Molecular Systems Biology 5; Article number 261; doi:10.1038/msb.2009.13
Citation: ***Molecular Systems Biology*** **5:261**

www.molecularsystemsbiology.com

molecular systems biology

GPIomics: global analysis of glycosylphosphatidylinositol-anchored molecules of *Trypanosoma cruzi*

Ernesto S Nakayasu[1], Dmitry V Yashunsky[2,4], Lilian L Nohara[1], Ana Claudia T Torrecilhas[3], Andrei V Nikolaev[2] and Igor C Almeida[1,*]

[1] Department of Biological Sciences, The Border Biomedical Research Center, University of Texas at El Paso, El Paso, TX, USA, [2] Division of Biological Chemistry and Drug Discovery, College of Life Sciences, The Wellcome Trust Biocentre, University of Dundee, Dundee, UK and [3] Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil
[4] On leave from Research Institute of Biomedical Chemistry RAMS, Moscow 119121, Russia.
* Corresponding author. Department of Biological Sciences, The Border Biomedical Research Center, University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968, USA. Tel.: +1 915 747 6086; Fax: +1 915 747 5808; E-mail: icalmeida@utep.edu

Received 5.9.08; accepted 23.2.09

Glycosylphosphatidylinositol (GPI) anchoring is a common, relevant posttranslational modification of eukaryotic surface proteins. Here, we developed a fast, simple, and highly sensitive (high attomole-low femtomole range) method that uses liquid chromatography-tandem mass spectrometry (LC-MS^n) for the first large-scale analysis of GPI-anchored molecules (i.e., the GPIome) of a eukaryote, *Trypanosoma cruzi*, the etiologic agent of Chagas disease. Our genome-wise prediction analysis revealed that approximately 12% of *T. cruzi* genes possibly encode GPI-anchored proteins. By analyzing the GPIome of *T. cruzi* insect-dwelling epimastigote stage using LC-MS^n, we identified 90 GPI species, of which 79 were novel. Moreover, we determined that mucins coded by the *T. cruzi* small mucin-like gene (TcSMUG S) family are the major GPI-anchored proteins expressed on the epimastigote cell surface. TcSMUG S mucin mature sequences are short (56–85 amino acids) and highly *O*-glycosylated, and contain few proteolytic sites, therefore, less likely susceptible to proteases of the midgut of the insect vector. We propose that our approach could be used for the high throughput GPIomic analysis of other lower and higher eukaryotes.
Molecular Systems Biology 7 April 2009; doi:10.1038/msb.2009.13
Subject Categories: membranes and transport; microbiology and pathogens
Keywords: global analysis; glycobiology; glycosylphosphatidylinositol; GPIomics; protein post-translational modifications

Introduction

Glycosylphosphatidylinositol (GPI) anchoring is a ubiquitous posttranslational modification (PTM) of proteins in lower and higher eukaryotes (McConville and Ferguson, 1993). In mammals, GPI biosynthesis is vital for embryonic development, and GPI-anchored proteins participate in important biological processes such as cell–cell interactions, signal transduction, endocytosis, complement regulation, and antigenic presentation (McConville and Ferguson, 1993; Orlean and Menon, 2007; Paulick and Bertozzi, 2008). In lower eukaryotes, particularly yeast and protozoa, GPI-anchored molecules have also been shown to have important biological functions (Ferguson, 1999; McConville and Menon, 2000; Orlean and Menon, 2007). In pathogenic protozoan parasites (e.g., *Trypanosoma cruzi*, *Trypanosoma brucei*, *Leishmania major*, and *Plasmodium falciparum*), for instance, GPI-anchored glycoconjugates may extensively coat the plasma membrane and are involved in many aspects of host–parasite interactions, such as adhesion and invasion of host cells, modulation and evasion from host immune response, and pathogenesis (McConville and Ferguson, 1993; Ferguson, 1999; McConville and Menon, 2000; Buscaglia *et al*, 2006; Gazzinelli and Denkers, 2006; Acosta-Serrano *et al*, 2007).

In *T. brucei*, the GPI biosynthesis has already been validated as a molecular target for development of new drugs against African sleeping sickness (Smith *et al*, 2004). To exploit GPIs as targets for the development of new therapies against other endemic protozoa (e.g., *T. cruzi*, *P. falciparum*, *Leishmania* spp.), a detailed, large-scale analysis of the GPI-anchored molecules expressed on the cell surface of these parasites (i.e., the GPIome) is of paramount importance. However, one of the

major hurdles to study the large-scale temporal and spatial expression of GPI-anchored molecules in eukaryotic cells is the dearth of a universal and straightforward approach for GPI analysis (Ferguson, 1992; Hooper, 2001). The difficulty to develop a universal method to analyze GPI and GPI-anchored proteins may be due to their complex structure. The general structure of a GPI anchor comprises a lipid tail containing either a phosphatidylinositol (PI) or an inositolphosphorylceramide (IPC) moiety, attached to a glycan core consisting of a glucosamine (GlcN) residue followed by three mannose (Man) residues. In the third Man distal from the GlcN residue, usually an ethanolaminephosphate (EtNP) group attaches the GPI to the C-terminus of the protein. Further modifications in the anchor may occur, such as extra glycan, EtNP and/or aminoethylphosphonate (AEP) residues substituting the glycan core, and/or an extra fatty acid (acyl) group attached to the *myo*-inositol ring, increasing the complexity of the GPI structure (Ferguson, 1999; McConville and Ferguson, 1993). Owing to their complex structure and amphiphilic nature, GPIs and GPI-anchored proteins are somewhat difficult to be extracted, purified, and fully characterized.

Elortza *et al* (2003) applied an approach to release GPI-anchored proteins from the plasma membrane by treatment with PI-specific phospholipases. Enzyme-treated and control samples were then separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and the differential bands were excised and analyzed by mass spectrometry (MS) (Elortza *et al*, 2003). Although this approach proved to be valuable in identifying GPI-anchored proteins, it did not provide any information on the GPI anchor structure itself. Another useful but somewhat limited approach is the extraction of GPI-anchored proteins with neutral detergents or 9% aqueous (aq.) *n*-butanol and further purification by hydrophobic interaction (HIC) or reverse phase chromatography (RPC) (Ferguson, 1992; Hooper, 2001). Both HIC and RPC methods separate biomolecules according to differences in their hydrophobicity. However, in contrast to RPC, HIC uses less denaturing buffers and/or organic solvents, and higher initial salt concentration to enhance the hydrophobic interactions between the biomolecule and the chromatography medium (stationary phase). On the other hand, RPC medium is generally more hydrophobic than that of a HIC medium, leading to stronger, more selective interactions between the biomolecule and the medium, and successful elution is achieved by the use of increasing concentration of organic solvent. A major drawback is that most stationary phases used for HIC and RPC have amphiphilic nature (i.e., hydrophobic groups attached to hydrophilic beads) and, therefore, their resolution is rather low. This might be due to the interaction of the resin with both hydrophobic (lipid) and hydrophilic (glycan) moieties of the GPI. Thus, here we hypothesize that the use of a more hydrophobic, polystyrene-divinylbenzene-based resin for RPC (e.g., POROS R1) (DePhillips *et al*, 1994; Whitney *et al*, 1998) could result in a much stronger, more selective interaction between the GPI lipid moiety and the covalently attached functional group (e.g., C4) of the stationary phase. This could considerably enhance the resolution of the process of purification of GPI-anchored molecules and free GPIs.

T. cruzi is the etiologic agent of Chagas disease or American trypanosomiasis, a neglected tropical disease that affects over 11 million people and causes an estimated 50 000 annual deaths in Latin America (Dias *et al*, 2002; Barrett *et al*, 2003; Moncayo and Ortiz Yanine, 2006). Lately, Chagas disease has become a public health threat for the US and European countries like Spain, where an increasing number of chronically *T. cruzi*-infected migrants from endemic countries are residing in (Bern *et al*, 2007; Piron *et al*, 2008). There are only two commercial drugs (Benznidazole and Nifurtimox) available for the treatment of Chagas disease, and both are partially effective and highly toxic. In addition, an increasing number of drug-resistant *T. cruzi* strains have been reported (Urbina and Docampo, 2003; Wilkinson *et al*, 2008). Moreover, thus far no human vaccine is available for treating or preventing Chagas disease (Garg and Bhatia, 2005; Dumonteil, 2007; Hotez *et al*, 2008). Therefore, there is an urgent need for new therapeutic targets against *T. cruzi*.

T. cruzi has four developmental stages or forms, two (i.e., epimastigote and metacyclic trypomastigote) dwelling in the triatomine insect vector, and two (i.e., amastigote and trypomastigote) in the mammalian host. The parasite can be transmitted by the bloodsucking insect vector (a Reduviidae, popularly known as the kissing bug), blood transfusion, organ transplantation, or congenitally. The vector-mediated transmission begins when metacyclic trypomastigotes present in the vector's excrement enter the host bloodstream through the insect's bite wound or exposed mucosal tissues. Metacyclic forms immediately invade a wide range of nucleated cells and transform into proliferative amastigote forms, which divide by binary fission and, 4–5 days later, transform into intracellular trypomastigote forms. These forms rupture the cell membrane and invade other surrounding cells, or enter the bloodstream to infect remote organs or tissues, or eventually another insect vector (Barrett *et al*, 2003).

T. cruzi GPI-anchored proteins are expressed in all developmental stages and encoded by thousands of members of multigene families, such as *trans*-sialidase (TS)/gp85 glycoprotein, mucin, mucin-associated surface protein (MASP), and metalloproteinase gp63. Some of these proteins (e.g., TS/gp85, mucins) have been shown to be essential for the infectivity of the parasite or its escape from the host immune response (Frasch, 2000; Acosta-Serrano *et al*, 2001, 2007; Buscaglia *et al*, 2006; Yoshida, 2006; Alves and Colli, 2008). In addition, GPI anchors from *T. cruzi* were shown to be strong proinflammatory molecules, being critical in the modulation of the host immune response against this parasite (Almeida and Gazzinelli, 2001; Gazzinelli and Denkers, 2006). Thus, GPI-anchored proteins and GPI anchors themselves seem to be very attractive targets for new therapies against Chagas disease.

Now we report the first large-scale analysis of the GPIome of a eukaryote, *T. cruzi*. Initially, we carried out a GPI-anchoring prediction analysis for all protein-encoding genes of the *T. cruzi* genome. Next, we developed a polystyrene-based (POROS R1) RPC to purify protein-free GPIs (glycoinositolphospholipids, GIPLs) and GPI-anchored proteins. This chromatographic step was then coupled to tandem MS, and used for the analysis of GIPLs and proteinase-released GPIs. Our results show that the GPIome of *T. cruzi* is much more complex than

was known earlier, and provide new insights into the biosynthesis of *T. cruzi* GPI anchors, which could be eventually exploited as potential therapeutic target in Chagas disease.

Results

Genome-wise prediction of GPI-anchored proteins from *T. cruzi*

To estimate the extent of *T. cruzi* proteins that might be GPI anchored, we performed a prediction analysis using the FragAnchor algorithm (Poisson *et al*, 2007). Proteins that will undergo GPI PTM contain a short hydrophobic C-terminal sequence that is removed before the *en bloc* addition of the complete GPI anchor by the transamidase complex (McConville and Ferguson, 1993; Ferguson, 1999). The newly formed C-terminal amino acid receiving the GPI is known as the ω (omega) site. The FragAnchor algorithm, thus, predicts the GPI-anchoring sites based on the short hydrophobic C-terminal amino acid sequence and the ω site. To test the sensitivity and accuracy of the prediction analysis, we used a database set containing all sequences of the *T. cruzi* mucin II (TcMUC II) family. This family of glycoproteins had their GPI-anchoring sites experimentally validated (Buscaglia *et al*, 2004) and, thus, could be used as the training set for our analysis. Out of 624 TcMUC II sequences, 598 (95.8%) were highly probable to be GPI anchored (Table I), showing that the sensitivity of the prediction is likely high. Next, to test the accuracy of the prediction, we removed a short hydrophobic sequence comprising the last 20 amino-acid residues at the C-terminus of the sequences and analyzed them with the FragAnchor algorithm. With this approach only two sequences (0.3%) were wrongly predicted to be GPI anchored (Table I), so the prediction could be considered very accurate.

The prediction analysis using the genome sequences (TcruziDB v5.0) lead to the estimation that 11.9% of the sequences have potential GPI-anchoring site (Table I). By performing the analysis using the *T. cruzi* sequences deposited on the GenBank, 11.7% of the sequences were predicted to be GPI anchored (Table I). Of those predicted GPI-anchored proteins, the members of MASP (1730 sequences), mucin (1262 sequences), and TS/gp85 (1160 sequences) multigene families were the most abundant among the sequences deposited in the GenBank. Amastin, gp63, mucin-like, and TolT were also represented by multiple sequences (Table I). TolT is an immunogenic surface protein found on *T. cruzi* trypomastigotes (Quanquin *et al*, 1999) with protein sequence identity and biochemical and biological traits in common with the bacterial TolA proteins, which are part of the Tol system of most gram-negative bacteria and are involved in the outer membrane stability (Lazzaroni *et al*, 2002). Taken together, our GPI-anchoring prediction data indicates that a high proportion of the *T. cruzi* genes possibly encode GPI-anchored proteins.

Table I Genome-wise prediction of putative GPI-anchored protein sequences in *T. cruzi*

Database	Number of sequences	Highly probable GPI-anchored sequences	% Total
TcruziDB v5.0	19 613	2304	11.9
GenBank *T. cruzi* sequences	41 478	4941	11.7
GeneDB TcMUC II sequences	624	598	95.8
GeneDB TcMUC $II^{-20\,AA}$ sequences[a]	624	2	0.3

Classification of the predicted *T. cruzi* GPI-anchored proteins from GenBank

Protein family	Highly probable GPI-anchored sequences	% Total potentially GPI-anchored proteins
MASP (mucin-associated proteins)	1730	35.0
Mucins	1262	25.5
TcMUC I	100	2.0
TcMUC II	1126	22.8
TcMUC III	2	0.0
TcSMUG S	20	0.4
TcSMUG L	14	0.3
TS/gp85 glycoproteins	1160	23.5
trans-Sialidase	1102	22.3
gp90	22	0.4
gp82	15	0.3
Complement regulatory protein	9	0.2
Tc85/gp85	8	0.2
Flagellum-associated protein	3	0.1
SAPA	1	0.0
Other families	789	16.0
Amastin	216	4.4
gp63	143	2.9
Mucin-like	86	1.7
TolT	14	0.3
Hypothetical proteins	207	4.2
Others	123	2.5
Total	4941	100

[a]These sequences had 20 amino-acid residues from the C-terminal removed before the GPI prediction analysis, and served as negative controls for the prediction.

Implementing a polystyrene-based RPC to purify free and protein-linked GPIs

Next, we developed a polystyrene-based RPC and tested it using a synthetic GPI (Yashunsky *et al*, 2006) containing several by-products resulting from unsaturated fatty acid oxidation or removal during its prolonged storage as nuclear magnetic resonance (NMR) sample in $[^2H]_6$-DMSO solution. The GPI sample was loaded onto a ziptip manufactured with POROS R1 resin (C4 groups linked to polystyrene-divinylbenzene beads) and eluted step-wisely with increasing (40, 50, 60, and 70%) 2-propanol concentrations. Each fraction was then analyzed by negative-ion mode electrospray ionization (ESI)-MS (Materials and methods). The results clearly showed an efficient separation between the *bona fide* GPI species and *lyso*-GPIs, oxidized GPIs, and other by-products (Figure 1).

Then, we tested the suitability and efficiency of POROS R1 for the purification of GPI-anchored proteins. We analyzed an organic (9% *n*-butanol) extract enriched with GPI-anchored proteins from epimastigote forms of *T. cruzi* (Almeida *et al*,

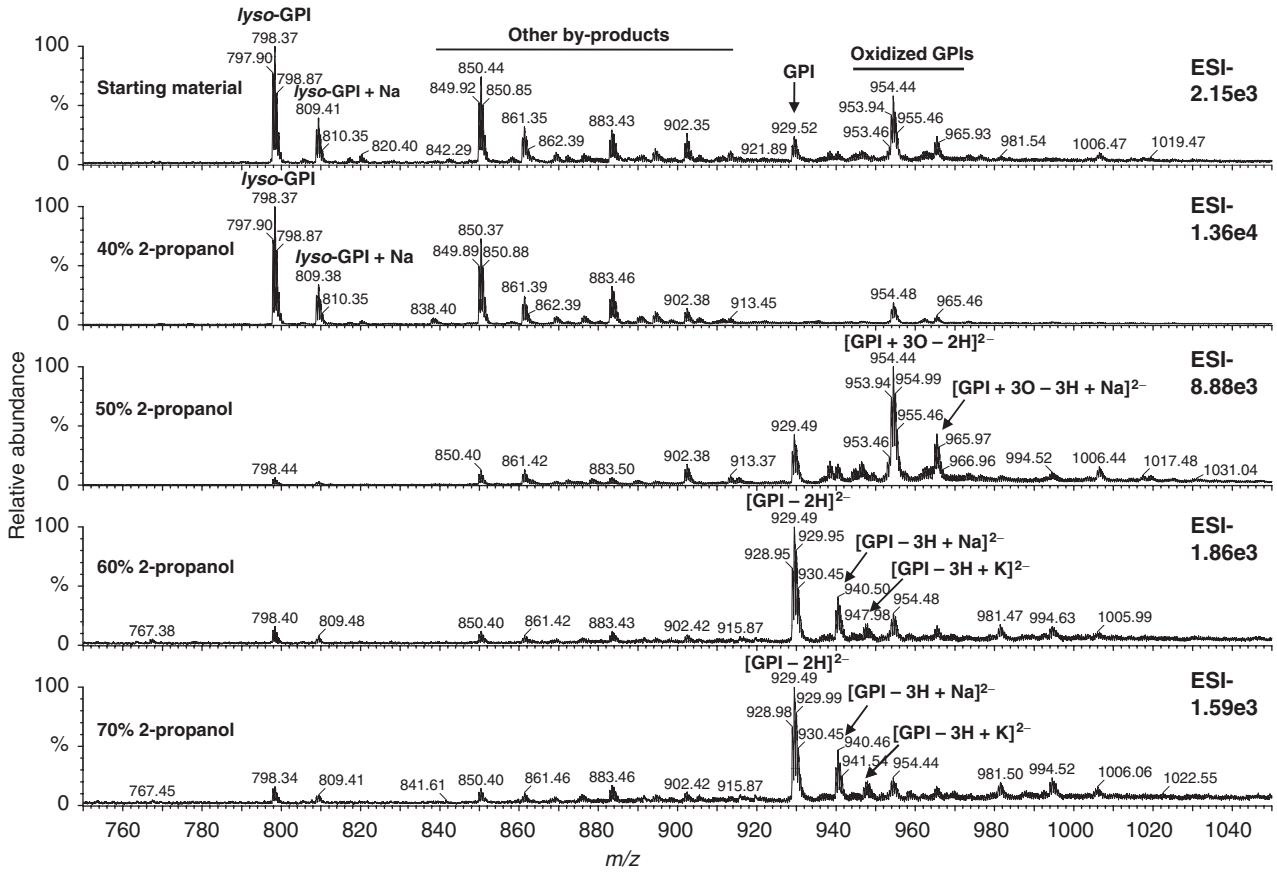

Figure 1 Fractionation of the synthetic trypomastigote mucin GPI and by-products using POROS R1 ziptip. The synthetic mucin GPI (Man-[EtNP]Man-Man_2-[AEP]GlcN-InsP-1-*O*-alkyl-C16:0-2-*O*-acyl-C18:2-Gro) was purified through POROS R1 50 resin as described in Materials and methods. The top panel represents the MS spectrum of the starting material, whereas the bottom four panels represent the MS spectra of the ziptip fractions eluted with 40, 50, 60, and 70% 2-propanol, respectively. All observed GPI species are doubly charged, plus or minus sodium adduct. *lyso*-GPI, GPI species missing the C18:2-fatty acid at C-2 of the glycerol backbone; *m/z*, mass-to-charge ratio.

2000) (Figure 2A). This extract was spiked with small amount of biotin-labeled epimastigote mucins, and fractionated in a solid-phase extraction cartridge prepared with POROS R1 resin. The fractions were then monitored by SDS–PAGE and chemiluminescent assay (CLA) (Figure 2A). By CLA, the epimastigote mucins were shown to be eluted mainly with 22.5% aq. *n*-propanol (Figure 2B). Consistently with this result, the silver-stained gel showed two broad bands in the 34–67 kDa range with typical behavior of epimastigote mucins (Acosta-Serrano *et al*, 2001; Buscaglia *et al*, 2006) (bands 9 and 10, Supplementary Figure 1). These bands were excised, digested with trypsin, and sequenced by liquid chromatography (LC) coupled to MS^2. As expected, here we were unable to detect any peptide sequence, most probably because (1) epimastigote mucins are highly glycosylated (Acosta-Serrano *et al*, 2001; Buscaglia *et al*, 2006), and (2) their protein core seems almost completely resistant to different proteases (e.g., trypsin and proteinase K) (Almeida *et al*, unpublished observation). We have observed earlier that on trypsin treatment, epimastigote mucins only slightly shift their relative migration on SDS–PAGE, suggesting that small peptide fragment(s) is(are) removed from the protein backbone (Almeida *et al*, unpublished observation). We have also reported that epimastigote mucins are only partially degraded by proteinase K, releasing the intact GPI anchor previously linked to their C-terminal end (Almeida *et al*, 2000).

To further corroborate the presence of GPI-anchored proteins, the fraction eluted from the POROS R1 column with 22.5% aq. *n*-propanol was digested with proteinase K and analyzed by negative ion-mode ESI-MS. We could detect at least four ion species with *m/z* identical to purified eMUC-derived GPIs (eMUC-GPIs) (Previato *et al*, 1995; Serrano *et al*, 1995; Almeida *et al*, 2000) that were further confirmed by fragmentation analysis (Figure 2C and D). Taken together, our results validate the use of POROS R1 for the purification of free GPIs and GPI-anchored proteins.

Analysis of protein-free GPIs by LC-MS^n

To ultimately increase the sensitivity and the speed of our analysis, we packed a nanocapillary column with POROS R1 resin and coupled it to an LC-MS system. Then, we performed LC-MS^2 and LC-MS^3 analysis of organic (91% *n*-butanol, and combined chloroform/methanol and chloroform/methanol/water) extracts enriched for GIPLs (Figure 3A). GPIs are frequently analyzed in negative-ion mode ESI-MS, dissolved in

solutions/buffers with neutral pH. However, under these conditions, the silica tubing has negative charges, which interact with amine (GlcN, AEP, and/or EtNP) groups present in the GPI. To neutralize the negative charge of the silica tubing, we analyzed our GPI samples in positive-ion mode ESI-MS, in the presence of 0.2% formic acid (FA), used as the ion pair for the LC-MS analysis. With the LC-MS approach, we could detect and characterize at least 78 doubly charged ion species of GIPLs (Table II; Supplementary Table I; Supplementary Figure 2A–C). Of those, 70 were found to be novel species

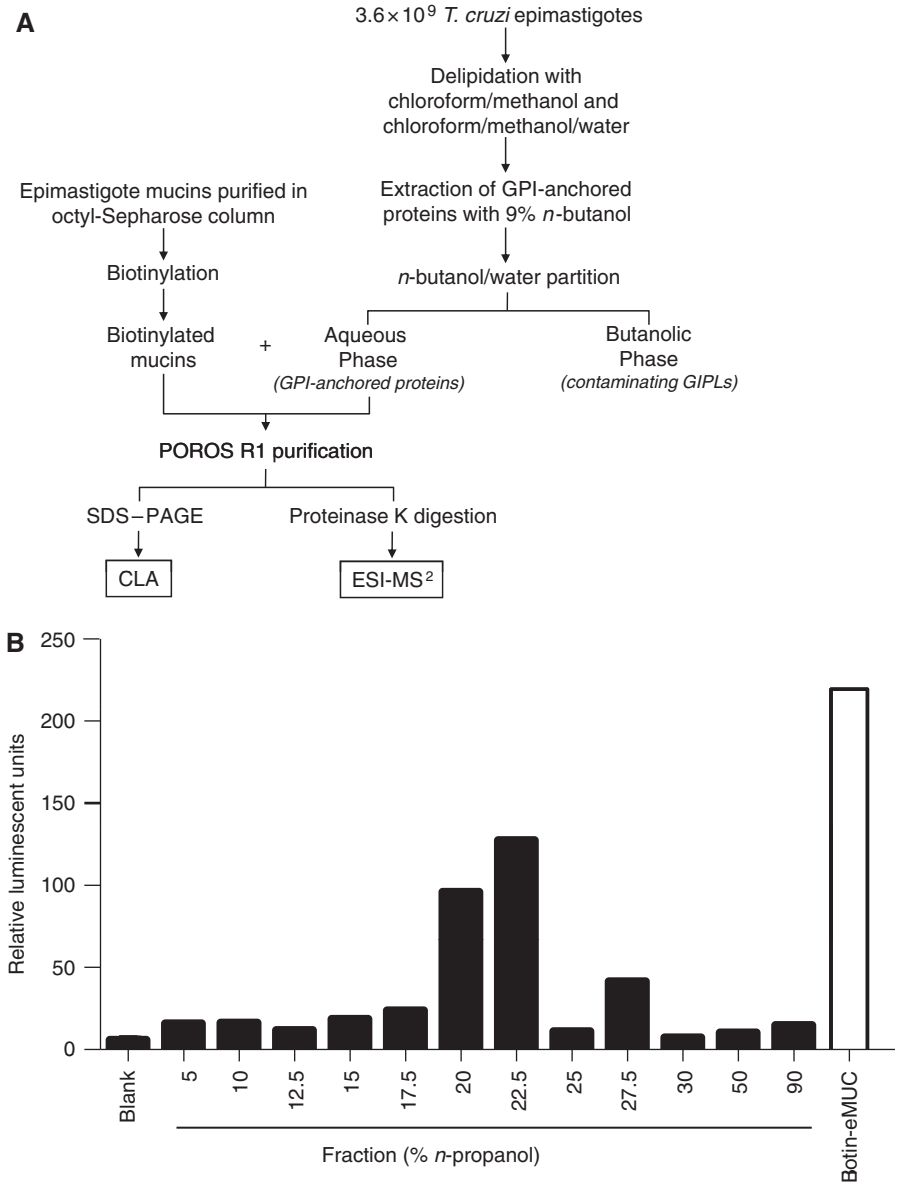

Figure 2 Purification of GPI-anchored proteins from *T. cruzi* epimastigotes using POROS R1 50 cartridge. (**A**) Schematic representation of the methodology used for the purification and analysis of GPI-anchored proteins. Briefly, epimastigote cells were delipidated with chloroform/methanol (1:1, v/v), followed by chloroform/methanol/water (1:2:0.8, v/v/v). Then, GPIs were extracted with 9% *n*-butanol and partitioned with *n*-butanol/water solution. The aqueous phase, rich in GPI-anchored proteins, was spiked with biotinylated epimastigote mucins (biotin-eMUC), purified in an octyl-Sepharose column. The GPI-rich extract was fractionated in a POROS R1 50 solid-phase extraction cartridge and the eluted fractions were analyzed by chemiluminescent assay (CLA) and SDS–PAGE, or digested with proteinase K and analyzed by electrospray ionization-tandem mass spectrometry (ESI-MS/MS). (**B**) CLA of POROS R1 50 fractions. Biotin-eMUC, biotinylated epimastigote mucins. (**C**) The fraction eluted with 22.5% *n*-propanol (equivalent to 2.7×10^7 epimastigote cells) was digested with proteinase K, purified in a POROS R1 50 ziptip, and analyzed in negative-ion mode using an ESI-QTOF-MS (Micromass Qtof-1, Waters). The top panel represents the MS spectrum of the purified epimastigote mucin (p-eMUC-GPI) fraction, whereas the bottom panel is the MS spectrum of the synthetic *T. cruzi* epimastigote mucin GPI (s-eMUC-GPI) (Man-[EtNP]Man-Man_2-GlcN[AEP]-InsP-1-*O*-C16:0-alkyl-2-*O*-C16:0-acyl-Gro) (Yashunsky *et al*, 2006). The asterisk indicates another species of GPI, previously characterized as [AEP]Man_4-GlcN[AEP]-InsP-1-*O*-alkyl-C16:0-2-*O*-acyl-C16:0-Gro (Almeida *et al*, 2000). (**D**) ESI-QTOF-MS^2 spectrum of parent ion at *m/z* 917.0. The top panel represents the MS^2 spectrum of the purified *T. cruzi* epimastigote mucin GPIs (p-eMUC-GPIs), whereas the bottom panel represents the MS^2 spectrum of the synthetic eMUC-GPI (s-eMUC-GPI) purified by POROS R1. Noteworthily, the same ion species are observed in both spectra, confirming the assignment of the major parent ion at *m/z* 917.03 in (C). AAG, alkylacylglycerol; AEP, aminoethylphosphonate; Alk, alkyl; Gro, glycerol; GroP, 3-*O*-phosphoglycerol; Hex, hexose; HexN, hexosamine; Ins, *myo*-inositol; InsP, *myo*-inositolphosphate; M, molecular mass; *m/z*, mass-to-charge ratio; PI, phosphatidylinositol.

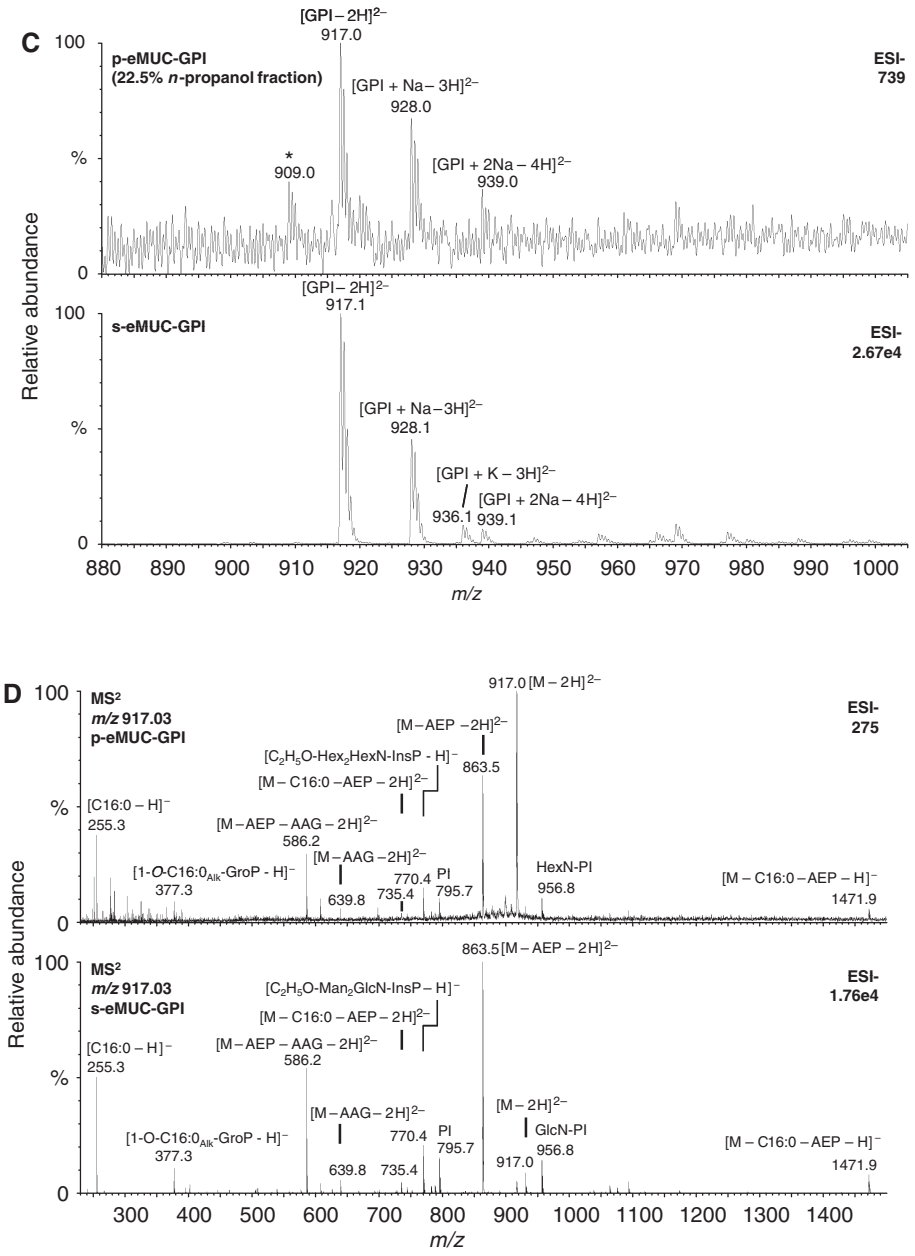

Figure 2 Continued.

(Supplementary Table I). The extracted-ion chromatograms for GIPLs obtained by either extraction procedure (9% *n*-butanol or combined chloroform/methanol and chloroform/methanol/water) (Figure 3A) clearly indicate that the majority (92–98%) of GIPLs contain ceramide (Cer) and only a small fraction (2–8%) have alkylacylglycerol (AAG) in the lipid moiety (Table II). In MS^2, typically the fragmentation (collision-induced dissociation, CID) of the doubly charged GIPL parent-ion gave rise to fragment- or daughter-ions corresponding to the loss of Cer ($[M-Cer+H]^+$), and to a series of ions containing AEP or EtNP residue attached to multiple (n=1–5) hexose (Hex) residues (e.g., Supplementary Figure 2C3, 2C5, 2C7, and 2C9). On the other hand, fragment-ions corresponding to the neutral loss of Hex residue(s) and AEP or EtNP were also highly abundant. We could also find a fragment corresponding to the inositolphosphate (InsP) attached to AEP-HexN (AEP-HexN-InsP) at *m/z* 529.3 (e.g., Supplementary Figure 2C3, 2C5, 2C7, and 2C9). Interestingly, some species had only one AEP residue and no EtNP. In these species, the AEP residue could be attached to either a Hex or HexN residue. For instance, the fragmentation of the GIPL species AEP-Hex_6-HexN-InsP-C16:0/d18:0-Cer observed at *m/z* 1012.49 showed diagnostic fragment ions corresponding to the AEP-HexN attached to IPC (AEP-HexN-IPC) (*m/z* 1050.48) and Hex_2-AEP (*m/z* 432.22) (Supplementary Figure 2C22).

Still in the case of ceramide-containing GIPLs, the two most abundant daughter-ions observed by MS^2 analysis were the one resulting from the neutral loss of ceramide

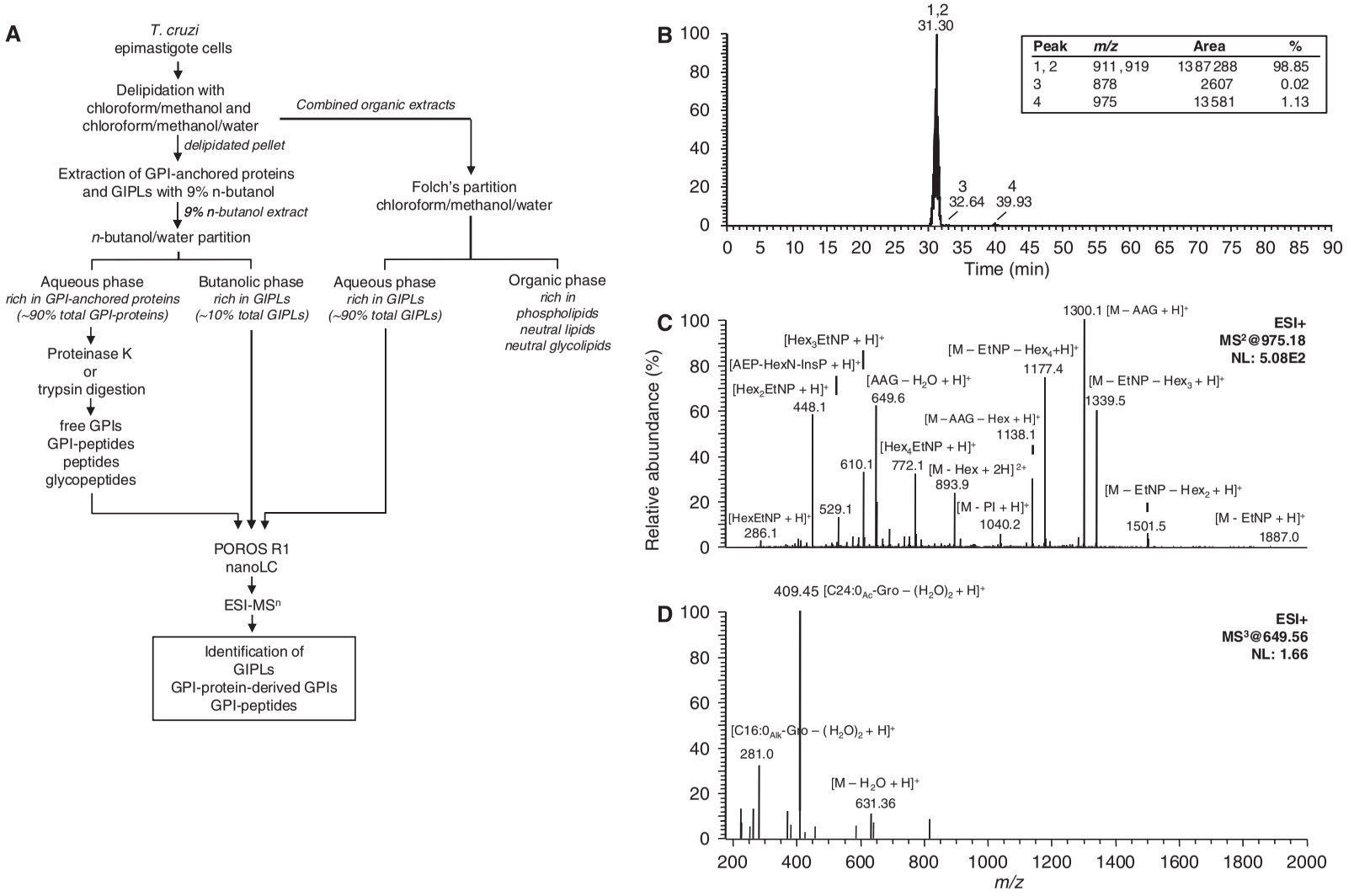

Figure 3 (**A**) General GPIomic approach. Organic extracts rich in GIPLs and GPI-anchored proteins were obtained as described in Materials and methods. The fraction rich in GIPLs was directly analyzed by LC-MS^2 and LC-MS^3, whereas the fraction rich in GPI-anchored proteins were digested with proteinase K or trypsin before LC-MS^2 and LC-MS^3 analyses. In both cases, the LC step was carried out using a POROS R1 10 column (75 μm × 10 cm) and the MS analysis was performed using a LTQXL ESI-linear ion-trap-MS. The resulting spectra were analyzed manually for annotation and assignment of the GPI species. (**B**) Data-dependent acquisition (DDA) (no dynamic exclusion enabled) LC-MS^2 analysis of epimastigote GPIs. The extracted-ion chromatogram was plotted for four major GPI species. The plotted ion species correspond to the dehydrated alkylacylglycerol (AAG—H_2O) and the loss of alkylacylglycerol moiety (M—AAG). The insert shows details about the *m/z* and the area of each peak. (**C**) Annotated MS^2 spectrum from GPI structure at *m/z* 975.18 corresponding to EtNP-Hex_4-[AEP]HexN-InsP-1-*O*-alkyl-C16:0-2-*O*-acyl-C24:0-Gro. (**D**) MS^3 spectrum for the dehydrated AAG fragment at *m/z* 649.56. (**E**) Proposed fragmentation and structure of the novel GPI species (EtNP-Hex_4-[AEP]HexN-InsP-1-*O*-alkyl-C16:0-2-*O*-acyl-C24:0-Gro) observed at *m/z* 975.18. Ac, acyl; Gro, glycerol.

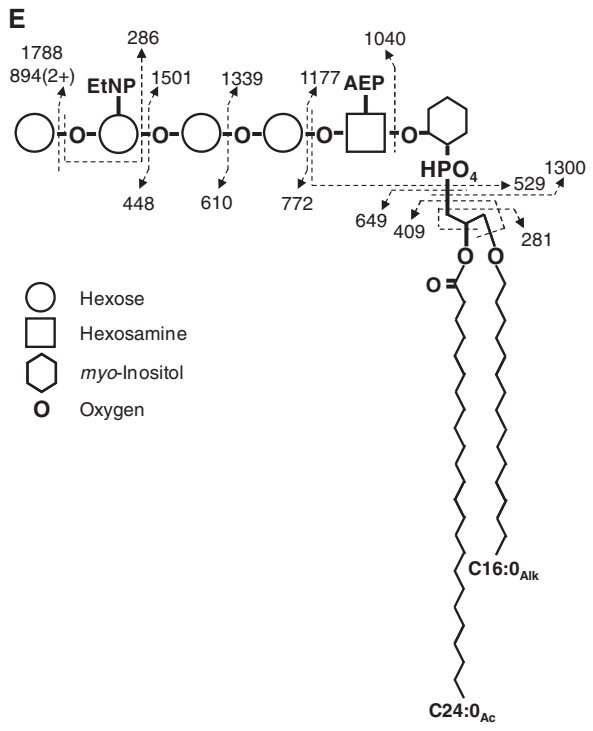

Figure 3 Continued.

([M−Cer + H]$^+$) and the dehydrated ceramide moiety itself ([Cer−H_2O + H]$^+$) (Supplementary Figure 2C3). Thus, to obtain more detail of the ceramide moiety, we performed MS^3 analysis of the two most abundant ions resulting from the loss of H_2O (at *m/z* 520.56 and 502.49). The ion species at *m/z* 520.56 gave rise to a C16:0-1-azirine derivative at *m/z* 280.24 and a dehydrated sphingoid (d18:1)-base derivative at *m/z* 264.41 (Hsu and Turk, 2000) (Supplementary Figure 2C4). 1-Azirine fatty acid derivatives were also frequent in MS^2 spectra (e.g., Supplementary Figure 2C3, 2C7, 2C9, and 2C12), thus, the structure of many ceramides could be determined without the need of MS^3 analysis. Most of the ceramides were dihydroxylated, but some of them had three hydroxyl groups instead (Table II; Supplementary Table I). The sphingoid base was invariably d18:0, d18:1, or d18:2, whereas the fatty acid varied in length from C16:0 to C26:0. Surprisingly, we found a few ceramide species containing odd-carbon number fatty acids (OCFAs), such as C23:0 and C25:0 (e.g., species 31, 33, 36, 40, 47, 60, 62–64, 66, and 67; Supplementary Table I). Interestingly, OCFAs have been described earlier in *T. cruzi* neutral glycosphingolipids (Barreto-Bergter *et al*, 1992). Other trypanosomatids, such as *Phytomonas* and *Herpetomonas*, have also been reported to biosynthesize *de novo* a range of *iso*-branched C18, C20, and C22 polyunsaturated fatty acids (Fish *et al*, 1982). In our case, we could not determine whether the observed OCFAs were linear or methyl-branched, or both. Furthermore, considering the calculated monoisotopic molecular mass, we cannot exclude the possibility that these fatty acids are actually oxidized C22:1 and C24:1 species. However, owing to their late retention time in the RPC (Supplementary Table I), more likely these fatty acids are OCFAs. Further studies are necessary to confirm the existence and precise structure of these unusual fatty acid species in *T. cruzi*.

A few GIPL species containing AAG as the lipid tail were also found (Table II; Supplementary Table I; Supplementary Figure 2A and B). Unlike the ceramide-containing GIPLs, MS^2 analysis of AAG-containing GIPL species revealed that fragments derived from the lipid tail were not always the most abundant ones (Supplementary Figure 2C25, 27–35). Therefore, we could not carry out data-dependent MS^3 analysis of all AAG-containing GIPL species. Trying to circumvent this issue, we performed data-independent MS^3 analysis by selecting the most abundant and frequent lipid moiety-derived ion species observed at *m/z* 537.6 (Supplementary Figure 2C27–30). Again, we observed that fragment ions derived from the AAG moiety had very low abundance and we could not obtain enough information to determine the fine composition of the lipid tail (data not shown). Thus, we annotated the less abundant AAG-containing GIPL species (e.g., C34:0-AAG, C34:1-AAG, C34:2-AAG, and C42:1-AAG) with the total number of carbons instead (Table II; Supplementary Table I). However, as in *T. cruzi* GPIs all 1-*O*-alkyl chains thus far described are C16:0 (Serrano *et al*, 1995; Carreira *et al*, 1996; Almeida *et al*, 2000; Previato *et al*, 2004), it is most likely that most if not all AAG-containing GIPL species observed in this study would have the same C16:0-alkyl chain composition.

To relatively quantify the GIPL species, we exploited the MS^2 total-ion chromatogram (MS^2 TIC) approach (Asara *et al*, 2008) (Materials and methods, and below). The quantification revealed that GIPL species containing ceramide (91.6–97.9%) are much more abundant than those having AAG (2.1–8.4%) as lipid tail (Table II; Supplementary Table I; Supplementary Figure 2).

Analysis of proteinase K-released GPIs by LC-MS^n

Next, the fraction rich in GPI-anchored proteins was digested with proteinase K and analyzed by LC-MS^2 and LC-MS^3 (Figure 3A). Proteinase K has a broad range of specificity and cleaves at the carboxyl terminus of the polypeptide, thus ensuing GPIs that are free of amino-acid residues (Almeida *et al*, 2000). The fragmentation pattern of proteinase K-released GPIs was very similar to that of the GIPLs (Figure 3C–E; Supplementary Figure 3). In fact, some of the AAG-containing GPIs had the same glycan and lipid structure as some of the GIPLs (Table II). Thus, we cannot discard the possibility that these AAG-containing GIPLs could be actually direct precursors of GPI anchors to be attached to proteins. It is also worth pointing out that all GPI structures derived from GPI-anchored proteins had AAG as the lipid moiety. Conversely, the majority (91.4–97.8%) of GIPLs had ceramide as the lipid tail (Table II; Supplementary Table I). It still remains to be determined why epimastigotes, in contrast to metacyclic trypomastigotes (Serrano *et al*, 1995), do not express ceramide-containing GPI anchors in their major surface glycoproteins (e.g., mucins). Furthermore, we could not detect any major GIPL species in the proteinase K-released GPI analysis, indicating that the sample was free of these glycolipids. The major fragments from MS^3 analysis of AAG were the ones resulting from the loss of either alkyl or acyl group. Noteworthily, all AAG structures we could determine had

Table II Summary of glycan, lipid, and peptide composition of the GPIome of *T. cruzi* epimastigotes characterized by LC-MS/MS

Overall data	GPI-anchored molecule (origin or treatment)[a]			
	GIPLs (CM and CMW)	GIPLs (BuOH)	Proteins (Proteinase K)	Proteins (Trypsin)
Inositol content (picomoles per 1e7 cells)	889.95 ± 49.02[b]	9.17 ± 1.51	0.43 ± 0.002	0.43 ± 0.002
Number of MS^2 spectra	1162	196	54	25
Number of identified species	70	44	11	5
	Relative quantification by MS2 TIC (% total)[c]			
Glycan moiety				
[AEP]Hex_4-[AEP]HexN-InsP	—	—	***14.31***	***11.43***
[AEP]Hex_5-[AEP]HexN-InsP	***17.61***	***46.24***	—	3.82
[AEP]Hex_6-[AEP]HexN-InsP	0.16	1.16	—	—
[EtNP]Hex_4-[AEP]HexN-InsP	0.12	0.01	***83.97***	***56.50***
[EtNP]Hex_5-[AEP]HexN-InsP	***26.68***	***35.47***	1.36	***19.93***
[EtNP]Hex_6-[AEP]HexN-InsP	0.09	0.17	0.18	—
[EtNP]Hex_4-HexN-InsP	—	—	0.19	8.31
[AEP]Hex_4-HexN-InsP or Hex_4-[AEP]HexN-InsP	1.31	0.16	—	—
[AEP]Hex_5-HexN-InsP or Hex_5-[AEP]HexN-InsP	0.03	—	—	—
[AEP]Hex_6-HexN-InsP or Hex_6-[AEP]HexN-InsP	**52.03**	**14.72**	—	—
Hex_7-[AEP]HexN-InsP	0.81	0.59	—	—
Hex_8-[AEP]HexN-InsP	1.13	1.48	—	—
Hex_7-HexN-InsP	0.03	—	—	—
NANA-Hex_7-[AEP]HexN-InsP	0.01	—	—	—
Total	100.00	100.00	100.00	100.00
Lipid moiety				
Alkyl-C16:0-acyl-C16:0-Gro	7.48	2.12	***96.24***	***91.69***
Alkylacyl-C33:0-Gro	—	—	0.18	—
Alkylacyl-C34-Gro	0.15	0.08	0.80	—
Alkylacyl-C34:1-Gro	0.66	—	—	—
Alkylacyl-C34:2-Gro	0.22	—	0.25	8.31
Alkyl-C16:0-acyl-C24:0-Gro	0.13	—	2.49	—
Alkylacyl-C42:1-Gro	—	—	0.04	—
C16:0/d18:0-Cer	***21.67***	4.80	—	—
C16:0/d18:1-Cer	***26.00***	***11.29***	—	—
C16:0/t18:0-Cer	0.30	—	—	—
C22:0/d18:0-Cer	—	0.64	—	—
C22:0/d18:1-Cer	0.01	0.08	—	—
C22:0/t18:1-Cer	0.23	—	—	—
C23:0/d18:0-Cer	0.10	—	—	—
C23:0/d18:1-Cer	0.21	0.75	—	—
C24:0/d18:0-Cer	***19.21***	***26.11***	—	—
C24:0/d18:1-Cer	***21.09***	***48.84***	—	—
C24:0/d18:2-Cer	0.12	0.16	—	—
C24:0/t18:0-Cer	0.05	0.14	—	—
C24:0/t18:1-Cer	—	1.14	—	—
C25:0/d18:0-Cer	0.42	0.39	—	—
C25:0/d18:1-Cer	0.52	1.31	—	—
C26:0/d18:0-Cer	0.64	0.92	—	—
C26:0/d18:1-Cer	0.79	1.22	—	—
Total	100.00	100.00	100.00	100.00
Peptide				
APTPGD	—	—	—	100.00

[a]CM and CMW, $CHCl_3$:MeOH (1:1, v/v) and $CHCl_3$:MeOH:water (1:2:0.8, v/v/v), respectively; BuOH, 1-butanol.
[b]The amount of GIPL-derived *myo*-inositol is overestimated because of the presence of large quantities of free phosphatidylinositol (PI) in this fraction.
[c]The most abundant (>10%) glycan or lipid structures are shown in bold italics.

alkyl-C16:0 at the *sn*-1 position and variable-size fatty acyl group at the *sn*-2 position (Figure 3C–E; Supplementary Figure 3).

In total, 11 species of GPI anchors were identified. Of those, only three had been described earlier (Previato *et al*, 1995; Serrano *et al*, 1995; Almeida *et al*, 2000) and nine were found to be novel but much less abundant species (Figure 3B; Table II; Supplementary Table I; Supplementary Figure 3). The MS^2 TIC quantification showed that the two major species [EtNP]Hex_4-[AEP]HexN-InsP-alkyl-C16:0-acyl-C16:0-Gro; [AEP]Hex_4-[AEP]HexN-InsP-alkyl-C16:0-acyl-C16:0-Gro) comprise about 95% of total protein-derived GPIs, whereas the least abundant species accounts for only 0.04% ([EtNP]Hex_4-[AEP]HexN-InsP-alkylacyl-C42:1-Gro) (Supplementary Table I). The MS^2 and MS^3 spectra, and schematic fragmentation for

one of the novel species ([EtNP]Hex_4-[AEP]HexN-InsP-alkyl-C16:0-acyl-C24:0-Gro), representing ~2.5% of the total protein-derived GPI anchors, is shown in Figure 3C–E.

Analysis of trypsin-released GPI-peptides by LC-MS^n

In *T. cruzi*, mucins are one of the major GPI-anchored antigens expressed on the parasite cell surface, and they are involved in the escape from host immune response as well as adhesion and invasion of host cells (Acosta-Serrano *et al*, 2001; Buscaglia *et al*, 2006; Acosta-Serrano *et al*, 2007). *T. cruzi* mucins are encoded by at least 863 genes, grouped in two major families (i.e., TcMUC and TcSMUG) (Buscaglia *et al*, 2006; Acosta-Serrano *et al*, 2007). We had shown earlier that we could identify the mucins expressed by the infective trypomastigote stage through the analysis of a short (3–4-mer) peptide containing the ω site still attached to the GPI anchor (GPI-peptide) (Buscaglia *et al*, 2004). Here, we exploited this approach to sequence the GPI-peptide derived from mucins of noninfective epimastigotes and, therefore, to determine which gene family(ies) is(are) expressed on this parasite stage. Although epimastigote mucins are somewhat resistant to different proteases (e.g., trypsin, proteinase K) as observed by SDS–PAGE (not shown), the C-terminal region seems to be less glycosylated, thus, susceptible to proteolytic digestion. Thus, the mucin-rich (9% *n*-butanol) extract from *T. cruzi* epimastigotes was digested with trypsin and analyzed by LC-MS^2 and LC-MS^3 (Figure 3A). On the basis of the structures identified from the proteinase K-treated samples, we could rapidly match several fragments, such as AAG (*m/z* 537.6), AEP-HexN-PI (*m/z* 1065.5), Hex(AEP)HexN-PI (*m/z* 1227.6), and Hex2(AEP)HexN-PI (*m/z* 1389.7), thus partially determining the GPI-peptide structure (Figure 4A). By combining the partially determined GPI-structure and the possible peptides that could be generated by tryptic digestion and being attached to a GPI anchor (using the GPI-anchoring prediction), one peptide candidate could be the APTPGD sequence from mucin TcSMUG S. Indeed, the fragmentation of GPI-peptide species showed abundant fragments corresponding to the mass of this peptide attached to dehydrated EtN, EtNP, or EtNP-Hex_{1-4} (Figure 4A), similar to the fragmentation pattern described elsewhere (Redman *et al*, 1994). However, the fragmentation of the peptide moiety in the MS^2 was very poor impairing the sequence confirmation. To ultimately sequence the peptide and determine the ω site, fragments corresponding to the peptide attached to dehydrated EtN or AEP were subjected to data-independent MS^3 analysis (Figure 4B; Supplementary Figure 4). The quality of MS^3 spectra were good enough to enable the *de novo* sequencing of the peptide. With this approach, we could sequence peptides attached to five different GPI structures, four of them also detected after proteinase K digestion (Figure 4A–C; Table II; Supplementary Table I; Supplementary Figure 4). Although five GPI species were found, all these species were attached to the same peptide sequence (APTPGD), which corresponds to the carboxyl terminus of the TcSMUG S subfamily of mucins from *T. cruzi* (El-Sayed *et al*, 2005; Buscaglia *et al*, 2006; Acosta-Serrano *et al*, 2007). This result was corroborated by amino-acid compositional analysis (not shown). The TcSMUG S subfamily comprises eight unique sequences (GenBank accession numbers XP_804663.1, XP_807370.1, XP_807371.1, XP_821038.1, XP_821039.1, XP_821040.1, XP_821041.1, and XP_821042.1), with predicted mature protein sequences varying from 56 to 85 amino acids. TcSMUG S sequences contain approximately 40% threonine in their composition (Buscaglia *et al*, 2004), suggesting that they could be heavily *O*-glycosylated and, therefore, quite resistant to protease digestion as we have already observed (data not shown).

Quantitative analysis of GPIs

The absolute quantification was done by analyzing the *myo*-inositol content of the samples (Materials and methods). As most of the identified GPIs have closely related structures and conserved ionizable groups (AEP, EtNP, GlcN, and phosphate), we speculated that their ionization efficiency by electrospray would be comparable. Thus, we would be able to relatively quantify the different species by the signal intensity of each GPI or GIPL species. First, we injected the synthetic GPI sGPI-C16:0 mixed with an equal amount of sGPI-C18:1 or sGPI-C18:2 (see Materials and methods for details). Although these GPI species have different lipid tails, their ionization efficiencies were almost identical (Supplementary Figure 5A and B). Next, we analyzed sample mixtures containing constant concentration of sGPI-C18:1 (5 μM), but variable concentration of sGPI-C16:0 (0.005–5 μM). The response in the MS was evaluated by comparing the signal intensities of the parent-ion or by MS^2 TIC approach (Asara *et al*, 2008). The quantification by comparing parent-ion signal intensities showed small errors (1.3–33.6%), but the response was linear only from 1:1 to 1:20 ratio (Supplementary Table II). On the other hand, we observed higher errors (9.5–48.2%) by MS^2 TIC approach, similar to the ones described by Asara *et al* (2008), but the linear response was shown to be much broader (from 1:1 to 1:500 ratio) (Supplementary Table II). As most of the identified GPI and GIPL species are poorly abundant, we decided to relatively quantify them by MS^2 TIC approach.

Combining both quantifications, we were also able to estimate absolute amounts of individual protein-derived GPI and GIPL species. To determine the sensitivity of our method, we performed LC-MS^2 analysis with different amounts of GPIs. First, we performed the analysis in a data-dependent acquisition (DDA) mode, fragmenting the five most abundant ions with dynamic exclusion. We detected GPIs from as low as 5×10^5 parasites, equivalent to 21.5 femtomoles of total GPIs and 3.1 femtomoles for the least abundant species, as quantified by *myo*-inositol analysis by GC-MS (data not shown). Furthermore, MS spectra were also collected by multiple-reaction monitoring (MRM) mode, targeting the two major species of proteinase K-released GPIs at *m/z* 911.42 and 919.23 (Supplementary Table I). With this approach, we detected and characterized GPIs from as low as 1×10^5 parasites, equivalent to 4.3 femtomoles of total GPIs and 611 attomoles for the least abundant species observed (Figure 5A and B). It is also worth to point out that epimastigote is much smaller compared with higher eukaryote cells. Thus, our method would require less number of mammalian cells for

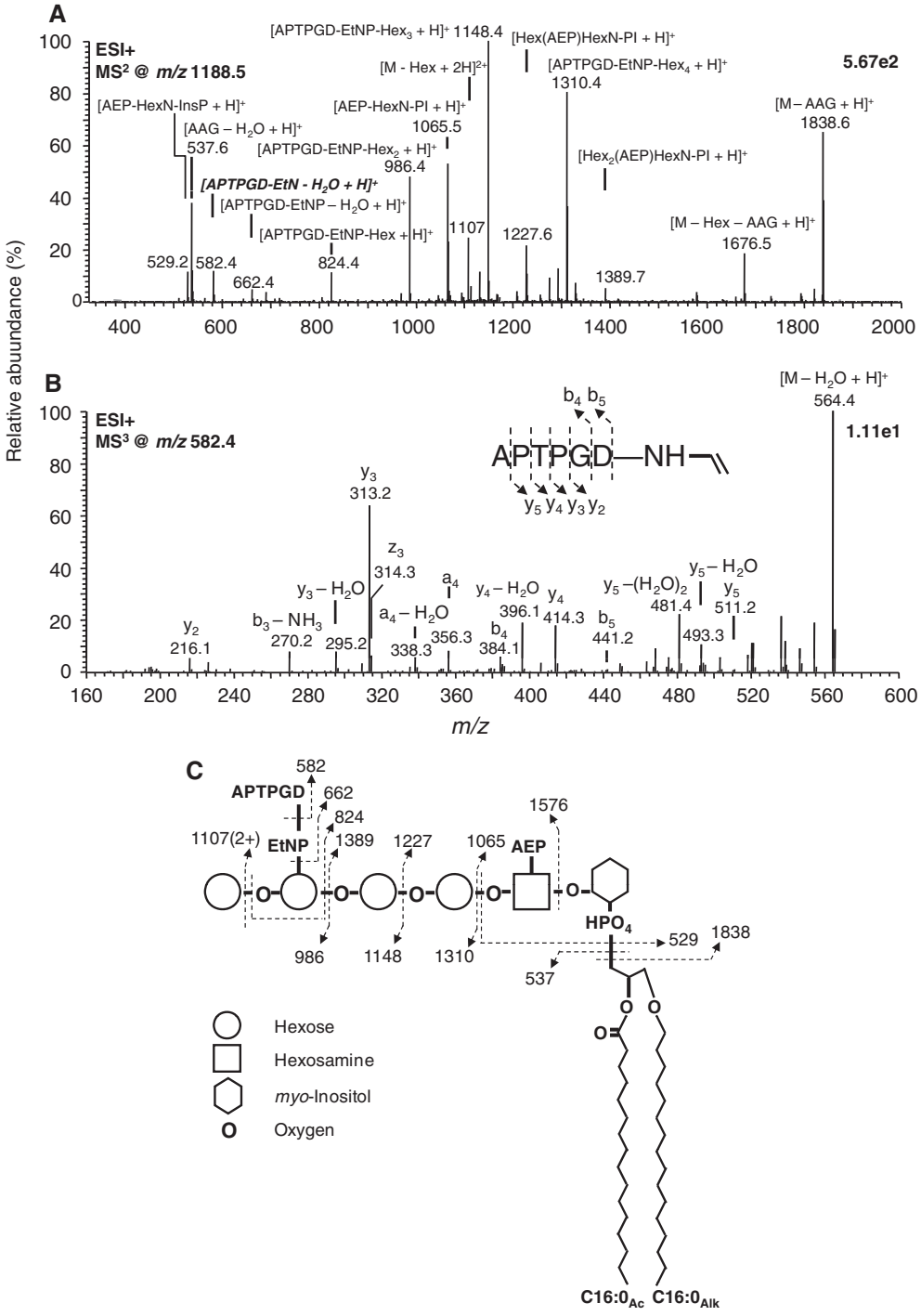

Figure 4 Analysis by LC-MS^2 and LC-MS^3 of GPI-peptides released by trypsin treatment. (**A**) MS2 spectrum of the GPI-peptide species at *m/z* 1188.5, which corresponds to APTPGD-EtNP-Hex$_4$-[AEP]HexN-InsP-1-*O*-alkyl-C16:0-2-*O*-acyl-C16:0-Gro. (**B**) MS3 spectrum of the fragment corresponding to the peptide attached to dehydrated EtN at *m/z* 582.4. The peptide fragments are indicated. (**C**) Proposed fragmentation and structure of the GPI-peptide APTPGD-EtNP-Hex$_4$-[AEP]HexN-InsP-1-*O*-alkyl-C16:0-2-*O*-acyl-C16:0-Gro. APTPGD, Ala-Pro-Thr-Pro-Gly-Asp.

doing the same analysis. A recent report characterized the GPI-anchor structure from *T. brucei* transferrin-receptor using 19.7 picomoles (Mehlert and Ferguson, 2008). Thus, our LC-MS approach is from ~6000- (for DDA) to ~30 000-fold (for MRM) more sensitive than current nonradioactive methods.

Retention times and column separation

The retention time of GPI species was determined as described in Materials and methods. Notably, the retention times of protein-derived GPIs and GIPLs containing the same lipid tail

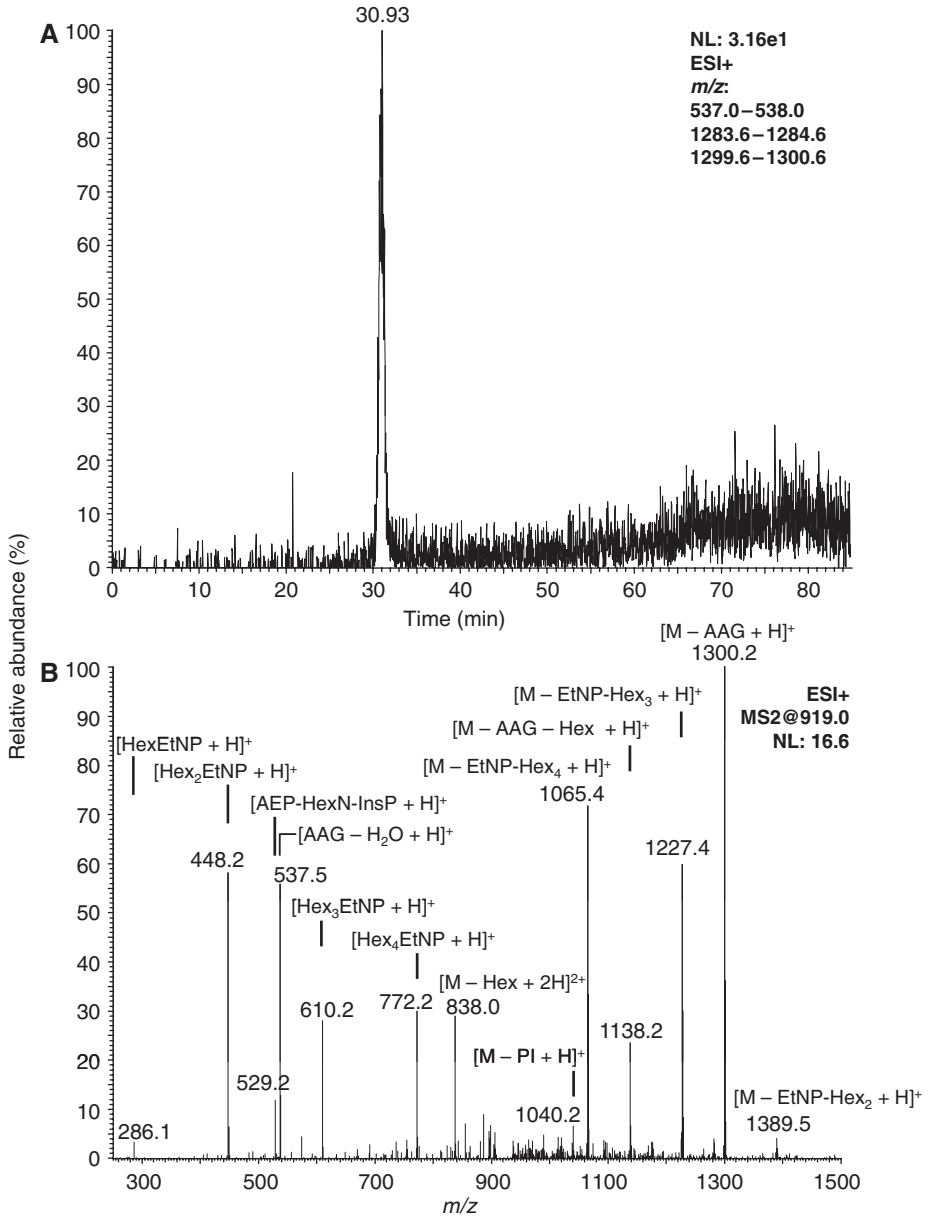

Figure 5 Analysis by LC-MRM-MS of proteinase K-released GPIs from *T. cruzi* epimastigotes. The two major GPI species originally observed at *m/z* 918.93 and 910.93 (protein-derived GPI species 1 and 2, respectively; Supplementary Table I) were subjected to MRM analysis (**A**) Extracted-ion chromatogram of the diagnostic ions corresponding to the dehydrated alkylacylglycerol (AAG—$H_2$0) ion-species at *m/z* 537.5, and the loss of alkylacylglycerol (M—AAG) at *m/z* 1300.1 and 1284.1. The peak detected at the retention time 30.93 min includes the two GPI species. For this experiment, the equivalent of 10^5 cells, corresponding to 4.3 femtomoles of total GPIs, was used. (**B**) MS^2 spectrum of the major GPI species ([EtNP]Hex_4-[AEP]HexN-InsP-1-*O*-alkyl-C16:0-2-*O*-acyl-C16:0-Gro) at *m/z* 919.

were very similar (Supplementary Table I), and apparently had no correlation with the size and different substituent groups (e.g., EtNP, AEP, Hex) attached to the glycan core. This finding indicates that the separation is likely given by the exclusive interaction of the lipid tail with the hydrophobic polystyrene-divinylbenzene-based C4 (POROS R1) resin. As expected, we also noticed that protein-derived GPIs and GIPLs with shorter, hydroxylated, and/or unsaturated lipid tails were eluted first from the column, compared with those with longer, nonhydroxylated, and/or saturated lipid tails (Supplementary Table I).

Discussion

Here, we present the first global analysis of GPI-anchored molecules from a human pathogenic eukaryote, *T. cruzi*. Our analysis included the genome-wise prediction of GPI-anchored proteins and the large-scale structural determination of protein-free and protein-linked GPIs. The prediction analysis showed that about 12% of *T. cruzi* genes possibly encode GPI-anchored proteins. This number is much higher compared with the African sleep sickness parasite, *T. brucei* (strain TREU927), which has only 132 (1.5%) of the 8750 sequences

possibly GPI anchored (data not shown). Poisson *et al* (2007) performed GPI-anchoring prediction for several organisms, including other early divergent eukaryotes *Plasmodium falciparum* (0.19%) and *Leishmania major* (1.02%); yeasts, *Saccharomyces cerevisiae* (0.95%) and *Cryptococcus neoformans* (0.90%); plants, *Arabidopsis thaliana* (0.83%) and *Gibberella zeae* (1.13%); and mammals, human (0.74%), and mouse (0.82%). Our study clearly shows that the number of potential GPI-anchored sequences is much higher in *T. cruzi* than in any other studied organisms.

The GPI-anchoring sequences in *T. cruzi* are concentrated in few large multigene families, such as MASP, mucin, TS/gp85, mucin-like, and gp63. Of these families, mucins, TS/gp85 glycoproteins, and gp63 metalloproteinases were shown to have important functions in the virulence and evasion/modulation of host immune response (Acosta-Serrano *et al*, 2007). Although plenty of information has been obtained with the completion of the genome project, little information is available about how many and which GPI-anchored molecules are expressed on the parasite surface. There is also a limitation of microarray (transcriptome) and real-time PCR data for *T. cruzi* gene expression, as in this organism and other trypanosomes, protein synthesis is mostly controlled by posttranscriptional regulation processes, like mRNA stabilization and degradation (Jager *et al*, 2007). In addition, microarray and real-time PCR analyses do not provide any information regarding absolute or relative expression, and structure of GPI-anchored proteins and GIPLs in different parasite stages and strains.

Large-scale GPIomic analysis in eukaryotes is limited due to the lack of a fast, simple, and sensitive method to characterize GPI-anchored molecules (Ferguson, 1992; Hooper, 2001). One of the limiting factors has been the absence of a more efficient method to purify GPIs. The current methods are based on HIC or RPC resins and have a restricted efficiency, which might be due to the amphiphilic character of these resins. Thus, we hypothesized that using a polystyrene-divinylbenzene-based C4 resin (e.g., POROS R1) (DePhillips *et al*, 1994; Whitney *et al*, 1998), which is essentially hydrophobic, would increase the chromatographic resolution of the RPC in the purification of GPI-anchored protein-linked and free GPIs (GIPLs). Indeed, our results proved that our hypothesis was correct. The resolution achieved with the POROS R1 RPC in our study was exceptional, and we could resolve and identify a remarkable number (90 species) of GIPLs and GPI-anchored protein-derived GPIs.

Then, to ultimate gain sensitivity and speed, we packed capillary columns with POROS R1 and analyzed *T. cruzi* epimastigote GPIs by LC-MS^n. First, we analyzed the protein-free GPI (GIPL) fraction. In total, we were able to identify 78 complete species of GIPLs. Of those, 70 were novel (Supplementary Table I). Up to now, only a few complete species of GIPLs had been detected and structurally characterized in *T. cruzi* (de Lederkremer *et al*, 1990, 1991; Previato *et al*, 1990; Almeida *et al*, 2000). Consistently, these species were also the most abundant found in this study in the GIPL sample extracted with 9% *n*-butanol.

In an earlier report, the glycan cores of GIPLs from Y strain epimastigote forms of *T. cruzi* were analyzed by NMR (Carreira *et al*, 1996). With that approach seven different structures for the glycan core were described. In this study, we found all those seven glycan cores and additional five novel ones. Three distinct series differing only in the number of hexose residues were found ([AEP]Hex_{5-6}-[AEP]HexN-InsP, [EtNP]Hex_{4-6}-[AEP]HexN-InsP, and AEP-Hex_{4-8}-HexN-InsP) (Table I). In addition, one minor species (comprising only 0.03% of the total GIPLs) without either EtNP or AEP (Hex_7-HexN-InsP) was observed (Table I; Supplementary Figure 2C63). Another minor species (0.01% of the total GIPLs) containing sialic acid (*N*-acetylneuraminic acid, NANA) (NANA-Hex_7-[AEP]HexN-InsP) was also detected and characterized by MS^2 analysis (Table I; Supplementary Figure 2C18). The presence of sialic acid-containing glycolipids in *T. cruzi* was proposed earlier, through *in vitro* radiolabeling using ^{3}H-NANA-labeled fetuin in epimastigote and trypomastigote cultures, but no structural data had been provided (Confalonieri *et al*, 1983; Zingales *et al*, 1987). Further structural and biosynthetic studies need to be carried out to establish the detailed structure and expression of the sialic acid-containing GIPLs in different parasite stages. Taken together, our results clearly show a greater diversity of GIPL species than was known earlier, therefore, indicating a higher complexity of the GPI biosynthesis in *T. cruzi*.

The analysis of proteinase K-released GPIs lead to the characterization of 11 different species. Of those, only three have been characterized earlier (Previato *et al*, 1995; Serrano *et al*, 1995; Almeida *et al*, 2000). Most of the glycan structures had either two AEP residues ([AEP]Hex_{4-5}-[AEP]HexN-InsP) or one AEP and one EtNP ([EtNP]Hex_{4-6}-[AEP]HexN-InsP). One species, however, had only one EtNP ([EtNP]Hex_4-HexN-InsP). Interestingly, the glycan cores containing one AEP residue (AEP-Hex_{4-8}-HexN-InsP) were shown to be highly abundant in GIPLs, but they were not detected in protein-derived GPI anchors, suggesting some differential steps during the biosynthesis of GIPLs and GPI anchors to be attached to proteins.

By analyzing the trypsin-released GPI-peptides, we could determine that the mucins expressed by the epimastigote forms of *T. cruzi* belong to the TcSMUG S subfamily. This subfamily of mucins comprises few genes that encode short polypeptides rich in Thr residues and, therefore, most likely to be highly *O*-glycosylated. The only other GPI-anchored polypeptide from epimastigotes with known amino-acid sequence is the NETNES (Macrae *et al*, 2005). This molecule has four glycosylation sites, but only 13 amino-acid residues. We were unable to detect NETNES in our analysis, probably due to the fact that this glycoconjugate does not have potential sites for trypsin digestion (Macrae *et al*, 2005). Furthermore, we cannot exclude that, akin to *T. brucei* variant surface glycoproteins (Ferguson *et al*, 1988), some *T. cruzi* GPI-anchored glycoproteins could not be extracted from cell membranes with 9% *n*-butanol. In this case, a stronger (more hydrophobic) extraction procedure using detergent (e.g., Triton X-100) should be used. This is currently under investigation in our laboratory.

Taken together, our results and others from the literature (Acosta-Serrano *et al*, 2001, 2007; Buscaglia *et al*, 2006) suggest that the epimastigote cell surface is heavily coated by a variety of GIPL structures, and to a less extent by short and highly glycosylated GPI-anchored polypeptides. GIPLs were

recently proposed to participate in the interaction of the parasite with the insect midgut (Nogueira *et al*, 2007). The absence of protease-cleavable polypeptide chains may also help protecting the parasite against insect digestive enzymes. On the other hand, bloodstream trypomastigotes were shown to express mucins from the TcMUC II subfamily that are composed by highly diverse sequences (Buscaglia *et al*, 2004). In agreement with these earlier data, our preliminary large-scale proteomic analysis of trypomastigotes suggests that this mammal-dwelling forms express hundreds, perhaps thousands, of GPI-anchored glycoproteins (including members of TS/gp85, MASP, and TcMUC II families) (Nakayasu *et al*, unpublished data). This high diversity of GPI-anchored proteins in trypomastigote forms may result in the lack of major immunodominant epitope(s) on the parasite surface, therefore, contributing for the immunoevasion of the parasite and its perpetuation within the mammalian host (Buscaglia *et al*, 2004).

In sum, here we showed that the GPIome of epimastigote forms of *T. cruzi* is highly complex. There were only few major GIPL (Carreira *et al*, 1996; Almeida *et al*, 2000) and eMUC-GPI (Previato *et al*, 1995; Serrano *et al*, 1995; Almeida *et al*, 2000) species characterized earlier, and there were no ω-sites described for *T. cruzi* epimastigote mucins. This high complexity of GIPLs and GPI-anchored proteins might be important for the interaction between the parasite and the insect host. This is the first comprehensive study of the characterization of free GPIs (GIPLs), protein-bound GPIs, and ω sites of a eukaryotic organism. This new methodology was shown to be simple and highly sensitive. Our method could be used for the high throughput screening (HTS) of potential GPI biosynthesis-targeting drugs against pathogenic parasites (Smith *et al*, 2004), as well as HTS of GPI-metabolism mutant cell lines. We also propose its use for the global analysis of the GPIome of other pathogenic eukaryotes and mammalian cells, including healthy and modified (cancer) cells, whose GPI anchor structures and roles are still mostly elusive (Paulick and Bertozzi, 2008).

Materials and methods

Prediction analysis of GPI-anchored proteins in the *T. cruzi* genome

The prediction analysis was carried out using the FragAnchor algorithm (Poisson *et al*, 2007) (http://navet.ics.hawaii.edu/~fraganchor/NNHMM/NNHMM.html) against the TcruziDB v5.0 (http://tcruzidb.org/tcruzidb/, downloaded on August 1, 2005), *T. cruzi* (downloaded on March 17, 2008) and *T. brucei* TREU927 (downloaded on August 13, 2008) sequences from the GenBank (http://www.ncbi.nlm.nih.gov/sites/entrez?db=Protein&itool=toolbar), and mucin TcMUC II sequences from GeneDB (http://www.genedb.org/, downloaded on August 20, 2005). Only sequences with high probability were accepted as potentially GPI-anchored proteins.

Solvents and salts

Otherwise indicated, all solvents and salts used in this study were of the highest quality available (HPLC or molecular biology grade), from Sigma-Aldrich, EMD Chemicals, or JT Baker.

T. cruzi culture

Noninfective epimastigote forms of *T. cruzi* (Y strain) were grown in liver infusion tryptose medium at 28°C for 3–4 days, as described earlier (Camargo, 1964). Cells were harvested and washed twice in phosphate-buffered saline (PBS), pH 7.4, and centrifuged at 2000 *g* at 4°C for 10 min. The parasite pellet containing 3.6×10^{10} was then store at −20°C until use.

Extraction of free GPIs and GPI-anchored proteins

The parasite pellet (3.6×10^{10} cells) was placed in an ice-bath, resuspended in 2 ml HPLC-grade water, and immediately transferred to a 15-ml PTFE-lined, screw capped glass tube (Supelco). Then, methanol, chloroform, and water were added to give 15 ml of the mixture chloroform:methanol:water (1:2:0.8, v:v:v). After homogenization in vortex (1 min), the suspension was centrifuged for 20 min at 2000 *g*. The organic phase was removed and transferred to a 100-ml glass flask, and the pellet was extracted three times with 10 ml chloroform:methanol (C:M, 2:1, v/v) solution, and twice with 10 ml chloroform:methanol:water (C:M:W, 1:2:0.8, v/v/v) solution. Between extractions, the organic phase was separated from the pellet by centrifugation at 2000 *g* for 20 min, at room temperature, and saved. All organic phases were combined, dried under N_2 stream, and the resulting organic extract was subjected to Folch's partition (Folch *et al*, 1957). The upper (aq.) phase of the partition, rich in free GPIs (GIPLs), was dried under N_2 stream and analyzed by LC-MS^2 as described below.

The resulting delipidated parasite pellet was dried under N_2 stream and extracted 3× with 4 ml 9% *n*-butanol, essentially as described (Almeida *et al*, 2000). The GPI-anchored proteins were separated from residual GIPLs by *n*-butanol:water (1:1, v/v) partition. The aq. phase, enriched with GPI-anchored proteins, was dried in a vacuum centrifuge (Vacufuge, Eppendorf) and stored at −20°C until use.

Biotinylation of epimastigote mucins and CLA

Epimastigote mucins were extracted and purified using octyl-Sepharose column (GE Healthcare), as described (Almeida *et al*, 2000). For biotinylation, the mucin sample was dried in a vacuum centrifuge, redissolved in 1 ml PBS containing 1 mM NHS-biotin (Pierce), and incubated for 2 h in an ice bath. The reaction was terminated by the addition of Tris–HCl pH 8.8 to a 50-mM final concentration, and the sample was dialyzed for 24 h against MilliQ ultrapure water (Millipore), at 4°C. The labeling efficiency was monitored by CLA. Briefly, biotinylated mucins were diluted 10-fold in 50 mM carbonate-bicarbonate buffer pH 9.6, immobilized on polystyrene 96-well chemiluminescent ELISA microplate (Nunc, Invitrogen) and incubated overnight at 4°C. After blocking for 2 h at 37°C with PBS-0.1% bovine serum albumin (PBS-A), 50 µl horseradish-conjugated streptavidin (1:1000 dilution) (Invitrogen) in PBS-A was added. The plate was washed 3× with PBS containing 0.1% Tween-20 (PBS-T) and developed with 50 µl chemiluminescent reagent (SuperSignal, Pierce) diluted five-fold in 200 mM carbonate/bicarbonate buffer, pH 9.6. Chemiluminescent readings were measured in a luminometer (Luminoskan Ascent, Labsystems, Finland).

Purification of synthetic GPIs

T. cruzi epimastigote and trypomastigote GPIs were synthesized as described earlier (Yashunsky *et al*, 2006) and stored as NMR samples in $[^2H]_6$-DMSO solution. Synthetic GPI samples were purified (from by-products appeared during storage) in reverse phase ziptips, built in 200-µl micropipette tips with a small piece of glass-fiber filter and 50 µl 40 mg/ml POROS R1 (C4) 50 resin (50-µm average particle size, Applied Biosystems) (Whitney *et al*, 1998) in 2-propanol. The ziptips were washed with 100 µl methanol and equilibrated with 5% 2-propanol/10 mM ammonium acetate (AA). About 1 nmol of synthetic GPI sample, containing several by-products, was resuspended in 100 µl of 5% 2-propanol/10 mM AA and after loading, the column was washed with the same buffer, followed by step-wise elution of increasing concentration (10–90%) of 100 µl 2-propanol containing

10 mM AA. Alternatively, we performed the ziptips by equilibrating the sample with 5% *n*-propanol and eluting with 80% *n*-propanol. The fractions were dried in a vacuum centrifuge and stored at −20°C until use.

Purification of GPI-anchored proteins from *T. cruzi* epimastigote extracts

Extracted GPI-anchored proteins from epimastigote forms were purified in POROS R1 solid-phase cartridges. The cartridges were assembled in solid-phase extraction supports with 1 ml POROS resin (40 mg/ml in 2-propanol). A glass fiber filter was laid over the resin bed to avoid disturbance during the sample loading and elution steps. The cartridge was washed with 4 ml 2-propanol and 4 ml methanol before being equilibrated with 4 ml 5% *n*-propanol. The purified mucin sample was redissolved in 5% *n*-propanol and spiked with biotinylated epimastigote mucins. After loading onto the cartridges, samples were eluted in increasing concentration (5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5, 30, 50, and 90%) of *n*-propanol. The fractions were dried in a vacuum centrifuge and analyzed by SDS–PAGE (Laemmli, 1970) stained with silver (Mortz *et al*, 2001), and by CLA as above. The bands from the gel were excised, digested with trypsin, and sequenced by LC-MS^2 (see below).

Protease digestion of GPI-anchored proteins

GPI-anchored proteins extracted with 9% butanol from the equivalent to 1.8×10^9 parasites (approximately 200 μg of total protein) were digested overnight at 37°C with 50 μg proteinase K (from *Tritirachium album*, lyophilized powder, ⩾30 units/mg protein, Sigma) or 2 μg sequencing-grade trypsin (Promega) in 100 mM ammonium bicarbonate, pH 8.0. The digested GPI samples were repurified in POROS R1 ziptips, as described above, and analyzed directly by ESI-MS^2 using a Qtof-1 ESI-MS, or LC-MS^n using a LTQXL ESI-linear ion trap-MS.

ESI-MS^2 analysis of GPI samples

Synthetic and purified GPI samples were resuspended in 50% *n*-propanol/10 mM AA and injected by infusion (500 nl/min) into an ESI-QTOF-MS (Qtof-1, Micromass, Waters). Spectra were collected in the negative-ion mode at 50–2000 mass-to-charge (*m/z*) range. The nanospray source was set at 1.5–2.5 kV, cone voltage 35 V, and the source and desolvation temperatures at 110 and 150°C, respectively. Each GPI species was subjected to fragmentation with collision energy from 30 to 50 eV.

LC-MS^n analysis of GPI samples

Capillary columns (75-μm internal diameter by 10-cm length) were packed with 10 mg/ml POROS R1 10 resin (10-μm particle size; Applied Biosystems) (suspension in 100% acetonitrile (ACN)) with a high-pressure apparatus at 500 psi (Gatlin *et al*, 1998). The quality of packing was monitored in a stereomicroscope. GPI samples were redissolved in 20% 2-propanol/0.2% FA and loaded onto the capillary column connected to a nanoHPLC system (1D-Plus, Eksigent). The samples were eluted in a gradient of 20–80% solvent B over 60 min (solvent A: 5% 2-propanol/0.2% FA; solvent B: 90% 2-propanol/0.2% FA) and directly analyzed in the linear ion-trap mass spectrometer (LTQ XL, Thermo Fisher Scientific). To compensate eventual decrease in the flow rate at high concentrations of organic solvent, the flow rate was set as a gradient from 350 to 450 nl/min within the same period of time. The ionization source was set to 1.9 kV and 200°C. The full MS scan was collected in positive mode from 400 to 2000 *m/z* range with a maximum injection time of 100 ms. The five most intense ions were submitted to fragmentation with 40% normalized collision energy and a maximum injection time of 150 ms. When the dynamic exclusion function was enabled, it was set to collect only twice and then to exclude for 30 s. To test the sensitivity of the method, the samples were also analyzed by the MRM approach. For this analysis, parent ions at *m/z* 911 and 919, which correspond to the major species of *T. cruzi* epimastigote mucin GPIs (Almeida *et al*, 2000), were constantly selected and fragmented. In addition, to determine the structure of lipid tail or the small peptide sequence attached to the GPI, these fragments were analyzed by MRM/MS^3. All the spectra were analyzed manually to assign the GPI structures.

Determination of retention times and relative quantification of GPI species

To determine the retention times, individual parent ions were plotted as extracted-ion chromatograms. The peaks were smoothed 13 times with the Gaussian method. The retention times were annotated as the top of individual peaks.

To standardize the quantitative analysis, synthetic GPIs [Man-[EtNP]Man-Man_2-[AEP]GlcN-InsP-1-*O*-alkyl-C16:0-2-*O*-acyl-C16:0-Gro (sGPI-C16:0), Man-[EtNP]Man-Man_2-[AEP]GlcN-InsP-1-*O*-alkyl-C16:0-2-*O*-acyl-C18:2-Gro (sGPI-C18:2), Man-[EtNP]Man-Man_2-[AEP]GlcN-InsP-1-*O*-alkyl-C16:0-2-*O*-acyl-C18:1-Gro (sGPI-C18:1)] were mixed in different ratios in 50% 2-propanol, 0.2% FA and analyzed in a linear ion-trap mass spectrometer (LTQXL, Thermo Fisher Scientific). The samples were injected with an automated system (Triversa Nanomate, Advion), set at 1.5 kV and 0.25 p.s.i. N_2 pressure, to avoid cross-contamination between samples. Full (enhanced scan rate, 100 ms maximum injection time) or MS^2 (normal scan rate, 3 a.m.u. isolation width, 35% normalized collision energy) scans were collected for 20 s. Each sample was run in triplicates. The relative quantification of the GPI species was done by the MS^2 TIC (Asara *et al*, 2008).

Identification of proteins separated by SDS–PAGE

The protein bands from the SDS–PAGE were excised and digested as described earlier (Shevchenko *et al*, 1996). The resulting peptides were purified in POROS R2 ziptips, as described elsewhere (Jurado *et al*, 2007). Resulting peptides were loaded into a capillary reverse phase column (PepMap, C18 3 μm, 15 cm × 75 μm, LC Packings, Dionex) connected to an Ultimate nanoHPLC (LC Packings, Dionex) in tandem to an ESI-QTOF-MS. The elution was performed with a flow rate of 300 nl/min in a gradient of 5–35% ACN/0.1% FA in 25 min, 35–72% ACN in 1 min, and hold at 72% ACN for 5 min. MS spectra in positive-ion mode were collected in the 400–1800 *m/z* range, and each peptide was fragmented (MS^2) for 3 s in the 50–2050 *m/z* range. MS data were converted into peak lists (PKL format) and searched against TcruziDB v5.0 (Aguero *et al*, 2006) (http://tcruzidb.org) using Mascot algorithm (Perkins *et al*, 1999) (Matrix Science). Database search parameters were the following: trypsin as the digesting enzyme (one missed cleavage site allowed); 500 ppm for peptide mass tolerance for monoisotopic ion; 0.8 Da for fragment mass tolerance; and carbamidomethylation of cysteine residues and oxidation of methionine residues as fixed and variable modifications, respectively. We validated only proteins with $P<0.05$, according to the software. An ion-score cutoff of 20 was set to ensure the quality of valid peptides and to remove redundant protein hits. Spectra with no peptide matching were subjected to *de novo* sequencing using PepSeq tool from the MassLynx v4.0 software (Waters).

Quantification of myo-inositol content by GC-MS

The *myo*-inositol content in the sample was determined as described (Ferguson, 1992). Briefly, 20 pmol of *scyllo*-inositol (Sigma) internal standard was added to each sample in a heat-cleaned (500°C, 2 h), one-end flame-sealed borosilicate 250-μl capillary microtube. An external standard containing 20 pmol of *myo*-inositol (Sigma) and 20 pmol of *scyllo*-inositol was also prepared. Samples were dried and subjected to strong acid hydrolysis with 50 μl of 6 M constant boiling HCl at 110°C for 24 h. The products were dried, redried on 20-μl methanol addition, and derivatized for 30 min with 15 μl freshly made trimethylsilyl (TMS) reagent (hexamethyldisilazane:trimethylchlorosilane:pyridine, 15:5:100, v:v:v). The inositol-TMS derivatives were analyzed on a TRACE GC Ultra gas chromatography system linked to a Polaris Q mass spectrometer (GC–MS, Thermo Fisher Scientific) using a Supelco SP2380™ column (30 m × 0.25 mm ID × 0.25 μm). The initial oven temperature was set to 100°C for 3 min and then a gradient of 15°C/min up to 250°C was applied. The final temperature of 250°C was kept for 5 min. The splitless injector was held at 200°C, the MS transfer line at

260°C, and the carrier gas flow at 0.7 ml/min. Extracted-ion chromatogram for *myo*-inositol was plotted using the diagnostic fragment ions at *m/z* 305 and 318. For the *myo*-inositol quantification, the chromatogram peaks were integrated and the following formula was applied: [area sample *myo* peak/area *scyllo* internal standard peak] × [amount of internal standard/(area standard *myo* peak/area standard *scyllo* peak)].

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb). Raw data files of LC-MS and GPI-prediction analyses are available from the ProteomeCommons repository (dataset at https://proteomecommons.org/dataset.jsp?i = 72069).

Acknowledgements

This study is dedicated to Professor Luiz RG Travassos in celebration of his 70th birthday. We thank Drs Sid Das (University of Texas at El Paso) and Alvaro Acosta-Serrano (University of Liverpool) for continuous support, valuable suggestions, and critical reading of the manuscript; and Dr Tiago Sobreira (University of Sao Paulo, Brazil) for assistance in the GPI prediction analysis. This work was funded by the grants 1R01AI070655 and 2S06GM008012-37 (to ICA) and 5G12RR008124 (to BBRC/Biology/UTEP) from the National Institutes of Health, and the Wellcome Trust (067089/Z/02/Z, to AVN). ESN was partially supported by the George A Krutilek memorial graduate scholarship from Graduate School, UTEP. LLN was partially supported by the Good Neighbor Scholarship, UTEP; and Florence Terry Griswold Scholarship I, Pan American Round Tables of Texas. We thank the Biomolecule Analysis Core Analysis at the Border Biomedical Research Center/Biology/UTEP (NIH grant # 5G12RR008124), for the access to the LC-MS instruments.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Acosta-Serrano A, Almeida IC, Freitas-Junior LH, Yoshida N, Schenkman S (2001) The mucin-like glycoprotein super-family of Trypanosoma cruzi: structure and biological roles. *Mol Biochem Parasitol* **114:** 143–150

Acosta-Serrano A, Hutchinson C, Nakayasu ES, Almeida IC, Carrington M (2007) Comparison and evolution of the surface architecture of trypanosomatid parasites. In *Trypanosomes: After the Genome*, Barry JD, Mottram JC, McCulloch R, Acosta-Serrano A (eds), pp 319–337. Norwich, UK: Horizon Scientific Press

Aguero F, Zheng W, Weatherly DB, Mendes P, Kissinger JC (2006) TcruziDB: an integrated, post-genomics community resource for Trypanosoma cruzi. *Nucleic Acids Res* **34:** D428–D431

Almeida IC, Camargo MM, Procopio DO, Silva LS, Mehlert A, Travassos LR, Gazzinelli RT, Ferguson MA (2000) Highly purified glycosylphosphatidylinositols from Trypanosoma cruzi are potent proinflammatory agents. *EMBO J* **19:** 1476–1485

Almeida IC, Gazzinelli RT (2001) Proinflammatory activity of glycosylphosphatidylinositol anchors derived from Trypanosoma cruzi: structural and functional analyses. *J Leukoc Biol* **70:** 467–477

Alves MJ, Colli W (2008) Role of the gp85/trans-sialidase superfamily of glycoproteins in the interaction of Trypanosoma cruzi with host structures. *Subcell Biochem* **47:** 58–69

Asara JM, Christofk HR, Freimark LM, Cantley LC (2008) A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* **8:** 994–999

Barreto-Bergter E, Vermelho AB, Hartmann R, Pohlentz G, Klein RA, Egge H (1992) Structural characterization of neutral glycosphingolipids from Trypanosoma cruzi. *Mol Biochem Parasitol* **51:** 263–270

Barrett MP, Burchmore RJ, Stich A, Lazzari JO, Frasch AC, Cazzulo JJ, Krishna S (2003) The trypanosomiases. *Lancet* **362:** 1469–1480

Bern C, Montgomery SP, Herwaldt BL, Rassi Jr A, Marin-Neto JA, Dantas RO, Maguire JH, Acquatella H, Morillo C, Kirchhoff LV, Gilman RH, Reyes PA, Salvatella R, Moore AC (2007) Evaluation and treatment of Chagas disease in the United States: a systematic review. *JAMA* **298:** 2171–2181

Buscaglia CA, Campo VA, Di Noia JM, Torrecilhas AC, De Marchi CR, Ferguson MA, Frasch AC, Almeida IC (2004) The surface coat of the mammal-dwelling infective trypomastigote stage of Trypanosoma cruzi is formed by highly diverse immunogenic mucins. *J Biol Chem* **279:** 15860–15869

Buscaglia CA, Campo VA, Frasch AC, Di Noia JM (2006) Trypanosoma cruzi surface mucins: host-dependent coat diversity. *Nat Rev Microbiol* **4:** 229–236

Camargo EP (1964) Growth and differentiation in *Trypanosoma cruzi*. I. Origin of metacyclic trypanosomes in liquid media. *Rev Inst Med Trop Sao Paulo* **12:** 93–100

Carreira JC, Jones C, Wait R, Previato JO, Mendonca-Previato L (1996) Structural variation in the glycoinositolphospholipids of different strains of Trypanosoma cruzi. *Glycoconj J* **13:** 955–966

Confalonieri AN, Martin NF, Zingales B, Colli W, de Lederkremer RM (1983) Sialoglycolipids in Trypanosoma cruzi. *Biochem Int* **7:** 215–222

de Lederkremer RM, Lima C, Ramirez MI, Casal OL (1990) Structural features of the lipopeptidophosphoglycan from Trypanosoma cruzi common with the glycophosphatidylinositol anchors. *Eur J Biochem* **192:** 337–345

de Lederkremer RM, Lima C, Ramirez MI, Ferguson MA, Homans SW, Thomas-Oates J (1991) Complete structure of the glycan of lipopeptidophosphoglycan from Trypanosoma cruzi Epimastigotes. *J Biol Chem* **266:** 23670–23675

DePhillips P, Buckland B, Gbewonyo K, Yamazaki S, Sitrin R (1994) Reversed-phase high-performance liquid chromatography assay for recombinant acidic fibroblast growth factor in E. coli cell suspensions and lysate samples. *J Chromatogr A* **663:** 43–51

Dias JC, Silveira AC, Schofield CJ (2002) The impact of Chagas disease control in Latin America: a review. *Mem Inst Oswaldo Cruz* **97:** 603–612

Dumonteil E (2007) DNA vaccines against protozoan parasites: advances and challenges. *J Biomed Biotechnol* **2007:** 90520

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E *et al* (2005) The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. *Science* **309:** 409–415

Elortza F, Nuhse TS, Foster LJ, Stensballe A, Peck SC, Jensen ON (2003) Proteomic analysis of glycosylphosphatidylinositol-anchored membrane proteins. *Mol Cell Proteomics* **2:** 1261–1270

Ferguson MA (1992) Chemical and enzymatic analysis of glycosylphosphatidylinositol anchors. In *Lipid Modifications of Proteins: A Practical Approach*, Hooper NM, Turner AJ (eds), pp 191–230. Oxford, UK: IRL Press

Ferguson MA (1999) The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. *J Cell Sci* **112** (Part 17)**:** 2799–2809

Ferguson MA, Homans SW, Dwek RA, Rademacher TW (1988) Glycosylphosphatidylinositol moiety that anchors Trypanosoma brucei variant surface glycoprotein to the membrane. *Science* **239:** 753–759

Fish WR, Holz Jr GG, Beach DH (1982) Some Phytomonas and Herpetomonas species form unique iso-branched polyunsaturated fatty acids. *Mol Biochem Parasitol* **5:** 1–18

Folch J, Lees M, Sloane Stanley GH (1957) A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem* **226:** 497–509

Frasch AC (2000) Functional diversity in the trans-sialidase and mucin families in Trypanosoma cruzi. *Parasitol Today* **16:** 282–286

Garg N, Bhatia V (2005) Current status and future prospects for a vaccine against American trypanosomiasis. *Expert Rev Vaccines* **4:** 867–880

Gatlin CL, Kleemann GR, Hays LG, Link AJ, Yates III JR (1998) Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry. *Anal Biochem* **263:** 93–101

Gazzinelli RT, Denkers EY (2006) Protozoan encounters with Toll-like receptor signalling pathways: implications for host parasitism. *Nat Rev Immunol* **6:** 895–906

Hooper NM (2001) Determination of glycosyl-phosphatidylinositol membrane protein anchorage. *Proteomics* **1:** 748–755

Hotez PJ, Bottazzi ME, Franco-Paredes C, Ault SK, Periago MR (2008) The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. *PLoS Negl Trop Dis* **2:** e300

Hsu FF, Turk J (2000) Structural determination of sphingomyelin by tandem mass spectrometry with electrospray ionization. *J Am Soc Mass Spectrom* **11:** 437–449

Jager AV, De Gaudenzi JG, Cassola A, D'Orso I, Frasch AC (2007) mRNA maturation by two-step trans-splicing/polyadenylation processing in trypanosomes. *Proc Natl Acad Sci USA* **104:** 2035–2042

Jurado JD, Rael ED, Lieb CS, Nakayasu E, Hayes WK, Bush SP, Ross JA (2007) Complement inactivating proteins and intraspecies venom variation in Crotalus oreganus helleri. *Toxicon* **49:** 339–350

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680–685

Lazzaroni JC, Dubuisson JF, Vianney A (2002) The Tol proteins of Escherichia coli and their involvement in the translocation of group A colicins. *Biochimie* **84:** 391–397

Macrae JI, Acosta-Serrano A, Morrice NA, Mehlert A, Ferguson MA (2005) Structural characterization of NETNES, a novel glycoconjugate in Trypanosoma cruzi epimastigotes. *J Biol Chem* **280:** 12201–12211

McConville MJ, Ferguson MA (1993) The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. *Biochem J* **294** (Part 2)**:** 305–324

McConville MJ, Menon AK (2000) Recent developments in the cell biology and biochemistry of glycosylphosphatidylinositol lipids (review). *Mol Membr Biol* **17:** 1–16

Mehlert A, Ferguson MA (2008) Proteomic scale high-sensitivity analyses of GPI membrane anchors. *Glycoconj J* (in press)

Moncayo A, Ortiz Yanine MI (2006) An update on Chagas disease (human American trypanosomiasis). *Ann Trop Med Parasitol* **100:** 663–677

Mortz E, Krogh TN, Vorum H, Gorg A (2001) Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics* **1:** 1359–1363

Nogueira NF, Gonzalez MS, Gomes JE, de Souza W, Garcia ES, Azambuja P, Nohara LL, Almeida IC, Zingales B, Colli W (2007) Trypanosoma cruzi: involvement of glycoinositolphospholipids in the attachment to the luminal midgut surface of Rhodnius prolixus. *Exp Parasitol* **116:** 120–128

Orlean P, Menon AK (2007) Thematic review series: lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids. *J Lipid Res* **48:** 993–1011

Paulick MG, Bertozzi CR (2008) The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. *Biochemistry* **47:** 6991–7000

Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20:** 3551–3567

Piron M, Verges M, Munoz J, Casamitjana N, Sanz S, Maymo RM, Hernandez JM, Puig L, Portus M, Gascon J, Sauleda S (2008) Seroprevalence of Trypanosoma cruzi infection in at-risk blood donors in Catalonia (Spain). *Transfusion* **48:** 1862–1868

Poisson G, Chauve C, Chen X, Bergeron A (2007) FragAnchor: a large-scale predictor of glycosylphosphatidylinositol anchors in eukaryote protein sequences by qualitative scoring. *Genomics Proteomics Bioinformatics* **5:** 121–130

Previato JO, Gorin PA, Mazurek M, Xavier MT, Fournet B, Wieruszesk JM, Mendonca-Previato L (1990) Primary structure of the oligosaccharide chain of lipopeptidophosphoglycan of epimastigote forms of Trypanosoma cruzi. *J Biol Chem* **265:** 2518–2526

Previato JO, Jones C, Xavier MT, Wait R, Travassos LR, Parodi AJ, Mendonca-Previato L (1995) Structural characterization of the major glycosylphosphatidylinositol membrane-anchored glyco-protein from epimastigote forms of Trypanosoma cruzi Y-strain. *J Biol Chem* **270:** 7241–7250

Previato JO, Wait R, Jones C, DosReis GA, Todeschini AR, Heise N, Previato LM (2004) Glycoinositolphospholipid from Trypanosoma cruzi: structure, biosynthesis and immunobiology. *Adv Parasitol* **56:** 1–41

Quanquin NM, Galaviz C, Fouts DL, Wrightsman RA, Manning JE (1999) Immunization of mice with a TolA-like surface protein of Trypanosoma cruzi generates CD4(+) T-cell-dependent parasiticidal activity. *Infect Immun* **67:** 4603–4612

Redman CA, Green BN, Thomas-Oates JE, Reinhold VN, Ferguson MA (1994) Analysis of glycosylphosphatidylinositol membrane anchors by electrospray ionization-mass spectrometry and collision induced dissociation. *Glycoconj J* **11:** 187–193

Serrano AA, Schenkman S, Yoshida N, Mehlert A, Richardson JM, Ferguson MA (1995) The lipid structure of the glycosylphosphatidylinositol-anchored mucin-like sialic acid acceptors of Trypanosoma cruzi changes during parasite differentiation from epimastigotes to infective metacyclic trypomastigote forms. *J Biol Chem* **270:** 27244–27253

Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68:** 850–858

Smith TK, Crossman A, Brimacombe JS, Ferguson MA (2004) Chemical validation of GPI biosynthesis as a drug target against African sleeping sickness. *EMBO J* **23:** 4701–4708

Urbina JA, Docampo R (2003) Specific chemotherapy of Chagas disease: controversies and advances. *Trends Parasitol* **19:** 495–501

Whitney D, McCoy M, Gordon N, Afeyan N (1998) Characterization of large-pore polymeric supports for use in perfusion biochromatography. *J Chromatogr A* **807:** 165–184

Wilkinson SR, Taylor MC, Horn D, Kelly JM, Cheeseman I (2008) A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. *Proc Natl Acad Sci USA* **105:** 5022–5027

Yashunsky DV, Borodkin VS, Ferguson MA, Nikolaev AV (2006) The chemical synthesis of bioactive glycosylphosphatidylinositols from Trypanosoma cruzi containing an unsaturated fatty acid in the lipid. *Angew Chem Int Ed Engl* **45:** 468–474

Yoshida N (2006) Molecular basis of mammalian cell invasion by Trypanosoma cruzi. *An Acad Bras Cienc* **78:** 87–111

Zingales B, Carniol C, de Lederkremer RM, Colli W (1987) Direct sialic acid transfer from a protein donor to glycolipids of trypomastigote forms of Trypanosoma cruzi. *Mol Biochem Parasitol* **26:** 135–144