtained at Johns Hopkins. All 11 clones had virtually identical procyclin compositions (Fig. 6). They also resembled that of the 1988 strain (Fig. 2A), except that we could barely detect GPEET-procyclin species and the amount of EP3 procyclin was slightly more abundant (Fig. 6). These results indicate that there is little variation from cell to cell in the original population. We also studied whether or not the EP-procyclin repertoire changed with time in culture. In cells from cultures passaged for 2 years at New York University or 6 months at Johns Hopkins, we found that the repertoire was nearly identical to that in the 1988 strain. In cells from cultures passaged for 2 years at New York University and 1 additional clone obtained at Johns Hopkins, we found that the repertoire was nearly identical to that in the 1988 strain. However, in both cultures the GPEET-procyclin had virtually disappeared (not shown).

**DISCUSSION**

Since the discovery of *T. brucei* procyclin proteins (3, 29) and their encoding genes (4, 5, 11), there has been extensive study of the expression and regulation of the genes, their GPI anchors, and their N-glycan (13–15, 18, 19). However, there has been little characterization of the different procyclin polypeptides expressed on the cell surface. Of the various polypeptides predicted from the gene sequences, only an unfraccionated mixture of the EP-procyclin species (3, 15, 29, 30) and, more recently, GPEET-procyclin (13, 14, 18–20) have been demonstrated to exist as proteins. The identification of individual EP-procyclin polypeptides has been extremely difficult because they have very similar sequences and therefore are not resolved by SDS-polyacrylamide gel electrophoresis or other fractionation techniques, even after removal of their highly heterogeneous GPI anchors. Fractionation is made even more difficult because the proteins are almost impossible to detect by conventional protein stains or by absorption at 280 nm. It is, however, possible to detect the unfraccionated EP-procyclin species and GPEET-procyclin using specific monoclonal antibodies (3, 14, 20).

Using MALDI-TOF-MS we have been successful in identifying all of the procyclin species (both the EP and GPEET forms) in a clone of *T. brucei* strain 427. We used HF treatment to remove the GPI anchors and then, in a key reaction, used mild acid hydrolysis to cleave selectively the products of the EP3, EP1-2, and EP1-1 genes at Asp-Pro sequences. We found that the 427 clone expresses the products of three genes encoding EP forms (EP3, EP1-2, and EP1-1) as well as the product of the gene encoding GPEET-procyclin (GPEET). Surprisingly, we detected only a trace of EP-procyclin encoded by one of the EP2 genes, the EP2-2.

We detected several fragments derived from the full-length polypeptides. For all EP-procyclin species, we found fragments missing 10 residues from the N termini. These fragments varied in abundance in different preparations, and control experiments indicated that they probably formed during the aq.HF treatment by cleavage at the Thr<sup>10</sup>-Lys<sup>11</sup> bond. Consistent with this hypothesis, we detected no N-terminal fragment lacking this region in the ESI-mass spectrum, because this preparation had not been aq.HF-treated (not shown). In the case of GPEET-procyclin, we found polypeptides missing four residues from the N terminus. This fragment had been detected previously by MS (19) and also by Edman degradation (13, 14, 18). Because the protein used for Edman sequencing was intact, this fragment is not dependent on HF treatment and is probably present in the cell.

In agreement with previous results (18–20), we detected phosphorylation of GPEET-procyclin. Previous studies had revealed that six out of seven threonine residues in this protein are phosphorylated (19), but we detected only a single phosphate, presumably because of the longer duration of our aq.HF treatment (16 h compared to 8 h in the previous study). We did not detect phosphorylation of EP-procyclins in the mass spectra presented in this paper. We did detect singly phosphorylated
EP-procycins of Trypanosoma brucei

EP-procycins are the major procyclins expressed by the parasite Trypanosoma brucei. They are glycoproteins that play a crucial role in the parasitic life cycle, particularly in the bloodstream and procyclic forms. The expression of these proteins is regulated by complex mechanisms that include both transcriptional and post-transcriptional factors.

**Phosphorylation and Deglycosylation**

Phosphorylation of procyclins has been observed, with EP1-2 and EP1-1 containing a single N-glycan site. However, EP3 lacks Asp-Pro bonds, making it the only nonglycosylated EP-procycin.

**Regulatory Mechanisms**

The regulation of procyclin expression is a complex process that occurs at different levels. It is influenced by various factors such as the species, the stage of development, and the environment. The presence of a C-terminal fragment containing 21 EP repeats suggests the presence of another EP3 gene product.

**Allelic Variability**

Allelic variability has been demonstrated, with some clones expressing low levels of GPEET-procyclicin (GPEET) and others expressing high levels (13, 18). Allelic variability could aid in the determination of structure-function relationships.

**Acknowledgments**

We thank Terry Pearson, Kim Paul, Yasu Morita, Isabel Roditi, Jim Morris, David Jiang, and Peter Butikofer for their contributions.

---

**References**

1. Acosta-Serrano, R. N. Cole, A. Mehlert, J. D. Bangs, M. A. Ferguson, and P. T. Englund, manuscript in preparation.

2. A. Acosta-Serrano and P. T. Englund, unpublished observation.

---
helpful discussions and critical reading of the manuscript. We also thank Isabel Roditi and Christine Clayton for helpful suggestions on the procyclin nomenclature.

REFERENCES
1. Vickerman, K. (1969) J. Cell Sci. 5, 163–193
2. Cross, G. A. M. (1975) Parasitology 71, 393–417
3. Richardson, J. P., Beecroft, R. P., Tolson, D. L., Liu, M. K., and Pearson, T. W. (1988) Mol. Biochem. Parasitol. 31, 203–216
4. Mowatt, M. R., and Clayton, C. E. (1987) Mol. Cell. Biol. 7, 2838–2844
5. Roditi, I., Carrington, M., and Turner, M. (1987) Nature 325, 272–274
6. Rudenko, G., Le Blancq, S., Smith, J., Lee, M. G., Rattray, A., and Van der Ploeg, L. H. (1990) Mol. Cell. Biol. 10, 3492–3504
7. Dorn, P. L., Aman, R. A., and Boothroyd, J. C. (1991) Mol. Biochem. Parasitol. 44, 133–139
8. Jackson, D. G., Smith, D. K., Luo, C., and Elliott, J. F. (1993) J. Biol. Chem. 268, 1894–1900
9. Clayton, C. E., Fueri, J. P., Itzhaki, J. E., Bellofatto, V., Sherman, D. R., Wisdom, G. S., Vijayasarathy, S., and Mowatt, M. R. (1990) Mol. Cell. Biol. 10, 3036–3047
10. Rudenko, G., Bishop, D., Gottesdiener, K., and Van der Ploeg, L. H. (1989) EMBO J. 8, 4259–4263
11. Mowatt, M. R., Wisdom, G. S., and Clayton, C. E. (1989) EMBO J. 8, 4259–4263
12. Butikofer, P., Ruepp, S., Boschung, M., and Roditi, I. (1997) Mol. Biochem. Parasitol. 91, 131–143
13. Field, M. C., Menon, A. K., and Cross, G. A. M. (1991) EMBO J. 10, 2731–2739
14. Schurch, N., Schurch, N., Kurath, U., and Roditi, I. (1997) Mol. Biochem. Parasitol. 89, 109–121
15. Wilson, K., Uyetake, L., and Boothroyd, J. (1999) Exp. Parasitol. 91, 222–230
16. Hotz, H. R., Biebinger, S., Flaspohler, J., and Clayton, C. (1998) Mol. Biochem. Parasitol. 91, 131–143
17. Chung, H. M., Lee, M. G., Dietrich, P., &amp; Huang, J., and Van der Ploeg, L. H. (1993) Mol. Cell. Biol. 13, 3734–3743
18. Butikofer, P., Ruepp, S., Boschung, M., and Roditi, I. (1997) Biochem. J. 326, 415–423
19. Mehlert, A., Treumann, A., and Ferguson, M. A. (1999) Mol. Biochem. Parasitol. 98, 291–296
20. Butikofer, P., Vassella, E., Ruepp, S., Boschung, M., Civenni, G., Seebeck, T., Hemphill, A., Mockherye, N., Pearson, T. W., and Roditi, I. (1999) J. Cell Sci. 112, 1765–1775
21. Ruepp, S., Furger, A., Kurath, U., Renggli, C. K., Hemphill, A., Brun, R., and Roditi, I. (1997) J. Cell Biol. 137, 1369–1379
22. Lee, M. G. (1996) Mol. Cell. Biol. 16, 1229–1230
23. Brun, R., and Shonenberger, M. (1979) Acta Tropica 36, 289–292
24. Ferguson, M. A. J. (1993) in Glycobiology: A Practical Approach (Fukuda, M., and Kobata, A., eds) pp. 349–383, IRL Press, Oxford
25. Stebeck, C. E., Beecroft, R. P., Singh, B. N., Jardim, A., Olafson, R. W., Taylor, C., Prenest, K. D., and Pearson, T. W. (1995) Mol. Biochem. Parasitol. 71, 1–13
26. Ferguson, M. A. J., Hemans, S. W., Dwek, R. A., and Rademacher, T. W. (1988) Science 239, 753–759
27. Butikofer, P., Boschung, M., and Menon, A. K. (1995) Anal. Biochem. 229, 125–132
28. Inglis, A. S. (1983) Methods Enzymol. 91, 324–332
29. Field, M. C., Menon, A. K., and Cross, G. A. M. (1991) EMBO J. 10, 2731–2739
30. Field, M. C., Menon, A. K., and Cross, G. A. M. (1991) EMBO J. 10, 2731–2739
31. Wilson, K., Uyetake, L., and Boothroyd, J. (1999) Exp. Parasitol. 91, 222–230
32. Hotz, H. R., Biebinger, S., Flaspohler, J., and Clayton, C. (1998) Mol. Biochem. Parasitol. 91, 131–143
33. Chung, H. M., Lee, M. G., Dietrich, P., &amp; Huang, J., and Van der Ploeg, L. H. (1993) Mol. Cell. Biol. 13, 3734–3743
34. Roditi, I., and Clayton, C. (1999) Mol. Biochem. Parasitol., in press

EP-procyclins of Trypanosoma brucei 29771