

2008-01-01

Molecular Characterization of Trypanosoma cruzi and Shed Vesicle Components Involved in Host Immunomodulation and Cell Invasion

Ernesto Satoshi Nakayasu

University of Texas at El Paso, esnakayasu@miners.utep.edu

Follow this and additional works at: https://digitalcommons.utep.edu/open_etd



Part of the [Microbiology Commons](#), [Molecular Biology Commons](#), and the [Parasitology Commons](#)

Recommended Citation

Nakayasu, Ernesto Satoshi, "Molecular Characterization of Trypanosoma cruzi and Shed Vesicle Components Involved in Host Immunomodulation and Cell Invasion" (2008). *Open Access Theses & Dissertations*. 318.
https://digitalcommons.utep.edu/open_etd/318

This is brought to you for free and open access by DigitalCommons@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of DigitalCommons@UTEP. For more information, please contact lweber@utep.edu.

**Molecular Characterization of *Trypanosoma cruzi* and Shed Vesicle
Components Involved in Host Immunomodulation and
Cell Invasion**

ERNESTO SATOSHI NAKAYASU

Department of Biological Sciences

APPROVED:

Igor C. Almeida, Ph.D., Chair

Siddhartha Das, Ph.D.

Kristine Garza, Ph.D.

Mahesh Narayan, Ph.D.

Patricia D. Witherspoon, Ph.D.
Dean of the Graduate School

Copyright ©

by

Ernesto Satoshi Nakayasu

2008

Dedication

I dedicate this dissertation to my parents, to my father who currently lives in Sao Paulo and to the memory of my beloved mother Hathuyo Watanabe Nakayasu, who passed away on February 10th, 2004, at the age of 59.

**Molecular Characterization of *Trypanosoma cruzi* and Shed Vesicle
Components Involved in Host Immunomodulation and
Cell Invasion**

by

ERNESTO SATOSHI NAKAYASU, B. Sc.

DISSERTATION

**Presented to the Faculty of the Graduate School of
The University of Texas at El Paso
in Partial Fulfillment
of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY**

**Department of Biological Sciences
THE UNIVERSITY OF TEXAS AT EL PASO
December 2008**

Acknowledgements

First of all, special thanks for my uncle Masaru Nakayasu and Dr. Igor Almeida, not only for the personal and financial support, but even more important because they have invested in my career and they are very great mentors. I also want to thanks my family, including my father, siblings, uncles, aunties, cousins and grandmother, for the irestricted support allowing me to come to the U.S. for my doctoral studies.

I also want to acknowledge my collaborators, Drs. Mike Ferguson, Andrei Nikolaev, Dmitry Yashunsky and Douglas Lamont from University of Dundee, Scotland; Mr. Tiago Sobreira, and Drs. Ana Claudia Torrecilhas, Dr. Paulo Oliveira and Maria Julia Alves from University of Sao Paulo, Brazil; Dr. Fabio Gozzo from State University of Campinas, Brazil; Drs. Esteban Cordero, Lucina Gentil, Nobuko Yoshida and Jose Franco da Silveira from Federal University of Sao Paulo; Mr. Matthew Gaynor, Mr. Rafael Torres, Ms. Lilian Nohara, Dr. Luciane Ganiko, Dr. Alexandre Marques, Dr. Juan Noveron, Dr. Jeremy Ross from University of Texas at El Paso (UTEP).

Many thanks for Dr. Rosa Maldonado (UTEP), Dr. Siddhartha Das (UTEP), Dr. Mahesh Narayan (UTEP), Dr. Kristine Garza (UTEP), Dr. Alvaro Acosta-Serrano (University of Liverpool, UK), Dr. Daniel Lorenzini (Petrobras, Brazil), Dr. Nilson Assuncao (State University of Sao Paulo, Dr. Sirlei Daffre (University of Sao Paulo), Dr. Marcio Rodrigues (Federal University of Rio de Janeiro, Brazil), Dr. Leonardo Nimrichter (Federal University of Rio de Janeiro, Brazil) and Dr. Celso Sant'Anna (INMETRO, Brazil) for fruitfull comments and discussion; the personnel of the core facilities for the assistance and access to the equipments; for the personnel of the main office for assistance in burocratic issues; and for the friends from the Department of Biological Sciences, UTEP.

This work was supported by Georges A. Krutilek memorial scholarship (Graduate School, UTEP and National Institutes of Health (1R01AI070655, 5S06GM08012-37, and 5G12RR008124).

Abstract

Chagas disease caused by *Trypanosoma cruzi* is a devastating infectious disease with millions of cases in Latin America, and recently became a public health concern in United States and Europe. Although many efforts have been made for the development of an effective immunotherapy, currently there is no human vaccine for Chagas disease. Thus, the treatment is based only on two drugs that have limited efficacy and in some cases present severe side effects. One restriction for the rational approach to develop new therapies against this disease is the limited information about the proteins, glycolipids and protein posttranslational modifications expressed by different phylogenetic lineages, strains, and stages of the parasite. In this dissertation, I focused in the analysis of glycoconjugates of the *T. cruzi* surface and secreted vesicles, as well in the analysis of the parasite phosphoproteome. The results presented here demonstrated that the glycocalix of each stage of the parasite has major differences in the composition. The cell surface of the insect stages of the parasite is mainly composed by a highly diverse glycolipid coat, in addition to short highly glycosylated polypeptides. On the other hand, the surface coat of mammalian host-dwelling stages is composed mainly by hundreds of glycoproteins. These findings have many implications for the parasite survival in the insect and mammalian hosts. Next, by examining the phosphoproteins of the epimastigote stage of the parasite, over 200 phosphorylation sites were mapped in proteins with various functions. Taken together, the results from this dissertation brought new insights into *T. cruzi* physiology and virulence, which may have implications for the design of new therapies against Chagas disease.

Table of Contents

Acknowledgements	v
Abstract	vi
Table of Contents	vii
List of Tables	ix
List of Figures	xi
General Introduction	1
References.....	8
Chapter 1: The Proteome of <i>Trypanosoma cruzi</i> Shed Vesicles Involved in Host Cell Invasion	16
1.1 Abstract	17
1.2 Introduction.....	18
1.3 Material and Methods	22
1.4 Results	30
1.5 Discussion	34
1.6 Conclusions.....	41
1.7 Acknowledgments	42
1.8 References	43
Chapter 2: GPIomics: Global Analysis of Glycosylphosphatidylinositol-Anchored Molecules of <i>Trypanosoma cruzi</i>	73
2.1 Abstract	74
2.2 Introduction.....	75
2.3 Material and Methods	80
2.4 Results	88
2.5 Discussion	99
2.6 Acknowledgments	104
2.7 References	105
Chapter 3: Proteomic Analysis of Detergent-Solubilized Membrane Proteins from Insect-Developmental Forms of <i>Trypanosoma cruzi</i>	139
3.1 Abstract	140

3.2	Introduction.....	141
3.3	Material and Methods	144
3.4	Results and Discussion	150
3.5	Conclusions.....	165
3.6	Acknowledgments	166
3.7	References	167
Chapter 4: Improved proteomic approach for the discovery of potential vaccine targets in <i>Trypanosoma cruzi</i>		
4.1	Abstract	187
4.2	Introduction.....	188
4.3	Material and Methods	191
4.4	Results and Discussion	195
4.5	Conclusions.....	202
4.6	Acknowledgments	203
4.7	References	204
Chapter 5: Phosphoproteomic analysis of the human pathogen <i>Trypanosoma cruzi</i> at the epimastigote stage		
5.1	Abstract	265
5.2	Introduction.....	266
5.3	Material and Methods	269
5.4	Results and Discussion	273
5.5	Conclusions.....	283
5.6	Acknowledgments	284
5.7	References	285
Final Discussion and Concluding Remarks		309
References		315
Appendix		322
List of Publications		322
List of Abbreviations		326
Vita.....		331

List of Tables

Chapter 1

Table 1.1:56

Table 1.2:57

Table 1.3:58

Table 1.4:59

Table 1.5:63

Chapter 2

Table 2.1:115

Table 2.2:116

Table 2.3:117

Table 2.4:121

Chapter 3

Table 3.1:176

Table 3.2:177

Table 3.3:178

Table 3.4:181

Chapter 4

Table 4.1:211

Table 4.2:212

Table 4.3:213

Table 4.4:255

Table 4.5:260

Chapter 5

Table 5.1:295

Table 5.2:296

Table 5.3:297

Table 5.4:303

List of Figures

Chapter 1

Figure 1.1:66

Figure 1.2:67

Figure 1.3:68

Figure 1.4:70

Figure 1.5:71

Figure 1.6:72

Chapter 2

Figure 2.1:122

Figure 2.2:123

Figure 2.3:127

Figure 2.4:128

Figure 2.5:130

Figure 2.6:133

Figure 2.7:135

Figure 2.8:138

Chapter 3

Figure 3.1:183

Figure 3.2:184

Figure 3.3:185

Chapter 4

Figure 4.1:261

Figure 4.2:	262
-------------------	-----

Chapter 5

Figure 5.1:	305
-------------------	-----

Figure 5.2:	306
-------------------	-----

Figure 5.3:	307
-------------------	-----

Figure 5.4:	308
-------------------	-----

Final Discussion and Concluding Remarks

Figure 6.1:	320
-------------------	-----

Figure 6.2:	321
-------------------	-----

GENERAL INTRODUCTION

American trypanosomiasis or Chagas' disease is a neglected infectious disease caused by the protozoan parasite, *Trypanosoma cruzi*. Chagas' disease is a major public health problem in many Latin America countries, where up to 50,000 people may die every year due to complications in the acute or chronic phase of the disease (Dias *et al*, 2002; Stuart *et al*, 2008). Recent data estimate that there are about 11 million people infected with the disease (Dias *et al*, 2002; Stuart *et al*, 2008). In addition, due to the migration of asymptomatic infected people from endemic areas and lack of screening in blood banks, Chagas' disease is becoming a public health concern in the United States and Europe (Bern *et al*, 2007; Piron *et al*, 2008). Furthermore, increasing numbers of autochthonous cases of Chagas' disease have been reported in the United States (Tarleton *et al*, 2007). A recent survey of a major blood bank in El Paso, TX, diagnosed 2 infected patients in about 10,000 tested blood samples (Tobler *et al*, 2007).

Chagas' disease has two phases, the acute and the chronic. The acute phase is characterized by a high parasitemia and a strong inflammatory response. The clinical manifestations start in about one week and persist for 1-2 months. The symptoms are general malaise, such as fever, fatigue, body ache, diarrhea, and vomiting (Barrett *et al*, 2003; Stuart *et al*, 2008; Tarleton *et al*, 2007). The chronic phase is generally asymptomatic and associated with the progressive damage of the heart and gastrointestinal tract, leading to the eventual failure of these organs (Barrett *et al*, 2003; Stuart *et al*, 2008; Tarleton *et al*, 2007).

Although Chagas' disease has been discovered 100 years ago by Carlos Chagas (Chagas, 1909), the treatment of this infection is still based only on two drugs: benznidazole (Rochagan, Roche) and nifurtimox (Lampit, Bayer). These drugs have different efficacy during the acute or chronic phase of the disease and may cause severe side effects (Urbina and

Docampo, 2003). Although they have been used for many decades to treat patients with Chagas' disease, only recently the mechanism of action of these drugs was elucidated. Trypanosomes have a mitochondrial nitroreductase that activate the drugs into nitro-reduced derivatives of benznidazole or nifurtimox, which are highly reactive, leading to the damage of parasite macromolecules (Wilkinson *et al*, 2008).

The life cycle of *T. cruzi* comprises two stages in the insect vector and two stages in the mammalian host (e.g., humans). During a bloodmeal kissing bugs can ingest infective bloodstream trypomastigotes, which differentiate into non-infective epimastigotes that replicate in the midgut of insect. In the distal portion of the insect gut, epimastigotes transform into infective metacyclic trypomastigotes, which are eventually excreted with the feces during a new blood uptake by the insect vector. These metacyclic trypomastigotes can penetrate the bloodstream through the insect's bite wound or host exposed mucosal tissues, and infect several types of nucleated cells. Inside the cells, metacyclics differentiate into amastigotes, which reproduce by binary fission and a few days later transform into trypomastigotes before rupturing the cell and reaching the extracellular milieu and, eventually, the bloodstream. Cell derived- trypomastigotes can infect new host cells or be uptaken by the insect vector (Barrett *et al*, 2003; Stuart *et al*, 2008; Tarleton *et al*, 2007).

The circulating trypomastigote forms activate or down-regulate several components of host immune system. One of the major surface glycoproteins of trypomastigotes, namely glycosylphosphatidylinositol (GPI)-anchored mucins (or tGPI-mucins), and their GPI anchors (tGPIs), have the ability to induce a strong proinflammatory response in host macrophages through the activation of Toll-like receptor 2 (TLR2) and myeloid differentiation primary response gene 88 (MyD88) (Almeida *et al*, 2000; Campos *et al*, 2001; Campos *et al*, 2004). TLR2 and MyD88 engagement results in phosphorylation of different mitogen-activated protein

kinases (MAPKs) and related transcription factors in inflammatory macrophages, leading to the expression of proinflammatory cytokines, such as IL-12 and TNF- α (Almeida *et al*, 2000; Campos *et al*, 2001; Ropert *et al*, 2001; Ropert *et al*, 2003).

During the adaptive immune response, large titers of anti-*T. cruzi*-specific lytic antibodies are produced. The immunodominant antibodies in both acute and chronic phases of Chagas' disease are the anti- α Gal (Ch anti- α Gal) antibodies (Almeida *et al*, 1991; Gazzinelli *et al*, 1991), which recognize non-reducing end, terminal α -galactosyl epitopes expressed on tGPI-mucins (Almeida *et al*, 1994a, b; Almeida *et al*, 1993) and other glycoconjugates, such as Tc85 or gp85 glycoprotein (Couto *et al*, 1990). Ch anti- α Gal antibodies have the ability to lyse metacyclic trypomastigotes and cell-derived trypomastigotes, and control the infection (Almeida *et al*, 1994a, b; Almeida *et al*, 1991; Gazzinelli *et al*, 1991).

Another important *T. cruzi* molecule is the *trans*-sialidase (TS). TS is a GPI-anchored cell surface enzyme that transfers sialic acid residues from host to parasite glycoconjugates (Schenkman *et al*, 1994). The major acceptor of sialic acid transferred by TS are the tGPI-mucins (Schenkman *et al*, 1991). Furthermore, the addition of sialic acid onto tGPI-mucin O-linked oligosaccharides confers a protection to the parasite against the lytic activity of Ch anti- α Gal antibodies (Acosta-Serrano *et al*, 2001; Buscaglia *et al*, 2006; Pereira-Chioccola *et al*, 2000). TS and other homolog proteins comprise a multigene family known as TS/gp85, with more than 1400 genes in the *T. cruzi* genome (El-Sayed *et al*, 2005; Frasch, 2000). However, not all members of this superfamily have catalytic (TS) activity.

The non-catalytic members of TS/gp85 superfamily include glycoproteins with about 85 kDa (e.g., Tc85 or gp85) that interact with laminin and cytokeratin 18 leading to parasite attachment and invasion of host cells (Alves *et al*, 1986; Giordano *et al*, 1999; Magdesian *et al*, 2001). Following the invasion, non-catalytic TS/gp85 glycoproteins facilitate the escape of

trypomastigotes from the parasitophorous vacuole to the cytosol, where they differentiate into amastigotes and proliferate (Rubin-de-Celis *et al*, 2006; Schenkman *et al*, 1994). Another group of non-catalytic members of TS/gp85 superfamily comprises glycoproteins of approximately 160 kDa, such as complement regulatory proteins (CRP), which bind to complement factors C3b and C4b and trigger their inactivation by a protease (Norris, 1996; Norris *et al*, 1991; Norris and Schimpf, 1994). This protease is still not fully characterized, but preliminary zymogram studies showed that it has about 75 kDa, and its activity is inhibited by aprotinin, leupeptin, and EDTA (Norris, 1996). Recent work has shown that a cell surface metalloprotease (gp63) has about the same molecular mass (Cuevas *et al*, 2003). The gp63 homolog from *Leishmania* *ssp.* has the ability to inactivate the complement system (Yao *et al*, 2003); therefore, it may act as a complement inactivating enzyme in *T. cruzi* as well.

TS/gp85 glycoproteins and many other antigens, such as Tc-52 and cruzipain, have been reported to be secreted by trypomastigotes into culture medium (Affranchino *et al*, 1989; Aparicio *et al*, 2004; Ouaisi *et al*, 2002). Tc-52 has also the ability to activate TLR2 (Ouaisi *et al*, 2002). Monteiro *et al.* showed that TLR2 and bradykinin B₂ receptor cooperate during the induction of type-1 immunity, in a process dependent on cruzipain (Monteiro *et al*, 2006). On the other hand, trypomastigotes have been reported to secrete antigen-rich membrane vesicles into the culture medium (Goncalves *et al*, 1991). These vesicles were shown to strongly react to anti- α Gal antibodies and were named Tc α GalVes (Torrecilhas and Nakayasu *et al.*, unpublished data). Tc α GalVes were shown to be a potent activator of TLR2, and surprisingly, this activation resulted in high increase of host cell susceptibility to parasite infection. Interestingly, this susceptibility was abolished by treatment of Tc α GalVes with α -galactosidase, suggesting that α Gal epitopes are required in this process (Torrecilhas and Nakayasu *et al.*, unpublished data). However, little is known about the composition of

TcαGalVes. By immunological assays, Tc85 and αGal epitopes were shown to be present in TcαGalVes (Goncalves *et al*, 1991) (Torrecilhas and Nakayasu *et al.*, unpublished data). Thus, the initial proposal of my dissertation included the study of the molecular composition of TcαGalVes and the correlation of these structures with their biological activities.

My initial proposal had two major aims. In the **specific aim #1**, I had proposed to perform a detailed proteomic analysis of TcαGalVes to identify their antigens and virulence factors. As discussed above, both proinflammatory and virulence activities of TcαGalVes involving TLR2-mediated pathways are clearly related to protein post-translational modifications (PTMs), such as glycosylation (i.e., α-galactosylation) and GPI-anchoring of abundant *T. cruzi* proteins like mucins and TS/gp85 glycoproteins. Thus, in the **specific aim #2**, I intended to perform the analysis of protein post-translational modification (PTM) of TcαGalVes. Therefore, this aim included the identification and characterization of GPI-anchored glycoproteins and their GPI anchors, and free GPIs (glycoinositolphospholipids, GIPLs) of TcαGalVes. Also, we included in this aim the analysis of phosphoproteins of TcαGalVes that could be eventually be playing a role in signaling pathways triggered by these vesicles.

The results of my dissertation are presented in 5 chapters. Chapter 1 covers specific aim 1, whereas Chapters 3-5 cover specific aim 2. In **Chapter 1**, I show the detailed analysis of the proteome of TcαGalVes. Next, my original plan was to analyze the GPI-anchored proteins and GPI anchors from TcαGalVes, since we found several proteins that had putative GPI-anchoring sites. However, the methods available at the time to analyze GPIs were time-consuming and required large amounts of material, which could take up to one year just collecting enough the material to perform the experiment. Instead, in **Chapter 2** I show the development of a highly sensitive method to analyze GPI-anchored proteins and GPIs based

on polystyrene-divinylbenzene reverse-phase chromatography (RPC) coupled to tandem mass spectrometry (MSⁿ). We tested this technique to analyze GPI-anchored proteins and free GPIs (GIPs) from epimastigotes, since they are easier to grow in culture and obtain large amounts. Our analysis has shown that the GPIome (the collection of free and protein-linked GPIs) from epimastigote is much more complex than previously described.

In our proteomic analysis of TcαGalVes (Chapter 1), no mucin sequence was detected, even though the highly reactivity with Ch anti-αGal antibodies suggested that these glycoproteins were present in the vesicles. Moreover, these glycoproteins were detected with specific antibodies against their C-terminal region. The absence of these sequences in our proteomic analysis, might be due to two reasons: 1) these proteins are highly glycosylated, thus the generated peptides are difficult to analyze by standard or conventional proteomic approaches; 2) hundreds of TcMUC II mucin genes are expressed at the same time, decreasing the stoichiometry of individual proteins. To test the second hypothesis, in **Chapter 3**, we tested a efficient method to solubilize membrane proteins from insect-dwelling stages of *T. cruzi*, using a neutral detergent. Following the same issue to analyze membrane glycoproteins, in **Chapter 4** we developed a highly sensitive to analyze peptides by two-dimensional liquid chromatography coupled to mass spectrometry (2D LC-MS/MS). Using this approach to analyze the whole bloodstream trypomastigote cell lysate, we could identify several mucins and other important surface molecules, such as mucin-associate surface proteins (MASP) and TS/gp85 glycoproteins.

In the specific aim 2, I originally intended to map the phosphorylation sites of TcαGalVes proteins. However, due to the lack of experience to perform this kind of experiment, in **Chapter 4** we analyzed the phosphoproteome of whole epimastigote cell lysate. With this analysis we could unambiguously map 220 phosphorylation sites from 144

phosphoproteins. These results brought new insights about the parasite biological processes and possible drug targets.

In sum, despite the fact that I could not complete all sub-aims of specific aim 2 (i.e., GPI and phosphoproteome analysis of TccGalVes) as planned, the techniques developed in this dissertation can be used not only for eventually accomplishing those sub-aims, but also opened much broader perspectives for the structural and functional analysis of key molecules of *T. cruzi*.

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.
- Affranchino JL, Ibanez CF, Luquetti AO, Rassi A, Reyes MB, Macina RA, Aslund L, Pettersson U, Frasch AC (1989) Identification of a *Trypanosoma cruzi* antigen that is shed during the acute phase of Chagas' disease. *Mol Biochem Parasitol* **34**: 221-228.
- 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, Ferguson MA, Schenkman S, Travassos LR (1994a) GPI-anchored glycoconjugates from *Trypanosoma cruzi* trypomastigotes are recognized by lytic anti-alpha-galactosyl antibodies isolated from patients with chronic Chagas' disease. *Braz J Med Biol Res* **27**: 443-447.
- Almeida IC, Ferguson MA, Schenkman S, Travassos LR (1994b) Lytic anti-alpha-galactosyl antibodies from patients with chronic Chagas' disease recognize novel O-linked oligosaccharides on mucin-like glycosyl-phosphatidylinositol-anchored glycoproteins of *Trypanosoma cruzi*. *Biochem J* **304 (Pt 3)**: 793-802.

Almeida IC, Krautz GM, Krettli AU, Travassos LR (1993) Glycoconjugates of *Trypanosoma cruzi*: a 74 kD antigen of trypomastigotes specifically reacts with lytic anti-alpha-galactosyl antibodies from patients with chronic Chagas disease. *J Clin Lab Anal* **7**: 307-316.

Almeida IC, Milani SR, Gorin PA, Travassos LR (1991) Complement-mediated lysis of *Trypanosoma cruzi* trypomastigotes by human anti-alpha-galactosyl antibodies. *J Immunol* **146**: 2394-2400.

Alves MJ, Abuin G, Kuwajima VY, Colli W (1986) Partial inhibition of trypomastigote entry into cultured mammalian cells by monoclonal antibodies against a surface glycoprotein of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **21**: 75-82.

Aparicio IM, Scharfstein J, Lima AP (2004) A new cruzipain-mediated pathway of human cell invasion by *Trypanosoma cruzi* requires trypomastigote membranes. *Infect Immun* **72**: 5892-5902.

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 A, Jr., 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, Frasch AC, Di Noia JM (2006) Trypanosoma cruzi surface mucins: host-dependent coat diversity. *Nat Rev Microbiol* **4**: 229-236.

Campos MA, Almeida IC, Takeuchi O, Akira S, Valente EP, Procopio DO, Travassos LR, Smith JA, Golenbock DT, Gazzinelli RT (2001) Activation of Toll-like receptor-2 by glycosylphosphatidylinositol anchors from a protozoan parasite. *J Immunol* **167**: 416-423.

Campos MA, Closel M, Valente EP, Cardoso JE, Akira S, Alvarez-Leite JI, Ropert C, Gazzinelli RT (2004) Impaired production of proinflammatory cytokines and host resistance to acute infection with Trypanosoma cruzi in mice lacking functional myeloid differentiation factor 88. *J Immunol* **172**: 1711-1718.

Chagas C (1909) Nova tripanosomíase humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi*, n.g., n. Sp., agente etiológico de nova entidade morbida do homem. . *Mem Inst Oswaldo Cruz* **1**: 159-218.

Couto AS, Goncalves MF, Colli W, de Lederkremer RM (1990) The N-linked carbohydrate chain of the 85-kilodalton glycoprotein from Trypanosoma cruzi trypomastigotes contains sialyl, fucosyl and galactosyl (alpha 1-3)galactose units. *Mol Biochem Parasitol* **39**: 101-107.

Cuevas IC, Cazzulo JJ, Sanchez DO (2003) gp63 homologues in Trypanosoma cruzi: surface antigens with metalloprotease activity and a possible role in host cell infection. *Infect Immun* **71**: 5739-5749.

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.

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, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**: 409-415.

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

Gazzinelli RT, Pereira ME, Romanha A, Gazzinelli G, Brener Z (1991) Direct lysis of *Trypanosoma cruzi*: a novel effector mechanism of protection mediated by human anti-gal antibodies. *Parasite Immunol* **13**: 345-356.

Giordano R, Fouts DL, Tewari D, Colli W, Manning JE, Alves MJ (1999) Cloning of a surface membrane glycoprotein specific for the infective form of *Trypanosoma cruzi* having adhesive properties to laminin. *J Biol Chem* **274**: 3461-3468.

Goncalves MF, Umezawa ES, Katzin AM, de Souza W, Alves MJ, Zingales B, Colli W (1991) *Trypanosoma cruzi*: shedding of surface antigens as membrane vesicles. *Exp Parasitol* **72**: 43-53.

Magdesian MH, Giordano R, Ulrich H, Juliano MA, Juliano L, Schumacher RI, Colli W, Alves MJ (2001) Infection by *Trypanosoma cruzi*. Identification of a parasite ligand and its host cell receptor. *J Biol Chem* **276**: 19382-19389.

Monteiro AC, Schmitz V, Svensjo E, Gazzinelli RT, Almeida IC, Todorov A, de Arruda LB, Torrecilhas AC, Pesquero JB, Morrot A, Bouskela E, Bonomo A, Lima AP, Muller-Esterl W, Scharfstein J (2006) Cooperative activation of TLR2 and bradykinin B2 receptor is required for induction of type 1 immunity in a mouse model of subcutaneous infection by *Trypanosoma cruzi*. *J Immunol* **177**: 6325-6335.

Norris KA (1996) Ligand-binding renders the 160 kDa *Trypanosoma cruzi* complement regulatory protein susceptible to proteolytic cleavage. *Microb Pathog* **21**: 235-248.

Norris KA, Bradt B, Cooper NR, So M (1991) Characterization of a *Trypanosoma cruzi* C3 binding protein with functional and genetic similarities to the human complement regulatory protein, decay-accelerating factor. *J Immunol* **147**: 2240-2247.

Norris KA, Schrimpf JE (1994) Biochemical analysis of the membrane and soluble forms of the complement regulatory protein of *Trypanosoma cruzi*. *Infect Immun* **62**: 236-243.

Ouaissi A, Guilvard E, Delneste Y, Caron G, Magistrelli G, Herbault N, Thieblemont N, Jeannin P (2002) The *Trypanosoma cruzi* Tc52-released protein induces human dendritic cell maturation, signals via Toll-like receptor 2, and confers protection against lethal infection. *J Immunol* **168**: 6366-6374.

Pereira-Chioccola VL, Acosta-Serrano A, Correia de Almeida I, Ferguson MA, Souto-Padron T, Rodrigues MM, Travassos LR, Schenkman S (2000) Mucin-like molecules form a negatively charged coat that protects *Trypanosoma cruzi* trypomastigotes from killing by human anti-alpha-galactosyl antibodies. *J Cell Sci* **113 (Pt 7)**: 1299-1307.

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.

Ropert C, Almeida IC, Closel M, Travassos LR, Ferguson MA, Cohen P, Gazzinelli RT (2001) Requirement of mitogen-activated protein kinases and I kappa B phosphorylation for induction of proinflammatory cytokines synthesis by macrophages indicates functional similarity of receptors triggered by glycosylphosphatidylinositol anchors from parasitic protozoa and bacterial lipopolysaccharide. *J Immunol* **166**: 3423-3431.

Ropert C, Closel M, Chaves AC, Gazzinelli RT (2003) Inhibition of a p38/stress-activated protein kinase-2-dependent phosphatase restores function of IL-1 receptor-associate kinase-1 and reverses Toll-like receptor 2- and 4-dependent tolerance of macrophages. *J Immunol* **171**: 1456-1465.

Rubin-de-Celis SS, Uemura H, Yoshida N, Schenkman S (2006) Expression of trypomastigote trans-sialidase in metacyclic forms of *Trypanosoma cruzi* increases parasite escape from its parasitophorous vacuole. *Cell Microbiol* **8**: 1888-1898.

Schenkman S, Eichinger D, Pereira ME, Nussenzweig V (1994) Structural and functional properties of *Trypanosoma* trans-sialidase. *Annu Rev Microbiol* **48**: 499-523.

Schenkman S, Jiang MS, Hart GW, Nussenzweig V (1991) A novel cell surface trans-sialidase of *Trypanosoma cruzi* generates a stage-specific epitope required for invasion of mammalian cells. *Cell* **65**: 1117-1125.

Stuart K, Brun R, Croft S, Fairlamb A, Gurtler RE, McKerrow J, Reed S, Tarleton R (2008) Kinetoplastids: related protozoan pathogens, different diseases. *J Clin Invest* **118**: 1301-1310.

Tarleton RL, Reithinger R, Urbina JA, Kitron U, Gurtler RE (2007) The challenges of Chagas Disease-- grim outlook or glimmer of hope. *PLoS Med* **4**: e332.

Tobler LH, Contestable P, Pitina L, Groth H, Shaffer S, Blackburn GR, Warren H, Lee SR, Busch MP (2007) Evaluation of a new enzyme-linked immunosorbent assay for detection of Chagas antibody in US blood donors. *Transfusion* **47**: 90-96.

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

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 U S A* **105**: 5022-5027.

Yao C, Donelson JE, Wilson ME (2003) The major surface protease (MSP or GP63) of *Leishmania* sp. Biosynthesis, regulation of expression, and function. *Mol Biochem Parasitol* **132**: 1-16.

Chapter 1: The Proteome of *Trypanosoma cruzi* Shed Vesicles Involved in Host Cell Invasion

Ernesto S. Nakayasu ¹, Ana C. T. Torrecilhas ², Fabio C. Gozzo ³, Lilian L. Nohara ¹, Luciane Ganiko ¹, Juan C. Noveron ⁴, Douglas J. Lamont ⁵, Maria Julia M. Alves ², Michael A. J. Ferguson ⁵, and Igor C. Almeida ^{1*}

¹ *Department of Biological Sciences, The Border Biomedical Research Center, University of Texas at El Paso, El Paso, TX 79968;* ² *Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, SP 05599-970, Brazil;* ³ *National Laboratory of Synchrotron Light, Campinas, SP 13084-971, Brazil;* ⁴ *Department of Chemistry, University of Texas at El Paso (UTEP), El Paso, TX 79968; and* ⁵ *Post Genomics and Molecular Interactions Centre and Division of Biological Chemistry and Molecular Microbiology, School of Life Sciences, The Wellcome Trust Biocentre, University of Dundee, Dundee, DD1 5EH, UK.*

*Corresponding author: E-mail: icalmeida@utep.edu; Tel: (915)747-6086; Fax: (915)747-5808.

Keywords: Proteome, *Trypanosoma cruzi*, shed vesicles, host-cell invasion

1.1 Abstract

Infective trypomastigote forms of the protozoan parasite *Trypanosoma cruzi* release *in vitro* membrane-bound vesicles resembling mammalian cell exosomes. We have recently shown that these vesicles are able to induce potent host proinflammatory immune response and greatly enhance parasite virulence. Here, we have performed a detailed proteomic analysis of *T. cruzi* shed vesicles aiming at a better understanding of their molecular composition and function. Parasite vesicles were obtained from the conditioned medium of mammalian cell-derived trypomastigote forms and fractionated by gel-filtration, followed by affinity chromatography using immobilized anti- α -galactosyl antibodies from patients with chronic *T. cruzi* infection. Fractions enriched of *T. cruzi* vesicles were digested with three different proteolytic strategies to increase protein coverage. Resulting peptides were fractionated by strong cation-exchange chromatography followed by reverse-phase chromatography, and analyzed by electrospray ionization mass spectrometry. Using this approach, we have identified 110 *T. cruzi*-specific proteins. Over half (55%) of those were found to be *trans*-sialidase/gp85 glycoproteins, well-known to play a key role in host-cell adhesion and invasion by the parasite. We have also found several polypeptides related to mammalian exosomes, suggesting a similar biogenesis for these parasite vesicles. Our data clearly demonstrate that vesicles secreted by *T. cruzi* trypomastigotes contain the major virulence factors responsible for promoting the parasite entry into the host cell.

1.2 Introduction

Chagas' disease or American trypanosomiasis is caused by the protozoan parasite *Trypanosoma cruzi*. In Latin America, it is estimated that over 11 million people are already infected and approximately 120 million people are at risk of acquiring the disease. Of those infected, 13,000-50,000 may die every year due to complications during acute or chronic phase of the disease (Dias *et al*, 2002). Chagas' disease threat is not restricted to Latin America. Recent estimates indicate that at least 25,000-100,000 migrants from endemic countries living in the U.S. are thought to be chronically infected with *T. cruzi*, thus representing a latent risk for recipients of blood components and organs (Leiby, 2004). In fact, an increasing number of cases of Chagas' disease in the U.S. has been recently reported (CDC, 2006; Young *et al*, 2007).

T. cruzi is transmitted by hematophagous insects from the Reduviidae family, as well as by blood transfusion, organ transplantation, or congenital route (Barrett *et al*, 2003). In the vector-mediated transmission, infective metacyclic trypomastigote forms of the parasite, present in the insect excrement, enter the mammalian host bloodstream through the insect bite wound or exposed host mucosal tissues, and rapidly invade various types of nucleated cells. Once inside the cells, metacyclic forms escape the parasitophorous vacuoles and transform into the non-flagellate amastigote forms, which proliferate freely in the cytoplasm. Three to four days later the amastigotes convert into trypomastigote forms, which then rupture the cell plasma membrane and reach the extracellular milieu to infect other cells and tissues or, eventually, another vector (Barrett *et al*, 2003; Tyler and Engman, 2001).

T. cruzi trypomastigote cell surface is covered by a dense and diverse repertoire of glycosylphosphatidylinositol (GPI)-anchored glycoproteins, such as mucins and *trans*-sialidase (TS)/gp85 glycoproteins (Acosta-Serrano *et al*, 2001; Acosta-Serrano *et al*, 2006; Buscaglia *et*

al, 2006; Frasch, 2000). Mucins are the major surface glycoproteins of the cell surface of trypomastigote forms of *T. cruzi* (Buscaglia *et al*, 2004; Buscaglia *et al*, 2006). They are rich in O-linked oligosaccharides containing terminal, non-reducing α -galactosyl (α Gal) epitopes, which are the major targets of lytic, protective Chagasic anti- α Gal antibodies (Ch anti- α Gal), predominant during both acute and chronic phases of Chagas' disease (Almeida *et al*, 1994; Almeida *et al*, 1993; Almeida *et al*, 1991; Gazzinelli *et al*, 1991). Mucins are also the major acceptors of sialic acid residues transferred from host cell glycoproteins and glycopeptides by the parasite TS (Schenkman *et al*, 1994; Schenkman *et al*, 1991). The addition of sialic acid residues to the protozoan cell coat confers protection to the parasite against lytic Ch anti- α Gal (Pereira-Chioccola *et al*, 2000). Recent genomic data revealed that the mucin multigene family is composed of 863 members (201 pseudogenes) (El-Sayed *et al*, 2005). Mucins are highly polymorphic and their expression varies according to the life cycle stage of the parasite, and the diversity of expression may be related to the capacity of *T. cruzi* to efficiently evade the host immune response (Buscaglia *et al*, 2004; Buscaglia *et al*, 2006). Recently, we have demonstrated that trypomastigote-derived mucins (tGPI-mucins) and their GPI anchors (tGPIs) are strong proinflammatory molecules (Almeida *et al*, 2000). The activation of macrophages by these parasite molecules leads to the production of nitric oxide (NO) and proinflammatory cytokines through a Toll-like receptor 2 (TLR2)-mediated signaling pathway (Almeida and Gazzinelli, 2001; Campos *et al*, 2001; Ropert *et al*, 2001).

The TS/gp85 glycoproteins are encoded by the most abundant *T. cruzi* multigene family, comprising 1430 genes (693 pseudogenes), which can be subdivided into three major groups (El-Sayed *et al*, 2005; Frasch, 2000). The first group encompasses glycoproteins of 85 kDa (gp85), which were shown to be important in the processes of parasite attachment to the host-cell surface and cell invasion (Giordano *et al*, 1999; Magdesian *et al*, 2001). The second

group comprises proteins of 160 kDa, which are involved in the regulation of the host complement system (Norris *et al*, 1991). The third group is composed by enzymatically active TS members that transfer sialic acid residues from host to parasite glycoconjugates, which are involved in host-cell infection and parasite evasion from host immune defense (Frasch, 2000; Schenkman *et al*, 1994; Schenkman *et al*, 1991).

Although these major surface GPI-anchored glycoproteins were shown to be essential to the establishment of the infection and elicitation of host immune response, the mechanism by which these molecules are presented to the host cell receptors and co-receptors still remains unclear. Several *T. cruzi* antigens were shown to be secreted by the parasite to the culture medium (Affranchino *et al*, 1989; Ouaisi *et al*, 1990; Silva *et al*, 2004). Gonçalves *et al*. (Goncalves *et al*, 1991) reported that infective bloodstream trypomastigote forms of *T. cruzi* secrete antigens bound to small membrane vesicles. These vesicles were recently shown to be rich in α Gal epitopes (Tc α GalVes - *T. cruzi* α Gal-containing vesicles) and able to strongly stimulate proinflammatory response through the activation of TLR2 in a way similar to the tGPI-mucins and tGPIs (Torrecilhas and Nakayasu *et al*, unpublished data) (Campos *et al*, 2001). Furthermore, the preincubation of mammalian cells with Tc α GalVes prior to host cell invasion assay resulted in a 6-8-fold increase in the number of infected cells and of intracellular parasites per cell. This activity was also shown to be dependent on TLR2 (Torrecilhas and Nakayasu *et al.*, unpublished data).

To determine which *T. cruzi* proteins could be present on Tc α GalVes and play a role in the stimulation of host immune response and cell invasion, we performed a detailed proteomic analysis of these vesicles. Tc α GalVes were purified by gel-filtration and affinity chromatography with immobilized Ch anti- α Gal, and proteins were digested and analyzed by liquid chromatography coupled to mass spectrometry (LC-MS). The rate of false-positive

identifications was calculated in order to validate only reliable hits. With this approach, one hundred and ten *T. cruzi*-specific proteins were identified. Over half (55%) of these proteins are related to infection and pathogenesis, which is consistent with the virulent nature of TcαGalVes. Several proteins that might be related to the biogenesis of these vesicles were also identified.

1.3 Materials and Methods

1.3.1 Cell cultures and harvest of *T. cruzi* vesicles

Tissue culture cell-derived trypomastigote forms of *T. cruzi* (Y strain) were obtained 5 to 9 days after infection of monolayers of green monkey kidney LLC-MK₂ epithelial cells (ATCC, Manassas, VA), maintained at 37°C, under 5% CO₂ atmosphere, in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), as described elsewhere (Andrews and Colli, 1982). Total *T. cruzi* shed vesicles were obtained from the culture medium supernatant of trypomastigotes. Briefly, parasites were counted, centrifuged (15 min, 1,500-2,000g, 10°C), and resuspended in DMEM supplemented with 5% FBS (previously depleted of its own exosomes by ultracentrifugation at 200,000g, 30 min 4°C) (Wubbolts *et al*, 2003), at a concentration of 1x10⁸ parasites/mL medium. After incubation for 2-3 h at 37°C, under 5% CO₂ atmosphere, trypomastigotes were removed by centrifugation (10 min, 3000g, 10°C), and the supernatant containing the total shed material was filtered through a 0.45 µm cartridge (Goncalves *et al*, 1991). Under these conditions, there was no significant increase in the number of dead parasites up to 3 h of incubation, as assessed by flow cytometry following the labeling of the parasites with propidium iodide (Figure 1.1). This indicates that the parasites are viable and, most likely that the shed vesicles are not formed as a result of cell death.

1.3.2 Fractionation of *T. cruzi* vesicles

The total shed material (250 µL) was mixed with 200 mM ammonium acetate, pH 6.5 (250 µL) and loaded onto a Sepharose CL-4B column (1 x 40 cm, GE Healthcare, Piscataway, NJ) pre-equilibrated with 100 mM ammonium acetate, pH 6.5. The column was eluted with the equilibration buffer, in a flow rate of 0.2 mL/min using a peristaltic pump (GE Healthcare).

Fractions of 1 mL were collected, and then screened by chemiluminescent enzyme-linked immunosorbent assay (CL-ELISA) using Ch anti- α Gal or anti-TcMUCII antibodies, as described elsewhere (Almeida *et al*, 2000; Buscaglia *et al*, 2004). To eliminate a considerable amount of FBS proteins still present following the gel-filtration chromatography, Tc α GalVes-containing fractions were submitted to affinity chromatography using Ch anti- α Gal immobilized onto cyanogen bromide-activated Sepharose CL-4B (Almeida *et al*, 1994). Pooled Tc α GalVes-containing fractions were incubated overnight with Ch anti- α Gal-Sepharose beads, previously equilibrated in 100 mM ammonium acetate, pH 7.0 (buffer C). The resin was extensively washed with buffer C containing 0.5 M NaCl, and then with buffer C again. Tc α GalVes were eluted with 1 M propionic acid and immediately lyophilized. We found that even after purification by affinity chromatography, Tc α GalVes were still contaminated with FBS proteins and lipids. Over 100 FBS-derived proteins and glycoproteins were found by LC-MS/MS analysis (data not shown). Phospholipids, especially phosphatidylinositols (PIs), most likely bound to FBS proteins (e.g., bovine serum albumin), were also detected by ESI-MS (data not shown). These FBS-derived contaminants consistently precluded an accurate quantitative analysis of Tc α GalVes-specific proteins, carbohydrates, and *myo*-inositol (in GPI-anchors). Therefore, here we express the Tc α GalVes concentration as equivalents of trypomastigotes.

1.3.3 Atomic force microscopy analysis of shed vesicles

Vesicles obtained from 10^8 trypomastigotes by gel-filtration chromatography as described above were analyzed by atomic force microscopy (AFM) to assess their size and integrity. All tapping-mode AFM measurements were acquired using a Veeco Multimode Nanoscope III (Woodbury, NY), equipped with a 10 μ m E scanner. Commercial Si_3N_3 cantilevers with tip radius of <10 nm (Nanoscience Instruments, Inc.) and force constant of 40

N/m were used. Images were recorded with typical scan rates of 1.0-2.0 Hz. Vesicles were imaged topographically on freshly cleaved mica plates and analyzed with NanoScope™ Software.

1.3.4 Protein digestion

Proteins from TcαGalVes were digested using three different proteolytic strategies: a) trypsin in NH₄HCO₃ buffer containing urea (TU); b) double-digestion with trypsin and endoproteinase Glu-C (TG); and c), trypsin in NH₄HCO₃ buffer containing 60% methanol (TM). The TU proteolytic strategy was carried out as described elsewhere (Stone and Williams, 1996). TcαGalVes proteins (equivalent to vesicles secreted by 2x10⁹ parasites) were dissolved in 20 μL 400 mM NH₄HCO₃, pH 8.0, containing 8 M urea. Reduction of disulfide bonds was carried out by the addition of 5 μL 45 mM dithiotreitol (DTT) and incubation for 15 min at 50°C. Cysteine residues were alkylated using 5 μL 100 mM iodoacetamide (IA) for 30 min, at room temperature and protected from the light. Then, ultrapure water was added to obtain a final concentration of 1 M urea. The digestion was performed with trypsin (2 μg), for 24 h at 37°C. For the TG strategy, after digestion using TU protocol, endoproteinase Glu-C (2 μg) was added. After incubation for 24 h at 37°C, the reaction was terminated by addition of trifluoroacetic acid (TFA) at a final concentration of 0.046%. The TM proteolytic strategy was carried out as described elsewhere (Goshe *et al*, 2003; Russell *et al*, 2001). The same amount of TcαGalVes proteins (equivalent to 2x10⁹ parasites) was dissolved in 100 μL 50 mM NH₄HCO₃, pH 8.0/HPLC-grade methanol (40/60) (v/v) (BM buffer), and heated for 5 min at 95°C. Disulfide bounds were reduced with 5 mM DTT, for 30 min at 37°C. Cysteine residues were alkylated with 10 mM IA at room temperature for 30 min, in the dark. The incubation mixture was then diluted to a final volume of 300 μL in BM buffer. The digestion was carried

out with trypsin (2 µg) for 24 h at 37°C. The reaction was terminated by the addition of TFA at a final concentration of 0.046% for peptide desalting, or 300 µL 1% formic acid (FA) for peptide fractionation by strong cation exchange (SCX) chromatography.

1.3.5 Peptide desalting and ion exchange chromatography

For LC-MS analysis, digested peptides were desalted using an in-house C18 zip-tip built by placing 50 µL of 100 mg/mL C₁₈-resin (10-15 µm, 300Å, Vydac, Hesperia, CA) in isopropanol on the top of a very fine glass wool layer, inside a 200-µL pipette tip (Axygen, Union City, CA). The zip-tip was activated with methanol and equilibrated with 0.046% TFA prior to the sample loading. After washing with 0.046% TFA, peptides were eluted in 80% acetonitrile (ACN)/0.046% TFA. For 2D-LC-MS analysis, peptides digested by the TM (for ESI-QTOF-MS analysis) or TU (for ESI-Qtrap-MS analysis) strategy were loaded in an in-house-assembled SCX zip-tip, built as the C18 zip-tip described above, with the exception that 20 µL of SCX resin (POROS HS50, Applied Biosystems, Framingham, MA) were used instead of C₁₈ resin. The zip-tip was equilibrated with 25% ACN/0.5% FA, the sample was loaded, and the microcolumn washed with equilibration buffer. Peptides were eluted with a stepwise gradient (0, 50, 100, 200, 300, 400, and 500 mM for ESI-QTOF-MS analysis, and 0, 50, 100, 200, and 500 mM for ESI-Qtrap-MS analysis) of NaCl dissolved in the equilibration buffer. All samples were dried in a vacuum centrifuge (VacufugeTM, Eppendorf, USA) before LC-MS analysis.

1.3.6 Liquid chromatography-mass spectrometry (LC-MS) analysis

For LC-ESI-ion trap (IT)-MS analysis, peptides were dissolved in 20 µL of 0.05% FA, and 1 µL was loaded into an in-house-manufactured 10-cm capillary column, assembled with C₁₈ resin in silica-fused capillary (100 µm i.d., 360 µm o.d., Polymicro, Phoenix, AZ) (Gatlin *et*

al, 1998). The run was performed in a Ultimate nanoHPLC system (LC Packings, Dionex, Sunnyvale, CA) connected to a LCQ Duo ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA), equipped with a nanospray source (Thermo Electron). Peptides were eluted in a linear gradient of 0-40% solvent B (solvent A = 5% ACN/0.2% FA; solvent B = 80% ACN/0.2% FA), in 120 or 240 min at 500 nL/min flow rate. The nanospray source was set at 1.9 kV, and the desolvation temperature was 180°C. The MS mode scan spectra were collected in a range between 300 and 2000 atomic mass units (a.m.u.). Each ion with intensity higher than 5×10^5 counts was automatically subjected only once to a zoom scan (10-a.m.u. range), followed by a MS/MS scan (selection width of 2.0 a.m.u., and normalized collision energy of 35%), and then dynamically excluded for 1 min.

For LC-ESI-QTOF-MS or LC-ESI-Qtrap-MS analysis, peptides were recovered in 10 μ L 5% FA, and then 5-fold diluted in 0.1% TFA. Twenty microliters of each SCX fraction were loaded onto a C₁₈ pre-column (LC Packings, Dionex) using a FamosTM autosampler (LC Packings, Dionex), and washed with 0.1% FA. Peptide separation was carried out using a commercial capillary column (PepMap, C₁₈, 5 μ m 300Å, 75 μ m x 15 cm, LC Packings, Dionex) coupled to an Ultimate nanoHPLC system. The run was performed using a linear gradient of 5–40% solvent B' (solvent A' = 0.1% FA; solvent B' = 90% ACN/0.08% FA) over 200 min, at 200 nL/min flow rate. Eluted peptides were analyzed on-line using an ESI-QTOF-MS (QSTAR® XL Hybrid LC/MS/MS System, Applied Biosystems/MDS SCIEX, Foster City, CA), or an ESI-Qtrap-MS (Qtrap 4000, Applied Biosystems/MDS SCIEX), both equipped with nanospray sources (Applied Biosystems/MDS SCIEX). For ESI-QTOF-MS analysis, MS spectra were collected each second in a range between 400 and 2000 a.m.u. Each doubly- and triply-charged peptide, with intensity higher than 10 counts, was subjected to fragmentation only once, using rolling collision energy. Each MS/MS spectra were collected for

4 sec at 50–1800 a.m.u. range. For ESI-Qtrap-MS analysis, each ion with intensity above 5×10^4 counts was subjected to an enhanced resolution scan, at 250 a.m.u./sec rate, followed by a MS/MS scan, using rolling collision energy. Each MS/MS spectra were collected using a variable mass range according to peptide charge state, at 4000 a.m.u./sec rate.

1.3.7 Database search, protein identification and sequence analysis

Raw data from ESI-IT-MS were converted into DTA files using BioWorks version 3.0 software (Thermo Fisher Scientific). Only MS/MS spectra from peptides between 500 and 3500 Da and more than 15 fragment ions were accepted. ESI-QTOF-MS and ESI-Qtrap-MS raw data were converted into peak lists (TMP files) using default parameters of the standard Mascot script for Analyst (available in Analyst QS 1.0 for ESI-QTOF-MS, and in Analyst 1.4.1 for ESI-Qtrap-MS, Applied Biosystems). All tandem MS spectra were correlated to *T. cruzi* (19,613 sequences, TcruziDB version 5.0 (Aguero *et al*, 2006), downloaded from <http://tcruzidb.org>) and bovine protein sequences (51,591 sequences, downloaded on November 3, 2005, from GenBank, <http://ncbi.nlm.nih.gov>) using TurboSequest (Eng *et al*, 1994) (Bioworks 3.0, Thermo Fisher Scientific, San Jose, CA), Mascot 2.1 (Perkins *et al*, 1999) (Matrix Science, UK), and Phenyx (Phenyx Virtual Desktop 1.1.10) algorithms (Colinge *et al*, 2004; Colinge *et al*, 2003) (GeneBio, Switzerland). All the parameters for database searching are summarized in Table 1.1.

TurboSequest analysis was performed using a database containing TcruziDB version 5.0 added of contaminant protein sequences from bovine serum (456 sequences). These bovine sequences were selected from previous searches against bovine database using TurboSequest ($X_{\text{corr}} \geq 1.8$, 2.0 and 2.5 for singly, doubly- and triply-charged peptides, respectively; and a $\Delta C_n \geq 0.1$) and Phenyx (Z peptide score ≥ 6.0) softwares. For

TurboSequest analysis, cysteine carbamidomethylation and methionine oxidation were considered as fixed and variable modification, respectively. A tolerance of 2.0 Da was set for peptide and 1.0 Da for MS/MS fragments mass accuracy. In addition, we only regarded as valid peptides those with a $\Delta C_n \geq 0.1$; and an Xcorr equal or higher than 1.9, 2.2, and 3.7, for mono-, doubly- and triply-charged peptides, respectively (Washburn *et al*, 2001).

For Mascot analysis, cysteine carbamidomethylation was considered as a fixed modification and methionine oxidation as a variable modification. Peptide mass tolerance was set to 100 ppm for ESI-QTOF-MS; 200 ppm for ESI-Qtrap-MS; and 2.0 Da, for ESI-IT-MS data. MS/MS fragment mass tolerance was set at 1.0 Da to ESI-IT-MS and 0.8 Da for ESI-QTOF-MS and ESI-Qtrap-MS. Phenyx searches were performed in two rounds. In first round, the same tolerance parameters for Mascot searches were used, but only cysteine carbamidomethylation was considered as a modification. In second round, the methionine oxidation was included as a variable modification. In addition, ESI-QTOF-MS data tolerance was increased to 200 ppm, and semi-specific cleavages (for peptides that result from trypsin or Glu-C cleavage at one terminus and from an endogenous protease cleavage at the other terminus) were allowed. To estimate the false-discovery rate (FDR) of protein identification, Mascot and Phenyx analyses were performed in a database built with TcruziDB version 5.0 and contaminant sequences from bovine serum, added by the reverse TcruziDB version 5.0 database and 100,000 randomly generated sequences, forming a database with 139,682 sequences. Peptides and proteins that matched to forward *T. cruzi* and bovine sequences were considered true-positive identifications, and hits that matched with reverse *T. cruzi* and random sequences were considered false-positive identifications. The FDR was calculated as the ratio between the numbers of false-positive identifications over the total number of identifications. After FDR estimation, Mascot data were validated with an ion score cutoff ≥ 20

(Tables 1.2 and 1.3) and a minimum protein score from 36 to 52, depending on the dataset (Table 1.2). Phenyx data was validated with a minimum Z-peptide score from 5 to 7, and a protein score from 7 to 11, depending on the dataset (Table 1.2). In addition, the spectra from proteins identified by a single hit were manually checked and the poor-matching hits were deleted.

All valid *T. cruzi* protein sequences were compared to sequences deposited into GenBank using Blast tool (<http://www.ncbi.nlm.nih.gov/BLAST/>) (Altschul *et al*, 1990). Gene ontology (GO) annotation was assigned by similarity searches against the Swiss-Prot and TrEMBL databases (invertebrate taxonomy, which includes all eukaryotic entries except those from vertebrates, fungi, and plants) using GOblet tool, which is available online at <http://goblet.molgen.mpg.de> (Groth *et al*, 2004). Only GO results from proteins with e-value $\leq 1e-10$ for database search were accepted. This analysis was performed on May 12, 2007.

Identified proteins were also analyzed for prediction of potential glycosylphosphatidylinositol (GPI)-anchoring sites using the following algorithms: DGPI (http://129.194.185.165/dgpi/index_en.html), big-PI Predictor (http://mendel.imp.univie.ac.at/gpi/gpi_server.html) (Eisenhaber *et al*, 1999), and GPI-SOM (<http://gpi.unibe.ch/>) (Fankhauser and Maser, 2005).

1.4 Results

We initially fractionated the total shed vesicles from *T. cruzi* trypomastigotes by gel-filtration chromatography in Sepharose CL-4B. Eighty 1-ml fractions were collected and screened by CL-ELISA using Ch anti- α Gal and a pool of sera from patients with chronic Chagas' disease (ChS pool), as previously described (Almeida *et al.*, 2000). A major included peak, highly reactive with both Ch anti- α Gal and ChS pool, was observed. This reactivity was mostly abolished by treatment with green-coffee bean α -galactosidase (Torrecilhas and Nakayasu *et al.*, unpublished data). By atomic force microscopy, α -galactosyl-positive fractions showed to contain intact vesicles of ~200 nm in diameter (Figure 1.2). As expected, preliminary LC-MS analysis revealed that the α Gal-containing vesicular fractions were contaminated with numerous proteins from the FBS (data not shown). To further enrich the *T. cruzi* α Gal-containing vesicles, we carried out an affinity chromatography with immobilized Ch anti- α Gal. The acid eluate, highly enriched in α Gal-containing vesicles (Tc α GalVes), was lyophilized, digested with three different proteolytic strategies (TU, TG, and TM), and subjected to LC-MS/MS analysis. Collected spectra were correlated to *T. cruzi* and bovine sequences using Phenyx, Mascot, and TurboSequest algorithms (Figure 1.3).

To increase protein coverage, we have applied a combination of different digestion strategies, LC-MS analysis, and database search algorithms. Since the samples were not normalized by the number of runs and amount of protein, we were unable to compare the performance of different proteolytic strategies and mass spectrometers. However, a significant increase in protein coverage was observed when all data were combined. One example is the putative TS (TcruziDB accession number Tc00.1047053509187.10). Mascot search of ESI-QTOF-MS data resulted in the identification of 11 peptides for this protein, but the combination all of analyses increased the coverage to 20 peptides (Table 1.4). However, we were unable to

compare the performance of database search algorithms, because data collected with ESI-QTOF-MS and ESI-Qtrap-MS were not compatible to TurboSequest algorithm, but as shown in Figure 1.3, the use of these different database search algorithms increased the number of identified peptides and proteins. Results from Phenyx, Mascot, and TurboSequest analyses led to the identification of 54 (191 peptides), 48 (130 peptides), and 47 (52 peptides) proteins, respectively, totaling 110 *T. cruzi*-specific proteins (282 peptides) (Figure 1.4).

By using different proteolytic strategies, mass spectrometers, and database search algorithms, one major concern was the reliability of protein identifications. To determine the reliability of individual datasets, an estimation of the FDR for Phenyx and Mascot analyses was performed using a database containing *T. cruzi*, bovine, and random sequences. This approach made possible the simultaneous identification of proteins and estimation of their FDR. Proteins were validated with confidence interval greater than 95% (Table 1.2). Only one analysis (TcαGalVes proteins digested with TG strategy and analyzed by ESI-IT-MS/Phenyx) was allowed to have 6% of FDR, because it represented only one false-positive identification, out of 15 true-positive identifications, thus not affecting the global FDR. It was not possible to estimate the FDR from the TurboSequest analysis, since the BioWorks 3.0 version does not have the capability to assemble the data from different runs, and the number of proteins present in single runs was not sufficient for calculating the FDR. To validate the data from TurboSequest analysis, we applied the threshold parameters proposed by Washburn et al. (Washburn *et al*, 2001) and used to validate *Plasmodium sp.* proteome (Florens *et al*, 2002), which are far more strict than those proposed by Peng et al. (Peng *et al*, 2003). In addition, we manually validated the protein identifications based on single peptide hits, removing the poor-matching peptides, thus resulting in a lower FDR.

Another problem faced during the annotation of TcαGalVes proteomic analysis was that the Mascot analysis generated several redundant hits. To eliminate these redundant hits, we have tested different ion-score cutoffs (0, 15, or 20). We found that using an ion score cutoff of 20, practically all redundant hits were eliminated from the Mascot analysis (Table 1.3).

After calculating the FDR and eliminating redundant hits, combined results from Phenyx, Mascot, and TurboSequest analyses resulted in the identification of 110 proteins specific from *T. cruzi* (Table 1.3), and 120 bovine proteins from the cell culture medium (data not shown). TcαGalVes were shown to be rich in cell surface proteins, particularly TS/gp85 glycoproteins (n=61). All members from this superfamily were annotated as putative *trans*-sialidases in the GenBank during the genome annotation, which made it difficult to obtain more detailed information about these proteins. To better correlate identified TS/gp85 members to their functions, they were annotated not only for the top hit in Blast analysis, but also for other relevant hits (Table 1.4). With this approach, we have found proteins that have significant similarity to Tc85, gp90, TS, shed-acute-phase-antigen (SAPA), complement-regulatory protein (CRP), flagellum-associated protein, trypomastigote ligand, 85 kDa surface antigen, c71 surface protein, and surface protein/glycoprotein. As to other cell surface glycoproteins, we have also found 3 members of the gp63 family and 1 member of the mucin-associated surface protein (MASP) family. Members of TS/gp85, gp63, and MASP superfamilies were proposed to be GPI-anchored (El-Sayed *et al*, 2005). To verify this, all validated *T. cruzi* protein sequences were analyzed by the DGPI, GPI-SOM, and big-PI Predictor algorithms. With this approach we have found 42 sequences (39 members of TS/gp85, 2 members of gp63, and 1 member of MASP superfamilies) with potential GPI-anchoring sites (Table 1.4).

Another important class of proteins found in this proteomic analysis was the cytoskeleton-related proteins, such as alpha- and beta-tubulin, kinesin, and myosin (Table

1.4). These cytoskeleton proteins are usually observed in the proteomic analysis of exosomes (Thery *et al*, 2001; Wubbolts *et al*, 2003). We have also identified other proteins related to exosomes and vacuoles, such as heat shock protein 85 (HSP85), elongation factor 1-alpha, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), vacuolar ATP synthase subunit B, and tetratricopeptide repeat (TPR) protein (Table 1.4).

TcαGalVes were also shown to contain proteins related to flagellum, such as flagellum-associated protein (which belongs to the TS/gp85 superfamily), paraflagellar rod protein 3, and flagellar calcium-binding protein (FCaBP). We have also identified proteins with diverse functions, such as kinases, trypanothione peroxidase, trifunctional enzyme alpha subunit, putative transporter, amidinotransferase family protein, putative glutamic acid/alanine-rich protein, Viral A-type inclusion protein repeat, and Aardvark. Finally, 16 proteins were assigned as hypothetical proteins, having no similarity or other relevant Blast hit (Table 1.4).

To better understand their function, all identified proteins from TcαGalVes were submitted to gene ontology (GO) annotation using GOblast algorithm and searching sequences against the TrEMBL and Swiss-Prot databases. This analysis resulted in the annotation of molecular function, biological process, and cellular component of 79% of 110 identified proteins (Figure 1.5 and Table 1.5). TcαGalVes were shown to be rich in proteins related to pathogenesis (54%) (included in the GO:0044419 - *interspecies interaction between organisms*) and catalytic activity (68%), which includes hydrolases, peptidases, and kinases.

1.5 Discussion

In a previous study, *T. cruzi* trypomastigotes were shown to secrete proteins as membrane-bound vesicles to the culture medium (Goncalves *et al*, 1991). More recently, we have demonstrated that these vesicles strongly induce host proinflammatory response and increase the parasite ability to invade host cells (Torrecilhas and Nakayasu *et al.*, unpublished data). Thus far, the molecular composition of these vesicles has been poorly understood. Nevertheless, we have determined that they are rich in α Gal epitopes (Torrecilhas and Nakayasu *et al.*, unpublished data). *T. cruzi* α Gal-rich vesicles (or Tc α GalVes) were also recognized by the monoclonal antibody H1A10 (Torrecilhas and Nakayasu *et al.*, unpublished data), which is specific for Tc85 glycoprotein belonging to the TS/gp85 superfamily (Goncalves *et al*, 1991). To better understand the molecular composition and function of Tc α GalVes, we have performed an extensive proteomic analysis of these vesicles by LC-MS/MS.

To obtain a good coverage of the proteomic analysis, we have applied the combination of different proteolytic strategies, mass spectrometers, and database search algorithms. The application of different database search algorithms increased the number of identified proteins, which can be explained by the fact that different statistics are used by each database search algorithm to calculate peptide and protein scores. Mascot and Phenyx are probability-based algorithms, whereas TurboSequest is based on the quality of spectrum for a given sequence on the database (heuristic-based algorithm) (Kapp *et al*, 2005). Our results suggest the idea of complementarity of these algorithms, which was already observed when same sets of data were analyzed with Phenyx and TurboSequest (Heller *et al*, 2005), or Mascot and TurboSequest (Elias *et al*, 2005).

The application of various datasets could increase not only the proteome coverage, but also the number of false-positive identifications. To assure the quality of our analysis, the

number of false-positive identifications was estimated and the data were validated with a FDR $\leq 5\%$. Another problem faced during validation of our data was that the Mascot analysis included low-score hits as diagnostic peptides to validate new sequences. This procedure invariably leads to the identification of redundant proteins and artificially increases the proteome coverage. The redundant identification of proteins is a special concern in *T. cruzi* proteomic analysis, since this parasite possesses several multigene families with numerous members that share highly conserved domains (El-Sayed *et al*, 2005). Weatherly *et al*. (Weatherly *et al*, 2005) proposed to validate redundant protein hits as protein groups. This approach was also applied to validate a large-scale proteomic analysis of *T. cruzi* (Atwood *et al*, 2005). However, we have decided to validate only the most probable sequence. In case of more than one protein sharing the same peptides, only the first sequence that appeared in database was validated in our analysis. We have applied this criterion because the inclusion of redundant hits can lead to an overestimation of the number of identified proteins. This procedure may also lead to an increase in the number of “true-positive hits” by the addition of redundant proteins as valid hits, which can artificially decrease the FDR. To eliminate redundant hits from TcαGalVes proteomic analysis, the ion-score cutoff was increased to 20, which practically abolished redundant identifications.

Here, we have shown that the combination of data from multiple analyses resulted in the identification of 110 *T. cruzi*-specific proteins. The most representative group of identified proteins was the TS/gp85 family of glycoproteins. Since there is a correlation between the number of sequenced peptides and the abundance of proteins (Ishihama *et al*, 2005), it is also possible to state that these glycoproteins are the most abundant as well, with some members having up to 20 peptides identified (Table 1.4). TS activity has also been detected in TcαGalVes (Alves *et al.*, personal communication). In agreement with these results, we have

found some sequences that appear to have enzymatic activity, such as TS itself and SAPA. The latter is an active TS protein with a tandem amino acid repeat at the C-terminal region (Affranchino *et al*, 1989). This tandem amino acid repeat was proposed to increase SAPA's half-life in the host bloodstream (Buscaglia *et al*, 1999), to be a *T. cruzi* virulence factor *in vivo* (Chuenkova and Pereira, 1995; Risso *et al*, 2004), and to stimulate the production of IL-6 in intestinal microvascular endothelial cells and peripheral blood mononuclear cells (Saavedra *et al*, 1999).

Members from gp85 group of the TS/gp85 superfamily, such as Tc85-11, Tc85-16, surface glycoprotein gp90, 85 kDa surface antigen, surface protein, and surface glycoprotein, were also found in the TcαGalVes proteomic analysis. One of the best characterized proteins from this group is the Tc85, which was proposed to be an adhesion molecule (Giordano *et al*, 1999; Magdesian *et al*, 2001). Monoclonal antibodies against Tc85 partially inhibit the host-cell invasion by the parasite (Alves *et al*, 1986). Another important group from TS/gp85 superfamily comprises proteins with 160 kDa (Frasch, 2000). In our proteomic analysis, we have identified the flagellum-associated protein and CRP. The latter was shown to possess the ability to bind specifically to complement factors C3b and C4b (Norris *et al*, 1991; Norris and Schrimpf, 1994). Following CRP binding to C3b or C4b, a protease processes these complement factors and deactivate them. This protease has not yet been characterized but preliminary studies showed that it migrates as a protein of ~75 kDa in polyacrylamide zymogram gels, and its activity is inhibited by aprotinin, leupeptin, and EDTA (Norris, 1996). Recent studies have shown that gp63 from *T. cruzi* has about the same molecular mass (Cuevas *et al*, 2003), and that gp63 from *Leishmania* ssp. seems to have the ability to inactivate the complement-mediated lysis (Yao *et al*, 2003). Therefore, *T. cruzi* gp63 could be a good candidate as complement-inactivating enzyme. In addition, *T. cruzi* gp63 was also shown to be important for

host-cell invasion by the parasite (Cuevas *et al*, 2003). Interestingly, we have identified 3 sequences from gp63 in our proteomic analysis. Other papers reported a protease activity in supernatant of trypomastigote culture medium. One of these activities was inhibited by E-64, and was proposed to be from the major parasite cysteine proteinase cruzipain (Yokoyama-Yasunaka *et al*, 1994). In addition, the activity of cruzipain combined with vesicles secreted by trypomastigotes was reported to be an important virulence factor (Aparicio *et al*, 2004; Cazzulo, 2002; Scharfstein, 2006). Cruzipain activity remains in the supernatant of trypomastigote culture medium even after ultracentrifugation (Aparicio *et al*, 2004). In agreement with this result, we were unable to detect cruzipain by proteomic analysis of Tc α GalVes.

Torreilhas and Nakayasu *et al*. (unpublished data) have recently shown by immunoblotting and CL-ELISA with Ch anti- α Gal that Tc α GalVes are rich in α Gal epitopes. Thus, we anticipated that tGPI-mucins were also present in these vesicles. However, it was not possible to identify any tGPI-mucin sequence in our proteomic analysis. Previous studies with highly purified tGPI-mucins suggested that these molecules are difficult to analyze by conventional proteomic approaches, due to their amino acid composition, sequence diversity, and post-translational modifications (Buscaglia *et al*, 2004). In agreement with this latter study and our current data, the comprehensive proteomic analysis from all four *T. cruzi* stages has also failed to identify those proteins (Atwood *et al*, 2005). To further investigate the presence of tGPI-mucins in Tc α GalVes we performed an ELISA using antibodies specific for the C-terminus of TcMUCII mucins, and we were able to confirm the presence of these molecules in the vesicles (Figure 1.6).

Interestingly, we have found one sequence of MASP, which was also identified in the comparative proteomic analysis of the four *T. cruzi* life cycle stages (Atwood *et al*, 2005).

MASP is a multigene family, whose member sequences have potential GPI-anchoring, and O- and N-glycosylation sites (El-Sayed *et al*, 2005). The biological function of MASP still remains to be determined. A recent paper reported a family of serine-, alanine-, and proline-rich proteins (SAPs) (Baida *et al*, 2006), with sequence similarity to MASP. SAPs were shown to be expressed and secreted to the culture medium by metacyclic trypomastigotes forms, and also to play a role in host cell invasion and Ca^{2+} signaling (Baida *et al*, 2006).

Some pieces of evidence led us to the idea that Tc α GalVes contain GPI-anchored proteins: (i) members of TS/gp85 and gp63 were shown to be GPI-anchored (Abuin *et al*, 1996; Cuevas *et al*, 2003); (ii) Tc α GalVes induce proinflammatory response similar to tGPIs (Torrecilhas and Nakayasu *et al.*, unpublished data); and (iii) after treatment of Tc α GalVes with proteinase K and extraction with butanol/water partition, part of proinflammatory activity was recovered in the organic phase, which is enriched with GPI anchors (Torrecilhas and Nakayasu *et al.*, unpublished data). To identify protein sequences with potential GPI-anchoring sites, all sequences identified by proteomic analysis of Tc α GalVes were analyzed by DGPI, GPI-SOM, and big-PI Predictor algorithms. Forty-two proteins with potential GPI-anchoring sites were identified (Table 1.4). However, it still remains to be determined whether all and not only part of the members of those three superfamilies (i.e., TS/gp85, gp63, and MASP) found in Tc α GalVes are indeed GPI-anchored. The available GPI-prediction tools are not 100% accurate, and some sequences from the *T. cruzi* genome project appear to be incorrectly annotated or incomplete. Most of GPI-anchored-predicted proteins are members of TS/gp85 superfamily, but we have also identified two sequences from gp63 and one sequence from MASP. In agreement with these results, *T. cruzi* gp63 is located on the parasite surface and can be artificially released by the treatment with phosphatidylinositol-specific phospholipase C (PI-PLC) (Cuevas *et al*, 2003). Another surface glycoprotein, SAPA, was proposed to be shed

to the trypomastigote culture medium in a soluble form by the action of a *T. cruzi* PI-PLC (Rosenberg *et al*, 1991). Here we show that SAPA is shed to the culture medium attached to membrane-bound vesicles. Nevertheless, we do not discard the possibility that SAPA is secreted to the culture medium in both forms. More recently, however, we have shown that immunodominant shed antigens from *T. cruzi* trypomastigotes are eluted in very high relative molecular mass fraction from Sepharose CL-4B column, suggesting that most antigens secreted by *T. cruzi* are present as membrane-bound vesicles (Torrecilhas and Nakayasu *et al*, unpublished data).

Our finding of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (2 peptides) (Table 1.3), an abundant parasite glycolytic enzyme, could suggest a contamination of our vesicular preparation with glycosomes. However, this is unlikely to have occurred, because we were unable to detect any other glycolytic enzymes in our analyses. It is worth pointing out that, more recently, a series of studies have shown that GAPDH is a multi-functional protein involved in several other key cellular processes (e.g., transcription, tubulin assembly, and nitric oxide-mediated apoptosis) rather than just glycolysis. Therefore, here we could speculate that *T. cruzi* GAPDH may play a role in the pathogenesis of Chagas' disease in a similar fashion to that of GAPDH from *Streptococcus pyogenes* (Terao *et al*, 2006). The latter was shown to be a cell surface protein, also secreted to the culture medium, and able to bind to complement factor C5a, which is then deactivated by the bacterial protease ScpA. This process seemed to be essential for *S. pyogenes* evasion from neutrophils (Terao *et al*, 2006).

To better correlate the identified proteins to the virulent nature of TcαGalVes, all identified sequences were submitted to gene ontology analysis (Figure 1.5). In agreement with this idea, we have found that over half of the identified proteins were annotated as being involved in pathogenesis.

A remaining question is about the process of TcαGalVes formation or biogenesis. One possibility is that TcαGalVes may be formed as apoptotic bodies or through a biogenesis similar to exosomes (Keller *et al*, 2006; Thery *et al*, 2002). Exosomes are membrane-bound vesicles (30-100 nm) secreted by various types of mammalian cells. They are believed to have two major mechanisms of biogenesis: 1) reverse budding of plasma membrane domains; or 2) extracellular release of multivesicular bodies (MVBs), formed by invagination of the limiting membrane of late endosomes (Couzin, 2005; Keller *et al*, 2006; Thery *et al*, 2002). During the process of harvesting TcαGalVes, the parasites retain their viability (Figure 1.1), as well as infectivity (Goncalves *et al*, 1991), suggesting that cell death does not occur during the process of vesicle secretion. Supporting the idea of a controlled biogenesis, TcαGalVes were shown to be secreted mainly by the flagellar pocket (Torrecilhas and Nakayasu *et al*, unpublished data). In our proteomic analysis, we have found several proteins related to exosomes and vacuoles, such as HSP85, cytoskeleton proteins, elongation factors, GAPDH, and tetratricopeptide repeat (TPR)-containing protein. TPRs are found in a variety of proteins with diverse functions. For instance, HSP90 from mammalian cells binds to immunophilin through its TPR domain and also to dynein through a distinct domain, forming a complex with HSP70 and p23, which is responsible to transport the glucocorticoid receptor via microtubules from cytosol to the nucleus (Pratt *et al*, 2004). TPRs are also found in several proteins from the complex responsible to kinesin- and dynein-dependent intraflagellar transport of proteins in *Chlamydomonas reinhardtii* (Cole, 2003). TPR- containing protein VPS4 AAA-type ATPase interacts with endosomal sorting complexes required for transport (ESCRT) during the formation of MVBs. Thus, HSP85, cytoskeleton proteins, and TPR protein found in TcαGalVes may play a role in sorting proteins and/or in the vesicle formation (Scott *et al*, 2005). The biogenesis of TcαGalVes is currently under investigation.

1.6 Conclusions

The proteomic analysis of Tc α GalVes revealed a reasonable large number of proteins related to virulence, pathogenesis, and proinflammatory activity. These findings may explain the ability of these vesicles to modulate host inflammatory response and to increase parasite infectivity. We propose that Tc α GalVes components could be eventually explored as molecular targets for the development of new therapeutic approaches to prevent or treat Chagas' disease.

1.7 Acknowledgments

The work described was supported by the grants 1RO1AI070655 (to ICA) from the National Institute of Allergy and Infectious Diseases (NIAID) and 5G12RR008124 (to the Border Biomedical Research Center (BBRC)/University of Texas at El Paso-UTEP) from the National Center for Research Resources (NCRR), both components of the National Institutes of Health (NIH). The contents of this study are solely the responsibility of the authors and do not necessarily represent the official views of NIAID, NCRR, or NIH. We like to thank the Biomolecule Analysis Core Facility (BCAF)/UTEP for the use of the equipment. The BCAF is supported by grant number 5G12RR008124 to the BBRC/UTEP from the NIH/NCRR. This study was also supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (grant 98/10495-5), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, and The Wellcome Trust, UK. We thank Dr Nilson Assunção and Roberto Zangradi (USP) for the technical assistance, Dr Carlos Buscaglia for providing the anti-TcMUCII antibodies, and Dr Sirlei Daffre (USP) for the full access to the LCQ-Duo and continuous support.

1.8 References

- Abuin G, Couto AS, de Lederkremer RM, Casal OL, Galli C, Colli W, Alves MJ (1996) Trypanosoma cruzi: the Tc-85 surface glycoprotein shed by trypomastigotes bears a modified glycosylphosphatidylinositol anchor. *Exp Parasitol* **82**: 290-297.
- 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 (2006) Comparison and evolution of the surface architecture of trypanosomatid parasites. In *Trypanosomes: After the genome*: Horizon Press.
- Affranchino JL, Ibanez CF, Luquetti AO, Rassi A, Reyes MB, Macina RA, Aslund L, Pettersson U, Frasch AC (1989) Identification of a Trypanosoma cruzi antigen that is shed during the acute phase of Chagas' disease. *Mol Biochem Parasitol* **34**: 221-228.
- 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**: D428-431.
- 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, Ferguson MA, Schenkman S, Travassos LR (1994) Lytic anti-alpha-galactosyl antibodies from patients with chronic Chagas' disease recognize novel O-linked oligosaccharides on mucin-like glycosyl-phosphatidylinositol-anchored glycoproteins of *Trypanosoma cruzi*. *Biochem J* **304 (Pt 3)**: 793-802.

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

Almeida IC, Krautz GM, Krettli AU, Travassos LR (1993) Glycoconjugates of *Trypanosoma cruzi*: a 74 kD antigen of trypomastigotes specifically reacts with lytic anti-alpha-galactosyl antibodies from patients with chronic Chagas disease. *J Clin Lab Anal* **7**: 307-316.

Almeida IC, Milani SR, Gorin PA, Travassos LR (1991) Complement-mediated lysis of *Trypanosoma cruzi* trypomastigotes by human anti-alpha-galactosyl antibodies. *J Immunol* **146**: 2394-2400.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Alves MJ, Abuin G, Kuwajima VY, Colli W (1986) Partial inhibition of trypomastigote entry into cultured mammalian cells by monoclonal antibodies against a surface glycoprotein of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **21**: 75-82.

Andrews NW, Colli W (1982) Adhesion and interiorization of *Trypanosoma cruzi* in mammalian cells. *J Protozool* **29**: 264-269.

Aparicio IM, Scharfstein J, Lima AP (2004) A new cruzipain-mediated pathway of human cell invasion by *Trypanosoma cruzi* requires trypomastigote membranes. *Infect Immun* **72**: 5892-5902.

Atwood JA, 3rd, Weatherly DB, Minning TA, Bundy B, Cavola C, Opperdoes FR, Orlando R, Tarleton RL (2005) The *Trypanosoma cruzi* proteome. *Science* **309**: 473-476.

Baida RC, Santos MR, Carmo MS, Yoshida N, Ferreira D, Ferreira AT, El Sayed NM, Andersson B, da Silveira JF (2006) Molecular characterization of serine-, alanine-, and proline-rich proteins of *Trypanosoma cruzi* and their possible role in host cell infection. *Infect Immun* **74**: 1537-1546.

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

Buscaglia CA, Alfonso J, Campetella O, Frasch AC (1999) Tandem amino acid repeats from *Trypanosoma cruzi* shed antigens increase the half-life of proteins in blood. *Blood* **93**: 2025-2032.

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.

Campos MA, Almeida IC, Takeuchi O, Akira S, Valente EP, Procopio DO, Travassos LR, Smith JA, Golenbock DT, Gazzinelli RT (2001) Activation of Toll-like receptor-2 by glycosylphosphatidylinositol anchors from a protozoan parasite. *J Immunol* **167**: 416-423.

Cazzulo JJ (2002) Proteinases of *Trypanosoma cruzi*: potential targets for the chemotherapy of Chagas disease. *Curr Top Med Chem* **2**: 1261-1271.

CDC (2006) Chagas disease after organ transplantation--Los Angeles, California, 2006. *MMWR Morb Mortal Wkly Rep* **55**: 798-800.

Chuenkova M, Pereira ME (1995) *Trypanosoma cruzi* trans-sialidase: enhancement of virulence in a murine model of Chagas' disease. *J Exp Med* **181**: 1693-1703.

Cole DG (2003) The intraflagellar transport machinery of *Chlamydomonas reinhardtii*. *Traffic* **4**: 435-442.

Colinge J, Masselot A, Cusin I, Mahe E, Niknejad A, Argoud-Puy G, Reffas S, Bederr N, Gleizes A, Rey PA, Bougueleret L (2004) High-performance peptide identification by tandem

mass spectrometry allows reliable automatic data processing in proteomics. *Proteomics* **4**: 1977-1984.

Colinge J, Masselot A, Giron M, Dessingy T, Magnin J (2003) OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics* **3**: 1454-1463.

Couzin J (2005) Cell biology: The ins and outs of exosomes. *Science* **308**: 1862-1863.

Cuevas IC, Cazzulo JJ, Sanchez DO (2003) gp63 homologues in *Trypanosoma cruzi*: surface antigens with metalloprotease activity and a possible role in host cell infection. *Infect Immun* **71**: 5739-5749.

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.

Eisenhaber B, Bork P, Eisenhaber F (1999) Prediction of potential GPI-modification sites in proprotein sequences. *J Mol Biol* **292**: 741-758.

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Wortley 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, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado

CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**: 409-415.

Elias JE, Haas W, Faherty BK, Gygi SP (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* **2**: 667-675.

Eng JK, McCormack AL, Yates JRr (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. . *J Am Soc Mass Spectrom* **5**: 976-989.

Fankhauser N, Maser P (2005) Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics* **21**: 1846-1852.

Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N, Sacci JB, Tabb DL, Witney AA, Wolters D, Wu Y, Gardner MJ, Holder AA, Sinden RE, Yates JR, Carucci DJ (2002) A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**: 520-526.

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

Gatlin CL, Kleemann GR, Hays LG, Link AJ, Yates JR, 3rd (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, Pereira ME, Romanha A, Gazzinelli G, Brener Z (1991) Direct lysis of *Trypanosoma cruzi*: a novel effector mechanism of protection mediated by human anti-gal antibodies. *Parasite Immunol* **13**: 345-356.

Giordano R, Fouts DL, Tewari D, Colli W, Manning JE, Alves MJ (1999) Cloning of a surface membrane glycoprotein specific for the infective form of *Trypanosoma cruzi* having adhesive properties to laminin. *J Biol Chem* **274**: 3461-3468.

Goncalves MF, Umezawa ES, Katzin AM, de Souza W, Alves MJ, Zingales B, Colli W (1991) *Trypanosoma cruzi*: shedding of surface antigens as membrane vesicles. *Exp Parasitol* **72**: 43-53.

Goshe MB, Blonder J, Smith RD (2003) Affinity labeling of highly hydrophobic integral membrane proteins for proteome-wide analysis. *J Proteome Res* **2**: 153-161.

Groth D, Lehrach H, Hennig S (2004) GOblet: a platform for Gene Ontology annotation of anonymous sequence data. *Nucleic Acids Res* **32**: W313-317.

Heller M, Ye M, Michel PE, Morier P, Stalder D, Junger MA, Aebersold R, Reymond F, Rossier JS (2005) Added value for tandem mass spectrometry shotgun proteomics data validation through isoelectric focusing of peptides. *J Proteome Res* **4**: 2273-2282.

Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**: 1265-1272.

Kapp EA, Schutz F, Connolly LM, Chakel JA, Meza JE, Miller CA, Fenyo D, Eng JK, Adkins JN, Omenn GS, Simpson RJ (2005) An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. *Proteomics* **5**: 3475-3490.

Keller S, Sanderson MP, Stoeck A, Altevogt P (2006) Exosomes: from biogenesis and secretion to biological function. *Immunol Lett* **107**: 102-108.

Leiby DA (2004) Threats to blood safety posed by emerging protozoan pathogens. *Vox Sang* **87 Suppl 2**: 120-122.

Magdesian MH, Giordano R, Ulrich H, Juliano MA, Juliano L, Schumacher RI, Colli W, Alves MJ (2001) Infection by *Trypanosoma cruzi*. Identification of a parasite ligand and its host cell receptor. *J Biol Chem* **276**: 19382-19389.

Norris KA (1996) Ligand-binding renders the 160 kDa *Trypanosoma cruzi* complement regulatory protein susceptible to proteolytic cleavage. *Microb Pathog* **21**: 235-248.

Norris KA, Bradt B, Cooper NR, So M (1991) Characterization of a *Trypanosoma cruzi* C3 binding protein with functional and genetic similarities to the human complement regulatory protein, decay-accelerating factor. *J Immunol* **147**: 2240-2247.

Norris KA, Schimpf JE (1994) Biochemical analysis of the membrane and soluble forms of the complement regulatory protein of *Trypanosoma cruzi*. *Infect Immun* **62**: 236-243.

Ouaissi MA, Taibi A, Cornette J, Velge P, Marty B, Loyens M, Esteva M, Rizvi FS, Capron A (1990) Characterization of major surface and excretory-secretory immunogens of *Trypanosoma cruzi* trypomastigotes and identification of potential protective antigen. *Parasitology* **100 Pt 1**: 115-124.

Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* **2**: 43-50.

Pereira-Chioccola VL, Acosta-Serrano A, Correia de Almeida I, Ferguson MA, Souto-Padron T, Rodrigues MM, Travassos LR, Schenkman S (2000) Mucin-like molecules form a negatively charged coat that protects *Trypanosoma cruzi* trypomastigotes from killing by human anti-alpha-galactosyl antibodies. *J Cell Sci* **113 (Pt 7)**: 1299-1307.

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.

Pratt WB, Galigniana MD, Harrell JM, DeFranco DB (2004) Role of hsp90 and the hsp90-binding immunophilins in signalling protein movement. *Cell Signal* **16**: 857-872.

Risso MG, Garbarino GB, Mocetti E, Campetella O, Gonzalez Cappa SM, Buscaglia CA, Leguizamon MS (2004) Differential expression of a virulence factor, the trans-sialidase, by the main *Trypanosoma cruzi* phylogenetic lineages. *J Infect Dis* **189**: 2250-2259.

Ropert C, Almeida IC, Closel M, Travassos LR, Ferguson MA, Cohen P, Gazzinelli RT (2001) Requirement of mitogen-activated protein kinases and I kappa B phosphorylation for induction of proinflammatory cytokines synthesis by macrophages indicates functional similarity of receptors triggered by glycosylphosphatidylinositol anchors from parasitic protozoa and bacterial lipopolysaccharide. *J Immunol* **166**: 3423-3431.

Rosenberg I, Prioli RP, Ortega-Barria E, Pereira ME (1991) Stage-specific phospholipase C-mediated release of *Trypanosoma cruzi* neuraminidase. *Mol Biochem Parasitol* **46**: 303-305.

Russell WK, Park ZY, Russell DH (2001) Proteolysis in mixed organic-aqueous solvent systems: applications for peptide mass mapping using mass spectrometry. *Anal Chem* **73**: 2682-2685.

Saavedra E, Herrera M, Gao W, Uemura H, Pereira MA (1999) The Trypanosoma cruzi trans-sialidase, through its COOH-terminal tandem repeat, upregulates interleukin 6 secretion in normal human intestinal microvascular endothelial cells and peripheral blood mononuclear cells. *J Exp Med* **190**: 1825-1836.

Scharfstein J (2006) Parasite cysteine proteinase interactions with alpha2-macroglobulin or kininogens: differential pathways modulating inflammation and innate immunity in infection by pathogenic trypanosomatids. *Immunobiology* **211**: 117-125.

Schenkman S, Eichinger D, Pereira ME, Nussenzweig V (1994) Structural and functional properties of Trypanosoma trans-sialidase. *Annu Rev Microbiol* **48**: 499-523.

Schenkman S, Jiang MS, Hart GW, Nussenzweig V (1991) A novel cell surface trans-sialidase of Trypanosoma cruzi generates a stage-specific epitope required for invasion of mammalian cells. *Cell* **65**: 1117-1125.

Scott A, Gaspar J, Stuchell-Brereton MD, Alam SL, Skalicky JJ, Sundquist WI (2005) Structure and ESCRT-III protein interactions of the MIT domain of human VPS4A. *Proc Natl Acad Sci U S A* **102**: 13813-13818.

Silva AG, Silveira-Lacerda EP, Cunha-Junior JP, de Souza MA, Favoreto Junior S (2004) Immunoblotting analyses using two-dimensional gel electrophoresis of Trypanosoma cruzi excreted-secreted antigens. *Rev Soc Bras Med Trop* **37**: 454-459.

Stone KL, Williams KR (1996) Enzymatic digestion of proteins in solution and in SDS polyacrilamide gel. In *The protein protocol handbook.*, Walker JM (ed). Totowa, NJ: Humana Press Inc.

Terao Y, Yamaguchi M, Hamada S, Kawabata S (2006) Multifunctional Glyceraldehyde-3-phosphate Dehydrogenase of *Streptococcus pyogenes* Is Essential for Evasion from Neutrophils. *J Biol Chem* **281**: 14215-14223.

Thery C, Boussac M, Veron P, Ricciardi-Castagnoli P, Raposo G, Garin J, Amigorena S (2001) Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles. *J Immunol* **166**: 7309-7318.

Thery C, Zitvogel L, Amigorena S (2002) Exosomes: composition, biogenesis and function. *Nat Rev Immunol* **2**: 569-579.

Tyler KM, Engman DM (2001) The life cycle of *Trypanosoma cruzi* revisited. *Int J Parasitol* **31**: 472-481.

Washburn MP, Wolters D, Yates JR, 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**: 242-247.

Weatherly DB, Astwood JA, 3rd, Minning TA, Cavola C, Tarleton RL, Orlando R (2005) A Heuristic method for assigning a false-discovery rate for protein identifications from Mascot database search results. *Mol Cell Proteomics* **4**: 762-772.

Wubbolts R, Leckie RS, Veenhuizen PT, Schwarzmann G, Mobius W, Hoernschemeyer J, Slot JW, Geuze HJ, Stoorvogel W (2003) Proteomic and biochemical analyses of human B cell-derived exosomes. Potential implications for their function and multivesicular body formation. *J Biol Chem* **278**: 10963-10972.

Yao C, Donelson JE, Wilson ME (2003) The major surface protease (MSP or GP63) of *Leishmania* sp. Biosynthesis, regulation of expression, and function. *Mol Biochem Parasitol* **132**: 1-16.

Yokoyama-Yasunaka JK, Pral EM, Oliveira Junior OC, Alfieri SC, Stolf AM (1994) *Trypanosoma cruzi*: identification of proteinases in shed components of trypomastigote forms. *Acta Trop* **57**: 307-315.

Young C, Losikoff P, Chawla A, Glasser L, Forman E (2007) Transfusion-acquired *Trypanosoma cruzi* infection. *Transfusion* **47**: 540-544.

Table 1.1 Database search parameters.

ESI-QTOF-MS dataset	Mascot	Phenyx 1st round	Phenyx 2nd round	TurboSequest
Database	TcruziDBv5.0 (forward and reverse) bovine sequences random sequences	TcruziDBv5.0 (forward and reverse) bovine sequences random sequences	TcruziDBv5.0 (forward and reverse) bovine sequences random sequences	ND¶
Peptide mass tolerance	100 ppm	100 ppm	200 ppm	ND
Fragment mass tolerance	0.8 Da	0.8 Da	0.8 Da	ND
Fixed modification	cysteine carbamidomethylation	cysteine carbamidomethylation	cysteine carbamidomethylation	ND
Variable modification	methionine oxidation	none	methionine oxidation	ND
Enzyme specificity	K and R, but not before P	K and R, but not before P	semi-specific K and R, but not before P	ND
Max. # missed cleavage sites	1	1	1	ND
ESI-Qtrap-MS dataset				
Database	TcruziDBv5.0 (forward and reverse) bovine sequences random sequences	TcruziDBv5.0 (forward and reverse) bovine sequences random sequences	TcruziDBv5.0 (forward and reverse) bovine sequences random sequences	ND
Peptide mass tolerance	200 ppm	200 ppm	200 ppm	ND
Fragment mass tolerance	0.8 Da	0.8 Da	0.8 Da	ND
Fixed modification	cysteine carbamidomethylation	cysteine carbamidomethylation	cysteine carbamidomethylation	ND
Variable modification	methionine oxidation	none	methionine oxidation	ND
Enzyme specificity	K and R, but not before P	K and R, but not before P	semi-specific K and R, but not before P	ND
Max. # missed cleavage sites	1	1	1	ND
ESI-IT-MS dataset				
Database	TcruziDBv5.0 (forward and reverse) bovine sequences random sequences	TcruziDBv5.0 (forward and reverse) bovine sequences random sequences	TcruziDBv5.0 (forward and reverse) bovine sequences random sequences	TcruziDBv5.0(forward) bovine sequences
Peptide mass tolerance	2 Da	2 Da	2 Da	2 Da
Fragment mass tolerance	1 Da	1 Da	1 Da	1 Da
Fixed modification	Cysteine carbamidomethylation	Cysteine carbamidomethylation	Cysteine carbamidomethylation	Cysteine carbamidomethylation
Variable modification	methionine oxidation	none	methionine oxidation	methionine oxidation
Enzyme specificity	K and R, but not before P for TU and TM strategies; and K, R, and E for TG strategy	K and R, but not before P for TU and TM strategies; and K, R, and E for TG strategy	semi-specific K and R, but not before P for TU and TM strategies; and semi-specific K, R, and E for TG strategy	K and R, but not before P for TU and TM strategies; and K, R, and E for TG strategy
Max. # missed cleavage sites	1 for TU and TM strategies, and 3 for TG strategy	1 for TU and TM strategies, and 3 for TG strategy	1 for TU and TM strategies, and 3 for TG strategy	1 for TU and TM strategies, and 3 for TG strategy

Table 1.2 Statistics of TcαGalVes proteomic analysis.*

Analysis Type	Equipment	No. of Runs	No. of MS/MS Spectra	Digestion Strategy	Algorithm	Protein Score	Peptide Score	Protein Level			Peptide Level		
								Bovine	<i>T. cruzi</i>	FPR (%)	Bovine	<i>T. cruzi</i>	FPR (%)
1D-LC/MS	ESI-IT-MS	3	4253	TU	Mascot	≥ 45	≥ 20	28	9	2.6	259	21	0.7
1D-LC/MS	ESI-IT-MS	2	3167	TM	Mascot	≥ 45	≥ 20	20	2	0.0	183	4	0.0
1D-LC/MS	ESI-IT-MS	2	2005	TG	Mascot	≥ 52	≥ 20	7	1	0.0	69	1	0.0
2D-LC/MS	ESI-Qtrap-MS	5	9690	TU	Mascot	≥ 36	≥ 20	28	16	4.3	156	21	1.1
2D-LC/MS	ESI-QTOF-MS	7	8314	TM	Mascot	≥ 39	≥ 20	76	39	2.5	443	111	0.5
1D-LC/MS	ESI-IT-MS	3	4253	TU	Phenyx	≥ 11	≥ 7	34	30	3.0	375	101	1.9
1D-LC/MS	ESI-IT-MS	2	3167	TM	Phenyx	≥ 11	≥ 7	26	11	2.6	200	19	4.4
1D-LC/MS	ESI-IT-MS	2	2005	TG	Phenyx	≥ 11	≥ 7	13	2	6.2	99	6	2.7
2D-LC/MS	ESI-Qtrap-MS	5	9690	TU	Phenyx	≥ 8.7	≥ 6	36	10	4.3	169	16	1.6
2D-LC/MS	ESI-QTOF-MS	7	8314	TM	Phenyx	≥ 7	≥ 5	76	30	4.7	352	72	1.9
1D-LC/MS	ESI-IT-MS	3	4253	TU	TurboSequest	– ¶	– ¶	42	37	ND †	164	42	ND
1D-LC/MS	ESI-IT-MS	2	3167	TM	TurboSequest	– ¶	– ¶	30	18	ND	125	19	ND
1D-LC/MS	ESI-IT-MS	2	2005	TG	TurboSequest	– ¶	– ¶	11	11	ND	54	11	ND

* Confidence interval > 95%.

¶ TurboSequest results were validated with $\Delta Cn \geq 0.1$; and an Xcorr ≥ 1.9 , 2.2, and 3.7, for 1+, 2+ and 3+ peptides, respectively.

† ND, Not determined.

Table 1.3 Redundant hits from Mascot analysis.

Equipment (Digestion strategy) *	Ion score cutoff	Score	Total	Bovine	<i>T. cruzi</i>	Random	FPR (%)	Redundant
ESI-QTOF-MS (TM)	0	≥ 39	171	76	86	9	5.3	57
ESI-QTOF-MS (TM)	15	≥ 39	120	76	41	3	2.5	1
<i>ESI-QTOF-MS (TM)</i>	20	≥ 39	118	76	39	3	2.5	0
ESI-Qtrap-MS (TU)	0	≥ 36	76	34	38	4	5.3	31
ESI-Qtrap-MS (TU)	15	≥ 36	49	28	18	3	6.1	3
<i>ESI-Qtrap-MS (TU)</i>	20	≥ 36	46	28	16	2	4.3	0
ESI-IT-MS (TG)	0	≥ 52	9	8	1	0	0.0	1
ESI-IT-MS (TG)	15	≥ 52	8	7	1	0	0.0	1
<i>ESI-IT-MS (TG)</i>	20	≥ 52	8	7	1	0	0.0	1
ESI-IT-MS (TM)	0	≥ 45	24	20	3	1	4.2	2
ESI-IT-MS (TM)	15	≥ 45	22	20	2	0	0.0	0
<i>ESI-IT-MS (TM)</i>	20	≥ 45	22	20	2	0	0.0	0
ESI-IT-MS (TU)	0	≥ 45	45	29	14	2	4.4	8
ESI-IT-MS (TU)	15	≥ 45	38	28	9	1	2.6	1
<i>ESI-IT-MS (TU)</i>	20	≥ 45	38	28	9	1	2.6	1

* Parameters used for the validation of Mascot search analysis are shown in bold, italics.

Table 1.4 *T. cruzi* proteins identified by proteomic analysis of TcαGalVes.

							GPI-anchoring prediction		
TcruziDB accession number		No. of peptides	Blast top hit	NCBI*	Other relevant blast hit	NCBI *	DGPI	Big PI	GPI-SOM
TS/gp85 superfamily									
1	Tc00.1047053509187.10	20	trans-sialidase, putative	EAN84046	surface glycoprotein Tc-85/16	AAP32148			
2	Tc00.1047053507839.40	18	trans-sialidase, putative	EAN85078	surface glycoprotein Tc-85/16	AAP32148		Y	
3	Tc00.1047053507427.10	18	trans-sialidase, putative	EAN83233	trans-sialidase	BAA09334	Y	Y	Y
4	Tc00.1047053511779.60	16	trans-sialidase, putative	EAN85138	Flagellum-Associated Protein	CAA50287		Y	Y
5	Tc00.1047053505609.10	15	trans-sialidase, putative	EAN86711	surface glycoprotein Tc-85/16	AAP32148	Y		
6	Tc00.1047053503759.10	15	trans-sialidase, putative	EAN84030	surface glycoprotein Tc-85/16	AAP32148		Y	
7	Tc00.1047053506961.25	13	trans-sialidase, putative	EAN96545	shed-acute-phase-antigen	CAA40511	Y	Y	Y
8	Tc00.1047053426675.9	12	trans-sialidase, putative	EAN80725	surface glycoprotein	AAA19650	Y	Y	Y
9	Tc00.1047053511311.30	11	trans-sialidase, putative	EAN93342	surface glycoprotein Tc-85/16	AAP32148		Y	
10	Tc00.1047053507047.40	11	trans-sialidase, putative	EAN94520	surface glycoprotein Tc-85/16	AAP32148		Y	
11	Tc00.1047053510787.10	10	trans-sialidase, putative	EAN84377	shed-acute-phase-antigen	CAA40511		Y	Y
12	Tc00.1047053398477.10	9	trans-sialidase, putative	EAN82320	surface glycoprotein Tc-85/16	AAP32148		Y	
13	Tc00.1047053509513.10	9	trans-sialidase, putative	EAN83217	surface glycoprotein Tc-85/16	AAP32148		Y	
14	Tc00.1047053508485.10	8	trans-sialidase, putative	EAN84281	surface glycoprotein Tc-85/16	AAP32148			
15	Tc00.1047053506331.130	8	trans-sialidase, putative	EAN96370	surface glycoprotein Tc-85/16	AAP32148		Y	
16	Tc00.1047053509875.80	7	trans-sialidase, putative	EAN98221	surface glycoprotein Tc-85/16	AAP32148		Y	
17	Tc00.1047053508139.240	7	trans-sialidase, putative	EAN98605	surface glycoprotein Tc-85/16	AAP32148			
18	Tc00.1047053509495.30	7	trans-sialidase, putative	EAN85580	shed-acute-phase-antigen	CAA40511		Y	Y
19	Tc00.1047053507125.10	7	trans-sialidase, putative	EAN83508	surface glycoprotein GP90	AAM47176		Y	
20	Tc00.1047053511585.230	7	trans-sialidase, putative	EAN98531	surface glycoprotein GP90	AAM47176			
21	Tc00.1047053511911.10	7	trans-sialidase, putative	EAN94899	surface glycoprotein Tc85-11	AAD13347			
22	Tc00.1047053506003.39	7	trans-sialidase, putative	EAN82481	surface glycoprotein Tc85-11	AAD13347			
23	Tc00.1047053503447.20	6	trans-sialidase, putative	EAN88538	surface glycoprotein GP90	AAM47176			
24	Tc00.1047053510163.60	6	trans-sialidase, putative	XP_806966	c71 surface protein	AAQ74641		Y	
25	Tc00.1047053509549.10	6	trans-sialidase, putative	XP_815315	c71 surface protein	AAQ74641		Y	
26	Tc00.1047053507821.130	6	trans-sialidase, putative	EAN91828	surface glycoprotein Tc85-11	AAD13347		Y	
27	Tc00.1047053510111.20	6	trans-sialidase, putative	EAN85603	surface glycoprotein Tc85-11	AAD13347		Y	
28	Tc00.1047053510403.30	6	trans-sialidase, putative	EAN89081	85 kDa surface antigen precursor	Q03877		Y	
29	Tc00.1047053509281.20	6	trans-sialidase, putative	EAN83324	surface protein-2	AAO84044		Y	
30	Tc00.1047053506895.80	6	trans-sialidase, putative	EAN86623	trans-sialidase	AAA99443			
31	Tc00.1047053463279.20	5	trans-sialidase, putative	EAN82322	surface glycoprotein Tc-85/16	AAP32148			
32	Tc00.1047053506757.120	5	trans-sialidase, putative	EAN94707	trypomastigote ligand	AAS00641		Y	

33	Tc00.1047053401961.10	5	trans-sialidase, putative	EAN80840	trypomastigote ligand	AAS00641			
34	Tc00.1047053505609.40	5	trans-sialidase, putative	EAN86714	trypomastigote ligand	AAS00641		Y	
35	Tc00.1047053507357.100	4	trans-sialidase, putative	EAN90392	trypomastigote ligand	AAS00641		Y	
36	Tc00.1047053506129.30	4	trans-sialidase, putative	EAN91385	surface glycoprotein GP90	AAM47176			
37	Tc00.1047053506757.60	4	trans-sialidase, putative	XP_816556	c71 surface protein	AAQ74641		Y	
38	Tc00.1047053506751.50	4	trans-sialidase, putative	EAN99918	85 kDa surface antigen precursor	Q03877			
39	Tc00.1047053511757.100	3	trans-sialidase, putative	EAN89837	surface glycoprotein Tc-85/16	AAP32148			
40	Tc00.1047053508343.10	3	trans-sialidase, putative	EAN81325	shed-acute-phase-antigen	CAA40511			
41	Tc00.1047053507979.30	3	trans-sialidase, putative	XP_807499	shed-acute-phase-antigen	CAA40511		Y	Y
42	Tc00.1047053509629.10	3	trans-sialidase, putative	XP_805061	Flagellum-Associated Protein	CAA50287		Y	Y
43	Tc00.1047053503907.10	3	trans-sialidase, putative	EAN82985	complement regulatory protein, putative	EAN94610			Y
44	Tc00.1047053508297.7	3	trans-sialidase, putative	EAN87032	surface protein-2	AAO84046			
45	Tc00.1047053507747.180	3	trans-sialidase, putative	EAN96531	surface protein	BAA31361	Y	Y	Y
46	Tc00.1047053459061.10	2	trans-sialidase, putative	EAN81119	surface glycoprotein	AAA19649			Y
47	Tc00.1047053506613.40	2	trans-sialidase, putative	EAN94452	trypomastigote ligand	AAS00641		Y	
48	Tc00.1047053508161.20	2	trans-sialidase, putative	XP_809179	trypomastigote ligand	AAS00641			
49	Tc00.1047053509377.20	2	trans-sialidase, putative	XP_805751	Flagellum-Associated Protein	CAA50287		Y	
50	Tc00.1047053510543.70	2	trans-sialidase, putative	XP_816532	surface glycoprotein GP90	AAM47176			
51	Tc00.1047053507283.10	2	trans-sialidase, putative	EAN82970	c71 surface protein	AAQ74641			
52	Tc00.1047053507233.10	2	trans-sialidase, putative	EAN93383	complement regulatory protein	AAB49414		Y	
53	Tc00.1047053510847.10	2	trans-sialidase, putative	XP_812956	complement regulatory protein	XP_816461			Y
54	Tc00.1047053508977.120	2	trans-sialidase, putative	EAN90463	surface protein	BAA31361		Y	
55	Tc00.1047053507505.10	1	trans-sialidase, putative	EAN82718	Flagellum-Associated Protein	CAA50288		Y	
56	Tc00.1047053503705.19	1	trans-sialidase, putative	EAN83043	Flagellum-Associated Protein	CAA50287			
57	Tc00.1047053432995.9	1	trans-sialidase, putative	EAN80816	85 kDa surface antigen precursor	Q03877			
58	Tc00.1047053433733.10	1	trans-sialidase, putative	EAN80791	shed-acute-phase-antigen	CAA40511			
59	Tc00.1047053505363.10	1	trans-sialidase, putative	XP_804884	surface protein-2	AAO84045		Y	
60	Tc00.1047053506159.5	1	trans-sialidase, putative	EAN87108	surface antigen gp85 - T. cruzi	A45622			
61	Tc00.1047053505699.10	1	trans-sialidase, putative	XP_804506	surface antigen gp85 - T. cruzi	A45622	Y	Y	

Structural Proteins

62	Tc00.1047053506563.40	13	beta tubulin, putative	EAN94839					
63	Tc00.1047053411235.9	10	alpha tubulin	AAL75955					
64	Tc00.1047053510901.50	1	C-terminal motor kinesin, putative	EAN97167					
65	Tc00.1047053506485.70	1	hypothetical protein, conserved	EAN89623	myosin heavy chain [D. japonica]	BAA34955			
66	Tc00.1047053506295.30	1	hypothetical protein, conserved	EAN91002	similar to myosin [C. familiaris]	XP_861769			
67	Tc00.1047053506673.69	1	hypothetical protein, conserved	EAN88083	kinesin, putative [L. major]	CAJ04095			

Proteases										
68	Tc00.1047053507559.110	3	surface protease GP63	XP_811374				Y	Y	Y
69	Tc00.1047053511703.19	2	surface protease GP63	XP_811374						
70	Tc00.1047053506289.140	1	surface protease GP63	XP_818545				Y	Y	
71	Tc00.1047053508461.430	1	proteasome beta 2 subunit, putative	EAN99784						
Protein kinases/phosphatase										
72	Tc00.1047053506419.20	1	protein kinase, putative	EAN84637						
73	Tc00.1047053510421.180	1	casein kinase I, putative	EAN97312						
Flagellum associated proteins										
74	Tc00.1047053509617.20	5	paraflagellar rod protein 3, putative	EAN87979						
75	Tc00.1047053506749.20	1	flagellar calcium-binding protein	XP_804908						
Vacuole/exosome related proteins										
76	Tc00.1047053507713.30	3	heat shock protein 85, putative	EAN82629						
77	Tc00.1047053510119.20	1	elongation factor 1-alpha (EF-1-alpha), putative	EAN84978						
78	Tc00.1047053506943.50	2	glyceraldehyde 3-phosphate dehydrogenase	XP_812138						
79	Tc00.1047053509053.160	1	hypothetical protein, conserved	EAN91469	TPR repeat family protein [R. sphaeroides]	ZP_00918386				
80	Tc00.1047053506025.50	1	vacuolar ATP synthase subunit B	XP_808325						
Proteins with other functions										
81	Tc00.1047053487507.10	2	tryparedoxin peroxidase homologue	CAA09922						
82	Tc00.1047053511603.380	1	mucin-associated surface protein (MASP)	EAN98920				Y	Y	Y
83	Tc00.1047053508027.70	1	ATP-dependent RNA helicase, putative	EAN91446						
84	Tc00.1047053508277.230	1	nucleolar RNA binding protein, putative	EAN98432						
85	Tc00.1047053504163.30	1	40S ribosomal protein S9, putative	EAN99415						
86	Tc00.1047053508981.39	1	trifunctional enzyme alpha subunit	EAN84570	fatty oxidation complex, alpha subunit [B. bacteriovorus]	NP_968701				
87	Tc00.1047053509885.70	1	transporter, putative	EAN87055						
88	Tc00.1047053507089.260	1	CYC2-like cyclin, putative	EAN96940						
89	Tc00.1047053511635.10	1	histone H2b	CAA43297						
90	Tc00.1047053505989.110	1	hypothetical protein, conserved	EAN87824	amidinotransferase family protein [S. pomeroyi]	AAV95389				
91	Tc00.1047053503837.30	1	hypothetical protein, conserved	EAN84709	putative glutamic acid/alanine-rich protein [T. godfreyi]	CAD53379				
92	Tc00.1047053511467.50	1	hypothetical protein, conserved	EAN89560	Viral A-type inclusion protein repeat [E. histolytica]	XP_653447				

93	Tc00.1047053509047.60	1	hypothetical protein, conserved	EAN88089	Aardvark [D. discoideum]	AAG17931
94	Tc00.1047053506725.30	1	hypothetical protein	XP_810914	probable regulator of reproduction DopA [N. crassa]	CAD79673
<i>Proteins with unknown function</i>						
95	Tc00.1047053508601.110	1	hypothetical protein, conserved	EAN89958		
96	Tc00.1047053511495.20	1	hypothetical protein, conserved	EAN81903		
97	Tc00.1047053509441.10	1	hypothetical protein, conserved	EAN81666		
98	Tc00.1047053510901.190	1	hypothetical protein, conserved	EAN97181		
99	Tc00.1047053507053.80	1	hypothetical protein, conserved	EAN95843		
100	Tc00.1047053508897.50	1	hypothetical protein, conserved	EAN89037		
101	Tc00.1047053507093.320	1	hypothetical protein, conserved	EAN98708		
102	Tc00.1047053506459.280	1	hypothetical protein, conserved	EAN98547		
103	Tc00.1047053504021.90	1	hypothetical protein	EAN93765		
104	Tc00.1047053507851.60	1	hypothetical protein, conserved	EAN86970		
105	Tc00.1047053508073.20	1	hypothetical protein	EAN82384		
106	Tc00.1047053510265.80	1	hypothetical protein, conserved	EAN91976		
107	Tc00.1047053511217.90	1	hypothetical protein, conserved	EAN96478		
108	Tc00.1047053503751.10	1	hypothetical protein, conserved	EAN82027		
109	Tc00.1047053507809.80	1	hypothetical protein, conserved	EAN95570		
110	Tc00.1047053504109.100	1	hypothetical protein, conserved	EAN94988		

*NCBI – GenBank accession number.

Table 1.5 Gene ontology analysis of TcαGalVes proteins.

		GO		Name	Number of hits	Percentage
N	Level	category				
1	1	GO:0003674	molecular_function		87	79.1
2	2	GO:0003774	motor activity		3	2.7
3	3	GO:0003777	microtubule motor activity		2	1.8
4	2	GO:0003824	catalytic activity		75	68.2
5	3	GO:0004386	helicase activity		1	0.9
6	3	GO:0016491	oxidoreductase activity		2	1.8
7	3	GO:0016740	transferase activity		6	5.5
8	3	GO:0016787	hydrolase activity		67	60.9
9	3	GO:0016853	isomerase activity		2	1.8
10	2	GO:0004871	signal transducer activity		2	1.8
11	3	GO:0000155	two-component sensor molecule activity		2	1.8
12	2	GO:0005198	structural molecule activity		3	2.7
13	3	GO:0003735	structural constituent of ribosome		1	0.9
14	2	GO:0005215	transporter activity		3	2.7
15	3	GO:0005386	carrier activity		3	2.7
16	3	GO:0005489	electron transporter activity		2	1.8
17	3	GO:0015075	ion transporter activity		1	0.9
18	3	GO:0043492	ATPase activity, coupled to movement of substances		1	0.9
19	2	GO:0005488	binding		23	20.9
20	3	GO:0000166	nucleotide binding		14	12.7
21	3	GO:0003676	nucleic acid binding		3	2.7
22	3	GO:0005515	protein binding		7	6.4
23	3	GO:0043167	ion binding		7	6.4
24	3	GO:0046906	tetrapyrrole binding		2	1.8
25	3	GO:0048037	cofactor binding		1	0.9
26	2	GO:0016209	antioxidant activity		1	0.9
27	3	GO:0004601	peroxidase activity		1	0.9
28	2	GO:0045182	translation regulator activity		1	0.9
29	3	GO:0008135	translation factor activity, nucleic acid binding		1	0.9
30	1	GO:0005575	cellular_component		18	16.4

31	2	GO:0005623	cell	18	16.4
32	3	GO:0005622	intracellular	13	11.8
33	3	GO:0016020	membrane	5	4.5
34	3	GO:0042995	cell projection	2	1.8
35	2	GO:0043226	organelle	11	10.0
36	3	GO:0043227	membrane-bound organelle	5	4.5
37	3	GO:0043228	non-membrane-bound organelle	7	6.4
38	3	GO:0043229	intracellular organelle	11	10.0
39	2	GO:0043234	protein complex	9	8.2
40	3	GO:0000502	proteasome complex (sensu Eukaryota)	1	0.9
41	3	GO:0000786	nucleosome	1	0.9
42	3	GO:0005875	microtubule associated complex	3	2.7
43	3	GO:0016459	myosin	1	0.9
44	3	GO:0016469	proton-transporting two-sector ATPase complex	1	0.9
45	3	GO:0030529	ribonucleoprotein complex	1	0.9
46	1	GO:0008150	biological_process	82	74.5
47	2	GO:0007275	development	1	0.9
48	3	GO:0048066	pigmentation during development	1	0.9
49	2	GO:0007582	physiological process	78	70.9
50	3	GO:0008152	metabolism	75	68.2
51	3	GO:0040011	locomotion	1	0.9
52	3	GO:0050791	regulation of physiological process	1	0.9
53	3	GO:0050875	cellular physiological process	28	25.5
54	3	GO:0051179	localization	7	6.4
55	2	GO:0009987	cellular process	28	25.5
56	3	GO:0007154	cell communication	2	1.8
57	3	GO:0007155	cell adhesion	3	2.7
58	3	GO:0050794	regulation of cellular process	1	0.9
59	3	GO:0050875	cellular physiological process	28	25.5
60	2	GO:0043473	pigmentation	1	0.9
61	3	GO:0048066	pigmentation during development	1	0.9
62	2	GO:0050789	regulation of biological process	1	0.9
63	3	GO:0050791	regulation of physiological process	1	0.9

64	3	GO:0050794	regulation of cellular process	1	0.9
65	2	GO:0050896	response to stimulus	2	1.8
66	3	GO:0006950	response to stress	1	0.9
67	3	GO:0007610	behavior	1	0.9
68	3	GO:0009605	response to external stimulus	1	0.9
69	3	GO:0009628	response to abiotic stimulus	2	1.8
70	2	GO:0051704	interaction between organisms	59	53.6
71	3	GO:0044419	interspecies interaction between organisms	59	53.6

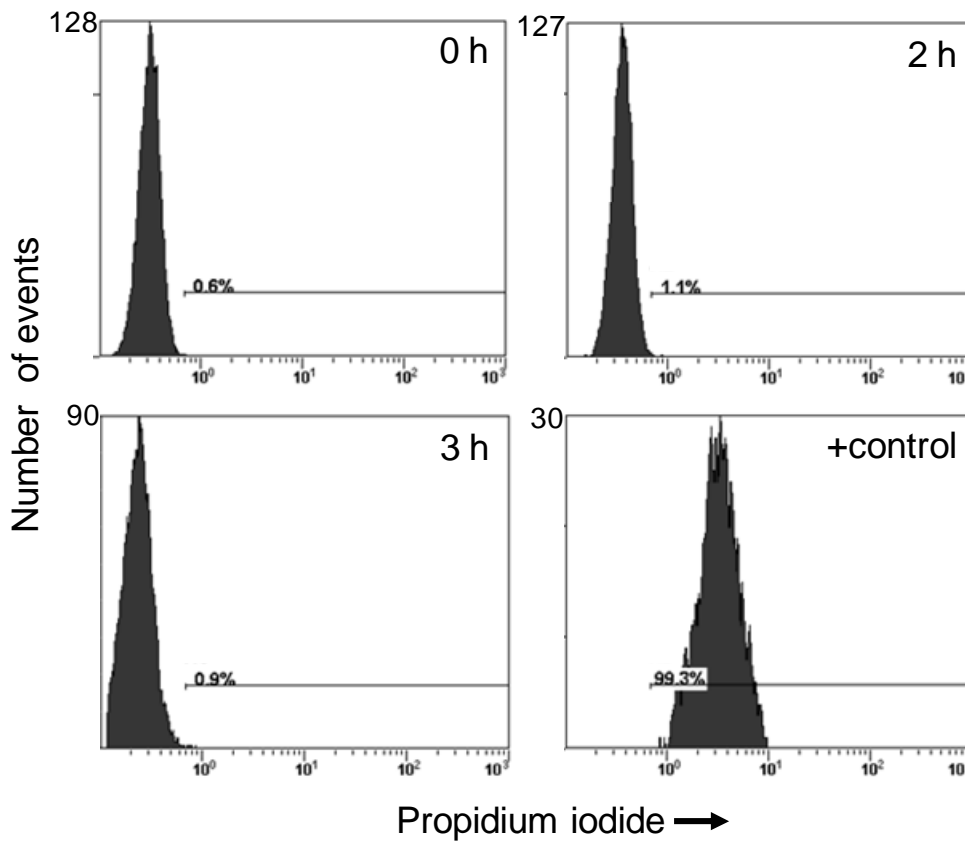


Figure 1.1 Flow cytometry analysis of *T. cruzi* trypomastigote viability following shedding procedure. Tissue culture cell-derived trypomastigotes were prepared as described in the section 1.3.1. Parasites (1×10^8 parasites/mL medium) were incubated for 0, 2, or 3 h at 37°C in the same conditions used to obtain the vesicles. Cell viability was assessed using the DeadDye reagent from the Cell Viability HitKit (Cellomics, Pittsburgh, PA). Briefly, parasites (1×10^7 cells per sample) were incubated with the DeadDye propidium iodide-containing solution for 15 min at 37°C following the manufacturer's recommendations. Parasites were then extensively washed with Cell Viability HitKit's washing buffer and fixed with 4% paraformaldehyde. Ten thousand cells were analyzed on a Cytomics FC 500 (Beckman Coulter, Miami, FL) flow cytometer. Parasites fixed with 4% paraformaldehyde prior to DeadDye staining were used as positive (+) control.

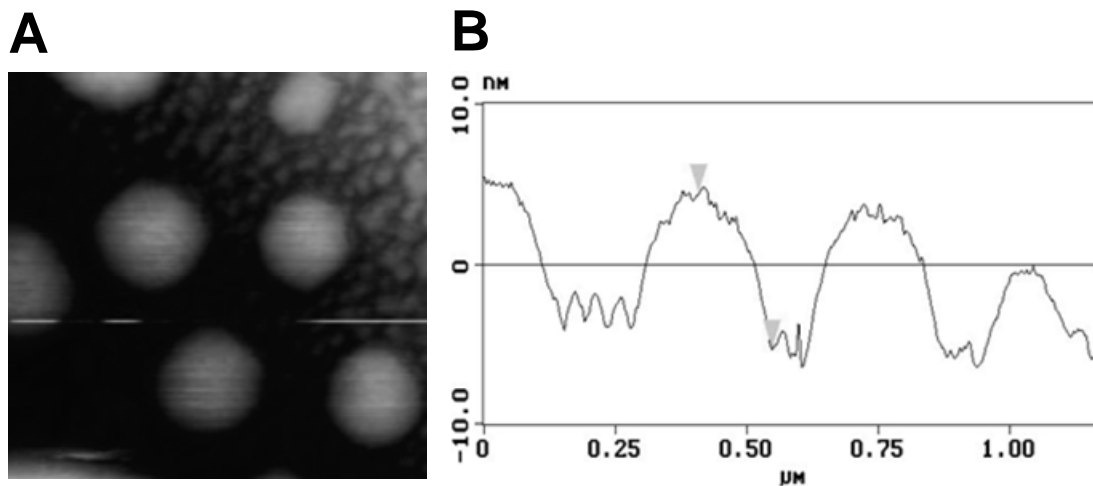


Figure 1.2 Atomic force microscopy analysis of *T. cruzi* trypomastigote vesicles obtained gel-filtration chromatography. (A) AFM scan top projection. (B) Vertical analysis of z- and y-axis.

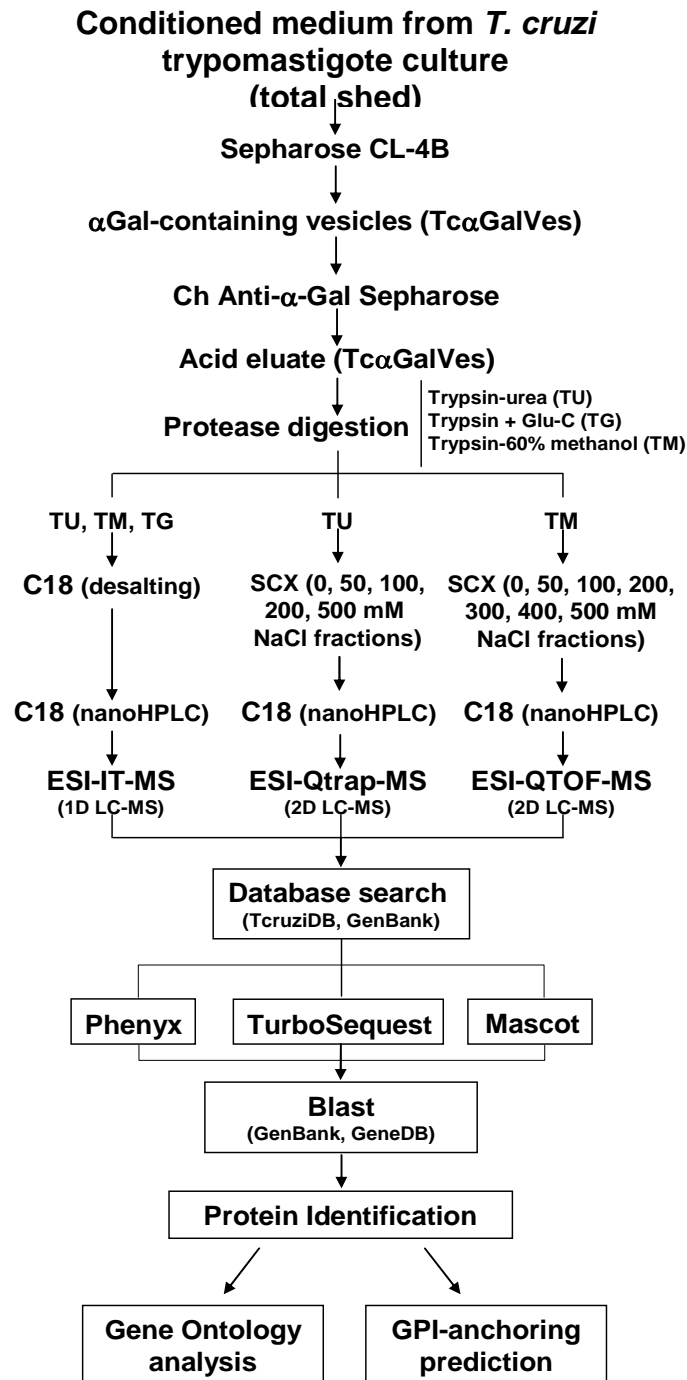


Figure 1.3 General strategy for TcαGalVes proteomic analysis. The vesicles were obtained from trypomastigote culture medium and purified by gel-filtration in Sepharose CL-4B and affinity chromatography with immobilized Ch anti-αGal. Ensuing peptides were desalted in a

C₁₈ zip-tip and analyzed by LC-MS using an ESI-IT-MS, or fractionated in a strong cation exchange (SCX) zip-tip, and analyzed by LC-MS using an ESI-Qtrap-MS or ESI-QTOF-MS. For ESI-Qtrap-MS analysis, peptides were eluted from SCX zip-tip with 0, 50, 100, 200, and 500 mM NaCl; and for ESI-QTOF-MS analysis, elution was performed with 0, 50, 100, 200, 300, 400, and 500 mM NaCl. All collected spectra were analyzed using Phenyx, TurboSequest, and Mascot algorithms. Identified sequences were subjected to Blast analysis and identified proteins were analyzed for gene ontology and GPI-anchoring prediction.

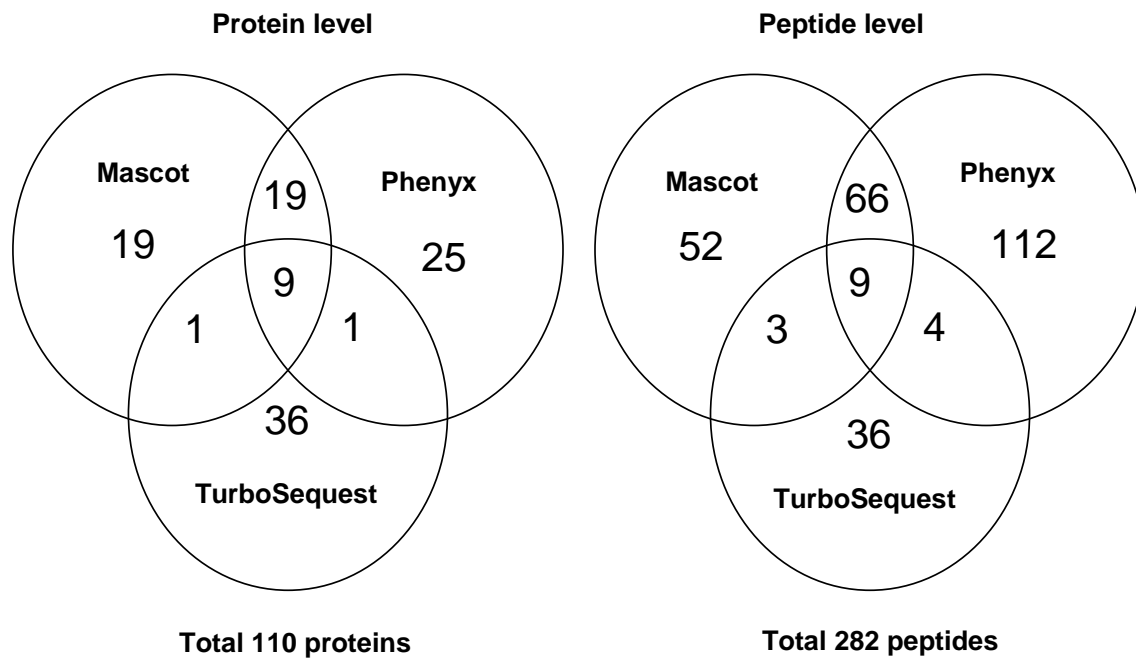


Figure 1.4 Comparative analysis of the number of *T. cruzi*-specific proteins and peptides identified using Mascot, Phenyx, and TurboSequest algorithms.

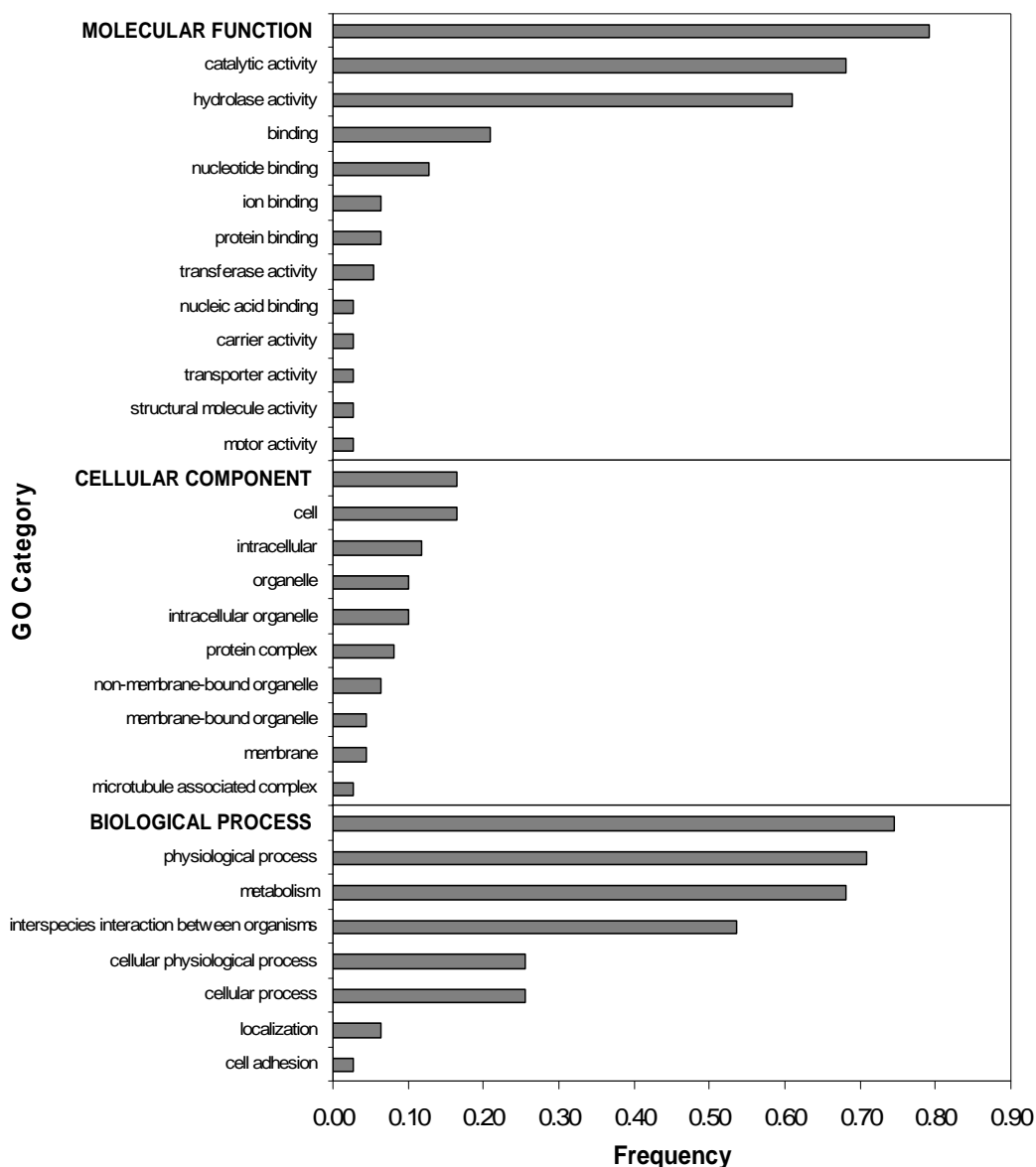


Figure 1.5 Gene ontology analysis. The graph represents the frequency of sequences that have a GO annotation for molecular function, biological process, and cellular component. Only the most representative (>3% of frequency) GO categories are shown in the graph. For the complete list, see Supplemental Table 4.

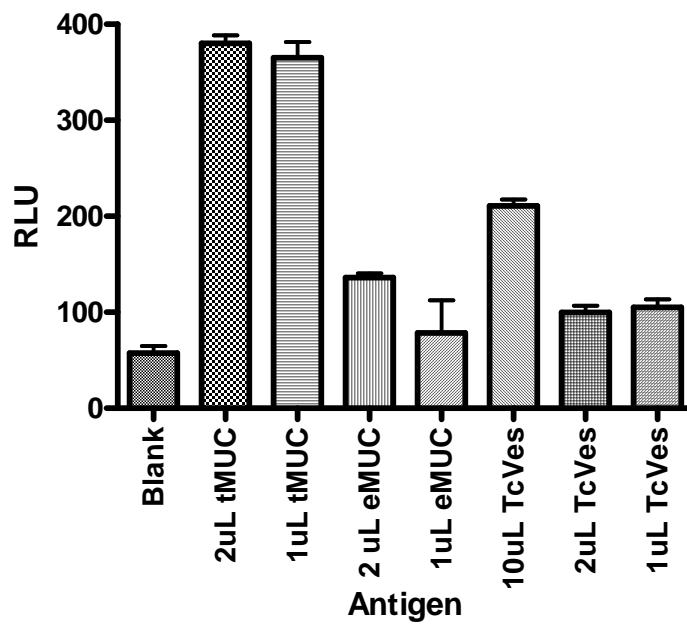


Figure 1.6 ELISA using anti-TcMUCII antibodies. Epimastigote (eMUC) and trypomastigote mucins (tMUC) were normalized by optical density at 214 nm (eMUC = 0.672 Abs, tMUC = 0.646 Abs). Tc α GalVes (TcVes) fraction was not normalized. Primary antibody anti-TcMUC II was used in 3 different 1/100 dilution and plate was incubated for 16h at 4°C. Secondary antibody anti-rabbit IgG conjugated with HRPO were diluted 1/2000 and incubated for 2.5h at 35°C. Reaction was developed with 1/5 diluted ECL reagent. RLU – relative luminescence unit.

Chapter 2: GPlomics: Global Analysis of Glycosylphosphatidylinositol-Anchored Molecules of *Trypanosoma cruzi*

Ernesto S. Nakayasu¹, Dmitry V. Yashunsky^{2,+}, Lilian L. Nohara¹, Ana Claudia T. Torrecilhas³, Andrei V. Nikolaev² and Igor C. Almeida^{1*}

¹Department of Biological Sciences, University of Texas at El Paso, El Paso, TX 79968, USA.

²Division of Biological Chemistry and Drug Discovery, College of Life Sciences, The Wellcome Trust Biocentre, University of Dundee, Dundee, DD1 5EH, UK. ³Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo SP 05508-900, Brazil.

⁺On leave from Research Institute of Biomedical Chemistry RAMS, Moscow 119121, Russia.

*Corresponding author. Department of Biological Sciences, University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968 USA. Telephone: +1 (915) 747-6086. Fax: +1 (915) 747-5808. E-mail: icalmeida@utep.edu

Running title: *Trypanosoma cruzi* GPlomics

Subject categories: Microbiology & Pathogens, Functional Genomics

Key words: Global analysis / GPlomics / Glycosylphosphatidylinositol / Glycobiology / Protein posttranslational modifications

2.1 Abstract

Glycosylphosphatidylinositol (GPI)-anchoring is a common, relevant posttranslational modification (PTM) of eukaryotic surface proteins. Here we developed a fast, simple, and highly sensitive method that employs liquid chromatography-tandem mass spectrometry (LC-MSⁿ) 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. By analyzing the GPIome of *T. cruzi* insect-dwelling epimastigote stage using LC-MSⁿ, we have identified 78 free and 11 protein-linked GPI species; of these, 72 and 8, respectively, were completely novel species. We have also identified the major GPI-anchored proteins expressed on the surface of epimastigotes as mucins from the *T. cruzi* small mucin-like gene (TcSMUG S) family. TcSMUG S mucins are highly O-glycosylated and, therefore, quite resistant to proteases. This make them very suitable to be expressed on the surface of epimastigote forms found in the protease-rich environment of the midgut of the insect-vector. We propose that our approach could be employed for the high throughput GPIomic analysis of other *T. cruzi* life cycle stages and pathogenic eukaryotes, as well as higher eukaryotes.

2.2 Introduction

Glycosylphosphatidylinositol (GPI)-anchoring is a ubiquitous post-translational 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 *et al*, 1993; Orlean and Menon, 2007; Paulick and Bertozzi, 2008). In lower eukaryotes, particularly yeast and protozoa, GPI-anchored molecules have also been shown to play important biological roles (Ferguson, 1999; McConville and Menon, 2000; Orlean *et al*, 2007). In pathogenic protozoan parasites (e.g., *Trypanosoma cruzi*, *Trypanosoma brucei*, *Leishmania major*, *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 (Acosta-Serrano *et al*, 2007; Buscaglia *et al*, 2006; Ferguson, 1999; Gazzinelli and Denkers, 2006; McConville *et al*, 1993; McConville *et al*, 2000).

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). In order 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, the major hurdle 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 GPIs 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 *et al*, 1993) .

Elortza and colleagues (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). While 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% 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 employs more polar and less denaturing buffers (e.g, acetate- and phosphate-based buffers) or organic solvents (e.g., *iso*- and *n*-propanol), and high initial salt concentration to enhance the relatively mild hydrophobic interaction between the hydrophobic moieties of the biomolecules and the chromatography medium. On the other hand, RPC medium is generally more hydrophobic than that of a HIC medium, leading to stronger interactions between the biomolecules and the medium, and

successful elution is achieved by the use of increasing concentration of non-polar, organic solvents (e.g., methanol, acetonitrile). The major drawback with this method is that most resins 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 more specific interaction between the GPI lipid moiety and the covalently attached functional group (e.g., C4) of the resin. This could considerably enhance the resolution of the process of purification of GPI-anchored molecules and free GPIs.

Trypanosoma 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 (Barrett *et al*, 2003; Dias *et al*, 2002). Lately, Chagas disease has become a public health threat for the U.S. 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 (Dumonteil, 2007; Garg and Bhatia, 2005). 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, popular 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 (Acosta-Serrano *et al*, 2001; Acosta-Serrano *et al*, 2007; Alves and Colli, 2008; Buscaglia *et al*, 2006; Frasch, 2000; Yoshida, 2006). 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 *et al*, 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 mass spectrometry, and employed for the analysis of GIPLs and proteinase-released GPIs. Our results show that the GPIome of *T. cruzi* is much more complex than was previously known, and provide new insights into the biosynthesis of *T. cruzi* GPI anchors, which could be eventually exploited as potential therapeutic target in Chagas disease.

2.3 Materials and Methods

2.3.1 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 Aug 1st, 2005), *T. cruzi* (downloaded on Mar 17th, 2008) and *T. brucei* TREU927 (downloaded on Aug 13th, 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 Aug 20th, 2005). Only sequences with high probability were accepted as potentially GPI-anchored proteins.

2.3.2 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 J.T. Baker.

2.3.3 *Trypanosoma cruzi* culture

Non-infective epimastigote forms of *T. cruzi* (Y strain) were grown in liver infusion tryptose (LIT) medium at 28°C for 3-4 days, as previously described (Camargo, 1964). Cells were harvested and washed twice in phosphate-buffered saline (PBS), pH 7.4, and centrifuged at 2,000g at 4°C for 10 min. The parasite pellet containing 3.6×10^{10} was then store at -20°C until use.

2.3.4 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 2,000g. 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 2,000g for 20 min, at room temperature, and saved. All organic phases were combined, dried under N₂ stream, and the resulting organic extract was subjected to Folch's partition (Folch *et al*, 1957). The upper (aqueous) phase of the partition, rich in free GPIs (GIPLs), was dried under N₂ stream and analyzed by LC-MS² as described below.

The resulting delipidated parasite pellet was dried under N₂ stream and extracted 3x 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 aqueous phase, enriched with GPI-anchored proteins, was dried in a vacuum centrifuge (Vacufuge, Eppendorf) and stored at -20°C until use.

2.3.5 Biotinylation of epimastigote mucins and chemiluminescent assay

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 chemiluminescent assay (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 3x with PBS containing 0.1% Tween-20 (PBS-T) and developed with 50 µl chemiluminescent reagent (SuperSignal, Pierce) diluted 5-fold in 200 mM carbonate/bicarbonate buffer, pH 9.6. Chemiluminescent readings were measured in a luminometer (Luminoskan Ascent, Labsystems, Finland).

2.3.6 Purification of synthetic GPIs

T. cruzi epimastigote and trypomastigote GPIs were synthesized as previously described (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.

2.3.7 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 sodium dodecyl sulfate-polyacrylamide gel electrophoresis (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-MS² (see below).

2.3.8 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-MS² using a Qtof-1 ESI-MS, or LC-MSⁿ using a LTQXL ESI-linear ion trap-MS.

2.3.9 Electrospray ionization-tandem mass spectrometry (ESI-MS²) 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.

For standardize the quantitative analysis, synthetic GPIs (Man-[EtNP]Man-Man₂-[AEP]GlcN-InsP-Gro-1-O-alkyl-C16:0-2-O-acyl-C16:0 (sGPI-C16:0), Man-[EtNP]Man-Man₂-[AEP]GlcN-InsP-Gro-1-O-alkyl-C16:0-2-O-acyl-C18:2 (sGPI-C18:2), Man-[EtNP]Man-Man₂-[AEP]GlcN-InsP-Gro-1-O-alkyl-C16:0-2-O-acyl-C18:1 (sGPI-C18:1)) were mixed in different ratios in 50% 2-propanol / 0.2% formic acid (FA) and analyzed in a linear ion-trap mass spectrometer (LTQ XL, Thermo Fisher Scientific). The samples were injected with an automated system (Triversa, Advion), set at 1.5 kV and 0.25 p.s.i. N₂ pressure, to avoid cross contamination between samples. Full (enhanced scan rate, 100 ms maximum injection time) or MS² (normal scan rate, 3 a.m.u. isolation width, 35% normalized collision energy) scans were collected for 20 sec. Each sample was run in triplicates.

2.3.10 Liquid chromatography-tandem mass spectrometry (LC-MSⁿ) analysis of GPI samples

Capillary columns (75- μ m internal diameter by 10-cm length) were packed with 10 mg/mL POROS R1 10 resin (10- μ m particle size; Applied Biosystems) (suspension in 100%

acetonitrile) with a high pressure apparatus at 500 psi (Gatlin *et al*, 1998). The quality of packing was monitored in a stereo microscope. 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. To compensate eventual decrease in the flow rate at high concentrations of organic solvent, the flow rate was set as a gradient from 350-450 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 msec. The five most intense ions were submitted to fragmentation with 40% normalized collision energy and a maximum injection time of 150 msec. When the dynamic exclusion function was enabled, it was set to collect only twice and then to exclude for 30 sec. To test the sensitivity of the method, the samples were also analyzed by the multiple-reaction monitoring (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/MS³. All the spectra were analyzed manually to assign the GPI structures.

2.3.11 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. The relative quantification of the GPI species was done by the MS² TIC (Asara *et al*, 2008).

2.3.12 Identification of proteins separated by SDS-PAGE

The protein bands from the SDS-PAGE were excised and digested as previously described (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 3µm, 15 cm x 75 µm, LC Packings, Dionex) connected to an Ultimate nanoHPLC (LC Packings, Dionex) in tandem to an ESI-QTOF-MS (Micromass Qtof-1, Waters). The elution was performed with a flow rate of 300 nL/min in a gradient of 5-35% acetonitrile (ACN)/0.1% formic acid (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 sec in the 50–2050 *m/z* range. MS data were converted into peak lists (PKL format) and searched against TcruidB v5.0 (Aguero *et al*, 2006) (<http://tcruidb.org>) using Mascot algorithm (Perkins *et al*, 1999) (Matrix Science). Database search parameters were the following: trypsin as the digesting enzyme (1 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).

2.3.14 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, 2h), 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 24h. The products were dried, redried upon 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 (30m x 0.25mm ID x 0.25μm). 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] x [amount of internal standard/(area standard *myo* peak/ area standard *scyllo* peak)].

2.4 Results

2.4.1 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 prior to the addition *en bloc* of the complete GPI anchor by the transamidase complex (Ferguson, 1999; McConville *et al*, 1993). 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), 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 2.1), 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 2 sequences (0.3%) were wrongly predicted to be GPI-anchored (Table 2.1), 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 2.1). 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 2.1). 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 ToIT (a surface protein with similarities with bacterial TolA) were also represented by multiple sequences (Table 2.1). From these results, we can conclude that a high proportion of the *T. cruzi* genes possibly encode GPI-anchored proteins.

2.4.2 Implementing a polystyrene-based reverse phase chromatography 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 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 2.1).

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*, 2000) (Materials and Methods, and Figure 2.2A). 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) (Materials and Methods, and Figure 2.2A). By CLA, the epimastigote mucins were shown to be eluted mainly with 22.5% aqueous *n*-propanol. Consistently with this result, the silver stained gel showed two broad bands with typical behavior of epimastigote mucins (Acosta-Serrano *et al*, 2001; Buscaglia *et al*, 2006) (Figure 2.2B, bands 9 and 10). These bands were excised, digested with trypsin, and sequenced by liquid chromatography coupled to tandem mass spectrometry (LC-MS²). We were unable to detect any peptide, probably because epimastigote mucins are highly glycosylated (Acosta-Serrano *et al*, 2007; Buscaglia *et al*, 2006), making them only partially susceptible to various proteases, including trypsin (data not shown). To further confirm 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 mode ESI-MS. We could detect at least 4 ion species with *m/z* identical to epimastigote mucin-derived GPIs (eMUC-GPIs) (Almeida *et al*, 2000; Previato *et al*, 1995; Serrano *et al*, 1995) that were further confirmed by fragmentation analysis (Figure 2.2C,D). Thus, we confirmed the excellent performance of POROS R1 in the purification of GPI-anchored proteins.

2.4.3 Analysis of protein-free GPIs by LC-MSⁿ

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-MS² and LC-MS³ analysis of organic (91% *n*-butanol and chloroform/methanol/water) extracts enriched for GPIs (Material and Methods, and Figure 2.3). 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 the amine (GlcN, AEP, and/or EtNP) groups of 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, used as the ion pair for the LC-MS analysis. With the LC-MS approach we could detect and characterize at least 78 species of GIPLs (Table 2 and 3). Of those, seventy-two were found to be novel, complete species (Table 2.3). In the MS² spectra, GIPL fragments corresponding to AEP or EtNP attached to multiple (n=1-5) hexose (Hex) residues were found (Figure 2.4). Fragments corresponding to the loss of Hex residue(s) attached to AEP or EtNP were also highly abundant (example in Figure 2.4). We could also find a fragment corresponding to the inositolphosphate (InsP) attached to AEP-GlcN (AEP-HexN-InsP). 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-Hex₆-HexN-InsP-C16:0/d18:0-Cer observed at *m/z* 1012.49 showed diagnostic fragment ions corresponding to the AEP-HexN attached to inositolphosphorylceramide (IPC) (AEP-HexN-IPC) (*m/z* 1050.45) and Hex₂-AEP (*m/z* 432.24) (Figure 2.4).

Still in the case of ceramide-containing GIPLs, the two most abundant ions observed by MS² analysis were the ceramide portion itself and the one resulting from the loss of ceramide (Figure 2.4). Thus, to obtain more detail from the ceramide moiety, we performed MS³ analysis of the two most abundant ions resulting from the loss of H₂O (at *m/z* 520.56 and 502.49). The ion species at *m/z* 520.56 gave rise to two 1-azirine C16:0-fatty acid derivatives at *m/z* 280.33 and 264.37 (Figure 2.4B). The 1-azirine fatty acid derivatives were also frequent in the MS² spectra, thus the structure of many ceramides could be determined without the need of MS³ analysis. Most of the ceramides were dihydroxylated, but some of them had three hydroxyl groups instead (Table 2.2 and 2.3, and Figure 2.4). The sphingoid base was invariably C18:0, C18:1, or C18:2, whereas the fatty acid varied in length from C16:0 to C26:0. Surprisingly, we

found a few ceramide species containing C23:0 and C25:0 (Table 2.2 and 2.3). To our best knowledge, odd- or branched-chain fatty acids have not yet been described in human pathogenic trypanosomatids (Lee *et al*, 2007). Nevertheless, other trypanosomatids such as *Phytomonas* and *Herpetomonas* have been reported to abundantly biosynthesize *de novo* a range of *iso*-branched, C18, C20 and C22, polyunsaturated fatty acids (Fish *et al*, 1982). 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 reverse phase chromatography, more likely these fatty acids are odd-chain or, most probably, monomethyl branched-chain species (Table 2.2 and 2.3). Further studies are necessary to confirm the existence of these unusual fatty acid species in *T. cruzi* and other pathogenic protozoan parasites.

A few GIPL species containing alkylacylglycerol (AAG) as the lipid tail were also found (Table 2.2 and 2.3, and Supplementary Figure 1C). Unlike the ceramide-containing GIPLs, MS² analysis of AAG-containing GIPL species revealed that fragments derived from the lipid tail were not always the most abundant ones (data not shown). Therefore, we could not carry out data-dependent MS³ analysis of all AAG-containing GIPL species. Trying to circumvent this issue, we performed data-independent MS³ analysis by selecting the most abundant and frequent lipid moiety-derived ion species observed at *m/z* 537.6 (data not shown). 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 2.2 and 2.3). However, since in *T. cruzi* GPIs all 1-O-alkyl chains thus far described are C16:0 (Almeida *et al*, 2000; Carreira *et al*, 1996; Previato *et al*, 2004; Serrano *et al*, 1995), 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 MS² total-ion chromatogram (MS² TIC) approach (Asara *et al*, 2008) (Material and Methods, and below). The quantification revealed that GIPL species containing ceramide (91.6-97.9%) are much more abundant than those having alkylacylglycerol (2.1-8.4%) as lipid tail (Table 2.2 and 2.3).

2.4.5 Analysis of proteinase K-released GPIs by LC-MSⁿ

Next, the fraction rich in GPI-anchored proteins was digested with proteinase K and analyzed by LC-MS² and LC-MS³ (Figure 2.2). 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 the GIPLs (Figure 2.5). In fact, some of the AAG-containing GPIs had the same glycan and lipid structure as some of the GIPLs (Table 2.2 and 2.3). 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 great majority (91.4-97.8%) of GIPLs had ceramide as the lipid tail (Table II and 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 MS³ analysis of AAG were the ones resulting from the loss of either alkyl or acyl group. Noteworthy, all AAG structures we could determine had

alkyl-C16:0 at the *sn*-1 position and variable- size fatty acyl group at the *sn*-2 position (Figure 2.5B).

In total, eleven species of GPI anchors were identified. Of those, only two had been previously described (Almeida *et al*, 2000; Previato *et al*, 1995; Serrano *et al*, 1995) and eight were found to be novel but much less abundant species (Figure 2.5, Table 2.3). The MS² TIC quantification showed that the two major species [EtNP]Hex₄-[AEP]HexN-InsP-Gro-alkyl-C16:0-acyl-C16:0; [AEP]Hex₄-[AEP]HexN-InsP-Gro-alkyl-C16:0-acyl-C16:0) comprise about 95% of total protein-derived GPIs, whereas the least abundant species accounts for only 0.04% ([EtNP]Hex₄-[AEP]HexN-InsP-Gro-alkylacyl-C42:1) (Table 2.3). The MS² and MS³ spectra, and schematic fragmentation for one of the novel species ([EtNP]Hex₄-[AEP]HexN-InsP-Gro-alkyl-C16:0-acyl-C24:0), representing ~2.5% of the total protein-derived GPI anchors, is shown in Figure 2.5.

2.4.6 Analysis of trypsin-released GPI-peptides by LC-MSⁿ

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, and adhesion and invasion of host cells (Acosta-Serrano *et al*, 2001; Acosta-Serrano *et al*, 2007; Buscaglia *et al*, 2006). *T. cruzi* mucins are encoded by 863 genes, grouped in two major families (i.e., TcMUC and TcSMUG) (Acosta-Serrano *et al*, 2007; Buscaglia *et al*, 2006). We had previously shown 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 non-infective epimastigotes and, therefore, to determine which gene family(ies) is(are) expressed on this parasite stage.

Although mucins are partially resistant to proteases, 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-MS² and LC-MS³ (Figure 2.2). Based on 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 2.6). 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-Hex₁₋₄ (Figure 2.6), similar to the fragmentation pattern described elsewhere (Redman *et al*, 1994). However, the fragmentation of the peptide moiety in the MS² 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 MS³ analysis (Figure 2.6). The quality of MS³ 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 2.6, Table 2.2 and 2.3). 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* (Acosta-Serrano *et al*, 2007; Buscaglia *et al*, 2006; El-Sayed *et al*, 2005). 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).

2.4.7 Quantitative analysis of GPIs

The absolute quantification was done by analyzing the *myo*-inositol content of the samples (Materials and Methods). Since the identified GPI structures are similar, we speculated that the ionization efficiency by electrospray of each species would be also very similar. Thus we would be able to relatively quantify the different species by signal intensities 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 material and methods for details). Although these GPI species have different lipid tails, their ionization efficiencies were almost identical (Figure 2.7). Next, we analyzed sample mixtures containing constant concentrations of sGPI-C18:1 (5 μ M), but varying the concentration of sGPI-C16:0 (0.005-5 μ M). The response in the MS was evaluated by comparing the signal intensities of the parent-ion or by MS² 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 in the range from 1:1 to 1:20 (Table 2.4). On the other hand, we found errors from 9.5 to 48.2% by MS² TIC approach, similar to the ones described by Asara *et al* (2008), but the linear-response range was shown to be much broader (1:1 to 1:500) (Table 2.4). Since most of the identified GPI and GIPL species are low abundant, we relatively quantified them by MS² 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-MS² analysis with different amounts of GPIs. First, we performed the analysis in 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 (Table 2.3). 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 2.8). 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 ~6,000- (for DDA) to ~30,000-fold (for MRM) more sensitive than current non-radioactive methods.

2.4.8 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 were very similar (Table 2.3), 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 to those with longer, non-hydroxylated, and/or saturated lipid tails (Table 2.3).

2.5 Discussion

Here we present the first global analysis of GPI-anchored molecules from a human pathogenic eukaryotic, *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 to 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 play key roles 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, since 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 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 (89 species) of GIPLs and protein-derived GPIs.

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

In a previous report, the glycan cores of GIPLs from Y strain epimastigote forms of *T. cruzi* were analyzed by nuclear magnetic resonance (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]Hex₅₋₆-[AEP]HexN-InsP, [EtNP]Hex₄₋₆-[AEP]HexN-InsP, and AEP-Hex₄₋₈-HexN-InsP) (Table 2.2). In addition, one minor species (comprising only 0.03% of the total GIPLs) without either EtNP or AEP (Hex₇-HexN-InsP) was observed (Table 2.2). Another minor species (0.01% of the total GIPLs) containing sialic acid (*N*-acetylneuraminic acid, NANA) (NANA-Hex₇-[AEP]HexN-InsP) was also detected and characterized by MS² analysis (Table 2.3). The presence of sialic acid-containing glycolipids in *T. cruzi* was proposed previously, through in vitro radiolabeling using ³H-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 previously known, indicating therefore a higher complexity of the GPI biosynthesis in *T. cruzi*.

The analysis of proteinase K-released GPIs led to the characterization of 11 different species. Of those, only two have been previously characterized (Almeida *et al*, 2000; Previato *et al*, 1995; Serrano *et al*, 1995). Most of the glycan structures had either two AEP residues ([AEP]Hex₄₋₅-[AEP]HexN-InsP) or one AEP and one EtNP ([EtNP]Hex₄₋₆-[AEP]HexN-InsP). One species, however, had only one EtNP ([EtNP]Hex₄-HexN-InsP). Interestingly, the glycan cores containing one AEP residue (AEP-Hex₄₋₈-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 (VSG) (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 employed. This is currently under investigation in our laboratory.

Taken together, our results and others from the literature (Acosta-Serrano *et al*, 2001; Acosta-Serrano *et al*, 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 previous 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 (Almeida *et al*, 2000; Carreira *et al*, 1996) and eMUC-GPI (Almeida *et al*, 2000; Previato *et al*, 1995; Serrano *et al*, 1995) species previously characterized, and there was 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 employed 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 *et al*, 2008).

2.6 Acknowledgements

We thank Mr. Tiago Sobreira (University of Sao Paulo, Brazil) for assistance in the prediction analysis. This work was funded by the grants 1R01AI070655 and 5G12RR008124 from the National Institutes of Health, and the Wellcome Trust (067089/Z/02/Z). ICA was also supported by grant 2S06GM00812-37 from the NIH/MBRS/SCORE Program. E.S.N. was supported by the George A. Krutilek memorial graduate scholarship from Graduate School, UTEP. 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.

2.7 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**: D428-431.
- 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.

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

Bern C, Montgomery SP, Herwaldt BL, Rassi A, Jr., 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 glycoposphatidylinositol 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, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B (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 glycosyl-phosphatidylinositol 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 (Pt 17)**: 2799-2809.

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

Fish WR, Holz GG, Jr., 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 JR, 3rd (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.

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 U S A* **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.

Lee SH, Stephens JL, Englund PT (2007) A fatty-acid synthesis mechanism specialized for parasitism. *Nat Rev Microbiol* **5**: 287-297.

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** (Pt 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.

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 glycopospholipids. *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*.

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 glycoprotein 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.

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 U S A* **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.

Table 2.1 Genome-wise prediction of GPI-anchoring sequences

Database	Number of sequences	Highly probable sequences	% from total
TcruziDB v5.0	19613	2304	11.9
GenBank <i>T. cruzi</i> sequences	41478	4941	11.7
GeneDB TcMUC II sequences	624	598	95.8
GeneDB TcMUC II -20 AA sequences*	624	2	0.3

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

Protein family	Number of sequences	% from total GPI-anchored proteins
<i>MASP</i>	1730	35.0
<i>Mucins</i>	1262	25.5
TcMUC I	1126	22.8
TcMUC II	100	2.0
TcMUC III	2	0.0
TcSMUG S	20	0.4
TcSMUG L	14	0.3
<i>TS/gp85</i>	1160	23.5
<i>trans</i> -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
<i>Other families</i>	789	16.0
Amastin	216	4.4
gp63	143	2.9
mucin-like	86	1.7
ToIT	14	0.3
hypothetical proteins	207	4.2
others	123	2.5

*These sequences had 20 of the C-terminal amino acid residues removed.

Table 2.2 Summary of glycan, lipid, and peptide composition of the GPIome of *T. cruzi* epimastigotes characterized by LC-MS/MS

	GPI-Anchored molecule (origin or treatment) *			
	GIPLs (CM and CMW)	GIPLs (BuOH)	Proteins (Proteinase K)	Proteins (Trypsin)
Overall data				
Inositol content (picomoles per 1e7 cells)	889.95±49.02 ¶	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 (% from total)				
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

* CM and CMW, CHCl₃:MeOH (1:1, v/v) and CHCl₃:MeOH:water (1:2:0.8, v/v/v), respectively; BuOH, 1-butanol.

¶ The amount of GIPL-derived *myo*-inositol is overestimated because of the presence of large quantities of free phosphatidylinositol (PI) in this fraction.

Table 2.3 Identification and quantification of GIPLs and protein-derived GPI anchors

GIPLs (from both extractions)				n-butanol extraction				C:M and C:M:W extractions				Proposed structure
Species	Calculated m/z	Fragmented m/z	Ion Species	RT (min)	# MS2 spectra	MS2 TIC	% from total MS2 TIC	RT (min)	# MS2 spectra	MS2 TIC	% from total MS2 TIC	
1	1000.45	1001.26	M+2H	-	0	0	0.00	20.92	3	3480	0.14	[EtNP]Hex5-[AEP]HexN-InsP-C34:0(OH)3-Cer
2	1019.98	1020.99	M+2H	-	0	0	0.00	ND*	4	3875	0.16	AEP-Hex6-HexN-InsP-C34:0(OH)3-Cer
3	991.45	992.16	M+2H	29.28	7	4567	3.40	21.38	87	207018	8.38	[EtNP]Hex5-[AEP]HexN-InsP-C16:0/d18:1-Cer
4	1173.03	1174.42	M+2H	-	0	0	0.00	21.48	7	9116	0.37	AEP-Hex8-HexN-InsP-C16:0/d18:1-Cer
5	983.45	983.90	M+2H	29.28	9	9390	6.98	21.41	45	146052	5.91	[AEP]Hex5-[AEP]HexN-InsP-C16:0/d18:1-Cer
6	1010.97	1011.15	M+2H	29.28	5	1223	0.91	21.48	120	271699	11.00	AEP-Hex6-HexN-InsP-C16:0/d18:1-Cer
7	1092.00	1092.97	M+2H	-	0	0	0.00	21.35	6	3597	0.15	AEP-Hex7-HexN-InsP-C34:1(OH)2-Cer
8	1064.48	1065.32	M+2H	-	0	0	0.00	ND	1	320	0.01	[AEP]Hex6-[AEP]HexN-InsP-C34:1(OH)2-Cer
9	848.92	849.13	M+2H	-	0	0	0.00	21.48	3	3129	0.13	AEP-Hex4-HexN-InsP-C16:0/d18:1-Cer
10	1185.03	1185.49	M+H+Na	-	0	0	0.00	21.88	5	4017	0.16	AEP-Hex8-HexN-InsP-C16:0/d18:1-Cer
11	992.46	992.36	M+2H	32.92	7	1483	1.10	21.85	61	166227	6.73	[EtNP]Hex5-[AEP]HexN-InsP-C16:0/d18:0-Cer
12	1093.00	1093.23	M+2H	-	0	0	0.00	21.95	8	5799	0.23	AEP-Hex7-HexN-InsP-C34:0(OH)2-Cer
13	1248.55	1250.00	M+H+Na	-	0	0	0.00	ND	1	229	0.01	NANA-Hex7-[AEP]HexN-InsP-C34:0(OH)2-Cer
14	1073.49	1073.87	M+2H	-	0	0	0.00	ND	6	2158	0.09	[EtNP]-Hex6-[AEP]HexN-InsP-C34:0(OH)2-Cer
15	1011.98	1012.49	M+2H	30.02	2	628	0.47	21.94	79	262191	10.61	AEP-Hex6-HexN-InsP-C34:0(OH)2-Cer
16	984.46	984.78	M+2H	29.75	7	4340	3.23	21.88	44	97905	3.96	[AEP]Hex5-[AEP]HexN-InsP-C16:0/d18:0-Cer
17	849.93	849.93	M+2H	-	0	0	0.00	21.91	3	3435	0.14	AEP-Hex4-HexN-InsP-C34:0(OH)2-Cer
18	1019.48	1019.76	M+2H	32.88	2	153	0.11	25.33	93	105670	4.28	AEP-Hex6-HexN-InsP-Gro-alkyl-C16:0-acyl-C16:0
19	999.96	1000.00	M+2H	32.95	5	1091	0.81	25.09	42	40140	1.63	[EtNP]Hex5-[AEP]HexN-InsP-Gro-alkylacyl-C32:0
20	1181.53	1181.73	M+2H	-	0	0	0.00	24.89	5	6264	0.25	AEP-Hex8-HexN-InsP-Gro-alkylacyl-C32:0
21	991.96	992.57	M+2H	32.92	7	1483	1.10	25.30	6	15330	0.62	[AEP]Hex5-[AEP]HexN-InsP-Gro-alkylacyl-C32:0
22	918.93	919.79	M+2H	ND	1	18.1	0.01	ND	5	3061	0.12	[EtNP]Hex4-[AEP]HexN-InsP-Gro-alkylacyl-C32:0
23	857.43	857.36	M+2H	34.40	1	109	0.08	25.90	15	14368	0.58	AEP-Hex4-HexN-InsP-Gro-alkylacyl-C32:0
24	869.43	870.70	M+2H	34.40	1	109	0.08	26.12	7	5479	0.22	AEP-Hex4-HexN-InsP-Gro-alkylacyl-C34:2
25	1032.49	1032.75	M+2H	-	0	0	0.00	ND	9	9695	0.39	AEP-Hex6-HexN-InsP-Gro-alkylacyl-C34:1
26	881.42	882.76	M+H+Na	-	0	0	0.00	ND	2	1196	0.05	AEP-Hex4-HexN-InsP-Gro-alkylacyl-C34:1
27	1032.99	1033.21	M+2H	-	0	0	0.00	ND	7	3639	0.15	AEP-Hex6-HexN-InsP-Gro-alkylacyl-C34:0
28	1052.51	1054.97	M+2H	-	0	0	0.00	ND	1	289	0.01	AEP-Hex6-HexN-InsP-C22:0/d18:1-Cer
29	1033.49	1034.34	M+2H	36.97	1	112	0.08	-	0	0	0.00	[EtNP]Hex5-[AEP]HexN-InsP-C22:0/d18:1-Cer
30	1034.51	1035.83	M+2H	37.13	3	221	0.16	-	0	0	0.00	[EtNP]Hex5-[AEP]HexN-InsP-C40:0-(OH)2-Cer

31	1040.51	1041.51	M+2H	38.32	2	101	0.08	30.39	3	1602	0.06	[EtNP]Hex5-[AEP]HexN-InsP-C23:0/d18:1-Cer
32	1026.51	1027.01	M+2H	36.93	3	646	0.48	-	0	0	0.00	[AEP]Hex5-[AEP]HexN-InsP-C40:0(OH)2-Cer
33	1060.03	1060.80	M+2H	-	0	0	0.00	-	5	2781	0.11	AEP-Hex6-HexN-InsP-C23:0/d18:1-Cer
34	1066.03	1066.83	M+2H	-	0	0	0.00	30.29	2	1719	0.07	AEP-Hex6-HexN-InsP-C24:0/d18:2-Cer
35	1047.21	1047.23	M+2H	-	0	0	0.00	30.11	2	617	0.02	[EtNP]-Hex5-[AEP]HexN-InsP-C42:2(OH)2-Cer
36	1032.51	1034.13	M+2H	37.51	5	913	0.68	ND	2	847	0.03	[AEP]Hex5-[AEP]HexN-InsP-C23:0/d18:1-Cer
37	1048.52	1048.92	M+2H	38.17	1	192	0.14	-	0	0	0.00	[AEP]Hex5-[AEP]HexN-InsP-C24:0/t18:0-Cer
38	1047.51	1047.66	M+2H	38.09	17	26098.7	19.41	30.53	42	136957	5.54	[EtNP]Hex5-[AEP]HexN-InsP-C24:0/d18:1-Cer
39	1076.04	1076.29	M+2H	-	0	0	0.00	ND	7	5774	0.23	AEP-Hex6-HexN-InsP-C22:0/t18:1-Cer
40	1061.03	1062.70	M+2H	-	0	0	0.00	ND	2	1420	0.06	AEP-Hex6-HexN-InsP-C23:0/d18:0-Cer
41	1057.02	1057.37	M+2H	-	0	0	0.00	ND	2	1164	0.05	[EtNP]Hex5-[AEP]HexN-InsP-C24:0/t18:0-Cer
42	1038.51	1039.64	M+2H	-	0	0	0.00	ND	1	660	0.03	[AEP]Hex5-[AEP]HexN-InsP-C42:2(OH)2-Cer
43	1039.51	1040.31	M+2H	38.01	17	23797	17.69	30.88	37	72953	2.95	[AEP]Hex5-[AEP]HexN-InsP-C24:0/d18:1-Cer
44	1229.09	1229.40	M+2H	37.7	6	1876	1.39	30.32	5	5921	0.24	AEP-Hex8-HexN-InsP-C24:0/d18:1-Cer
45	1056.52	1056.86	M+2H	37.28	6	1113	0.83	-	0	0	0.00	[EtNP]Hex5-[AEP]HexN-InsP-C24:0/t18:1-Cer
46	1066.53	1066.61	M+2H	38.28	12	12782	9.50	30.29	109	295565	11.97	AEP-Hex6-HexN-InsP-C24:0/d18:1-Cer
47	1033.51	1033.87	M+2H	-	0	0	0.00	ND	3	1063	0.04	[AEP]Hex5-[AEP]HexN-InsP-C23:0/d18:0-Cer
48	1148.06	1148.52	M+2H	38.01	2	213	0.16	30.60	10	7824	0.32	AEP-Hex7-HexN-InsP-C24:0/d18:1-Cer
49	1120.55	1121.55	M+2H	37.90	4	690	0.51	-	2	762	0.03	[AEP]Hex6-[AEP]HexN-InsP-C24:0/d18:1-Cer
50	1067.54	1067.9	M+2H	38.09	6	4124	3.07	30.88	85	280398	11.35	AEP-Hex6-HexN-InsP-C24:0/d18:0-Cer
51	1149.07	1149.03	M+2H	ND	2	579	0.43	ND	4	2330	0.09	AEP-Hex7-HexN-InsP-C24:0/d18:0-Cer
52	1094.55	1094.89	M+2H	-	0	0	0.00	ND	1	731	0.03	Hex7-HexN-InsP-C24:0/d18:0-Cer
53	1040.53	1041.36	M+2H	37.98	13	19197	14.27	30.39	37	92934	3.76	[AEP]Hex5-[AEP]HexN-InsP-C24:0/d18:0-Cer
54	1230.09	1230.37	M+2H	ND	2	111	0.08	30.42	2	2271	0.09	AEP-Hex8-HexN-InsP-C24:0/d18:0-Cer
55	1121.55	1122.13	M+2H	-	0	0	0.00	ND	3	808	0.03	[AEP]Hex6-[AEP]HexN-InsP-C24:0/d18:0-Cer
56	1048.52	1048.74	M+2H	38.24	9	11097.1	8.25	31.30	46	90559	3.67	[EtNP]Hex5-[AEP]HexN-InsP-C24:0/d18:0-Cer
57	986.01	987.47	M+2H	-	0	0	0.00	ND	1	365	0.01	AEP-Hex5-HexN-InsP-C24:0/d18:1-Cer
58	1054.52	1055.87	M+2H	39.14	4	698	0.52	ND	4	2908	0.12	[EtNP]Hex5-[AEP]HexN-InsP-C25:0/d18:1-Cer
59	904.98	905.23	M+2H	-	0	0	0.00	31.44	2	573	0.02	AEP-Hex4-HexN-InsP-C24:0/d18:1-Cer
60	1073.54	1074.46	M+2H	39.33	2	201	0.15	32.11	6	7961	0.32	AEP-Hex6-HexN-InsP-C25:0/d18:1-Cer
61	905.99	906.77	M+2H	-	0	0	0.00	ND	6	4117	0.17	AEP-Hex4-HexN-InsP-C42:0(OH)2-Cer
62	1248.10	1249.94	M+H+Na	-	0	0	0.00	ND	1	202	0.01	AEP-Hex8-HexN-InsP-C25:0/d18:0-Cer
63	1075.05	1076.70	M+2H	-	0	0	0.00	ND	10	6413	0.26	AEP-Hex6-HexN-InsP-C25:0/d18:0-Cer
64	1156.07	1157.06	M+2H	-	0	0	0.00	ND	1	352	0.01	AEP-Hex7-HexN-InsP-C25:0/d18:0-Cer
65	987.01	987.69	M+2H	-	0	0	0.00	ND	1	313	0.01	AEP-Hex5-HexN-InsP-C24:0/d18:0-Cer

66	1046.53	1047.61	M+2H	39.14	3	868	0.65	ND	2	2079	0.08	[AEP]Hex6-[AEP]HexN-InsP-C25:0/d18:1-Cer
67	1055.53	1055.90	M+2H	ND	1	157	0.12	ND	3	2088	0.08	[EtNP]Hex5-[AEP]HexN-InsP-C25:0/d18:0-Cer
68	1047.53	1047.73	M+2H	ND	2	364	0.27	ND	1	1290	0.05	[AEP]Hex5-[AEP]HexN-InsP-C25:0/d18:0-Cer
69	1080.55	1081.91	M+2H	ND	1	93.1	0.07	33.00	8	11360	0.46	AEP-Hex6-HexN-InsP-C26:0/d18:1-Cer
70	1053.53	1053.58	M+2H	40.45	4	925	0.69	ND	4	4118	0.17	[AEP]Hex5-[AEP]HexN-InsP-C26:0/d18:1-Cer
71	1061.53	1062.89	M+2H	40.41	3	626	0.47	33.04	4	3952	0.16	[EtNP]Hex5-[AEP]HexN-InsP-C26:0/d18:1-Cer
72	1081.55	1082.97	M+2H	ND	1	166	0.12	33.11	11	11650	0.47	AEP-Hex6-HexN-InsP-C26:0/d18:0-Cer
73	1062.53	1063.29	M+2H	40.53	1	123	0.09	ND	2	2195	0.09	[EtNP]Hex5-[AEP]HexN-InsP-C26:0/d18:0-Cer
74	1054.54	1054.78	M+2H	ND	4	946	0.70	33.56	3	1927	0.08	[AEP]Hex5-[AEP]HexN-InsP-C26:0/d18:0-Cer
75	1075.55	1075.94	M+2H	-	0	0	0.00	ND	5	3169	0.13	AEP-Hex6-HexN-InsP-Gro-alkylacyl-C40:0
76	1128.53	1128.76	M+2H	37.82	2	222	0.17	-	0	0	0.00	[EtnP]Hex6-[AEP]HexN-InsP-C24:0/d18:1-Cer
77	1075.53	1077.7	M+2H	38.05	2	421	0.31	-	0	0	0.00	AEP-Hex6-HexN-InsP-C24:0/t18:1-Cer
78	1046.51	1046.6	M+2H	ND	1	220	0.16	-	0	0	0.00	[EtnP]Hex5-[AEP]HexN-InsP-C24:0/d18:2-Cer
Total					196	134487	100.00		1162	2470070	100.00	
Protein-derived GPIs (from proteinase K digestion)												
Species	Calculated m/z	Fragmented m/z	Ion Species	RT (min)	# MS2 spectra	MS2 TIC	% from total MS2 TIC	-	-	-	-	
1	918.93	919.23	M+2H	31.27	23	47113.3	80.50	-	-	-	-	[EtNP]Hex4-[AEP]HexN-InsP-Gro-alkyl-C16:0-acyl-C16:0
2	910.93	911.42	M+2H	31.12	9	8316	14.21	-	-	-	-	[AEP]Hex4-[AEP]HexN-InsP-Gro-alkyl-C16:0-acyl-C16:0
3	876.92	877.78	M+2H	32.65	1	112	0.19	-	-	-	-	[EtNP]Hex4-HexN-InsP-Gro-alkylacyl-C34:2
4	974.99	975.18	M+2H	39.98	6	1460	2.49	-	-	-	-	[EtNP]Hex4-[AEP]HexN-InsP-Gro-alkyl-C16:0-acyl-C24:0
5	999.96	1000.66	M+2H	31.08	5	794	1.36	-	-	-	-	[EtNP]Hex5-[AEP]HexN-InsP-Gro-alkylacyl-C32:0
6	1080.99	1081.82	M+2H	31.05	3	104	0.18	-	-	-	-	[EtNP]Hex6-[AEP]HexN-InsP-Gro-alkylacyl-C32:0
7	932.95	933.17	M+2H	33.11	3	410	0.70	-	-	-	-	[EtNP]Hex4-[AEP]HexN-InsP-Gro-alkylacyl-C34:0
8	930.93	931.8	M+2H	ND	1	36.9	0.06	-	-	-	-	[EtNP]Hex4-[AEP]HexN-InsP-Gro-alkylacyl-C34:2
9	924.95	925.15	M+2H	33.07	1	57.1	0.10	-	-	-	-	[AEP]Hex4-[AEP]HexN-InsP-Gro-alkylacyl-C34:0
10	988.00	988.38	M+2H	40.22	1	20.9	0.04	-	-	-	-	[EtNP]Hex4-[AEP]HexN-InsP-Gro-alkylacyl-C42:1
11	925.94	926.31	M+2H	32.07	1	105	0.18	-	-	-	-	[EtNP]Hex4-[AEP]HexN-InsP-Gro-alkylacyl-C33:0
Total					54	58529.2	100.00	-	-	-	-	
Protein-derived GPI-peptides (from trypsin digestion)												
Species	Calculated m/z	Fragmented m/z	Ion Species	RT (min)	# MS2 spectra	MS2 TIC	% from total MS2 TIC					
1	1186.55	1188.00	M+2H	ND	9	2882.6	56.50	-	-	-	-	APTPGD[EtNP]Hex4-[AEP]HexN-InsP-Gro-Alkylacyl-C32:0
2	1269.07	1270.00	M+2H	ND	6	1017	19.93	-	-	-	-	APTPGD[EtNP]Hex5-[AEP]HexN-InsP-Gro-Alkylacyl-C32:0

3	1146.04	1146.71	M+2H	ND	4	424	8.31	-	-	-	-	APTPGD[EtNP]Hex4-HexN-InsP-Gro-Alkylacyl-C34:2
4	1180.05	1180.50	M+2H	ND	4	583	11.43	-	-	-	-	APTPGD[AEP]Hex4-[AEP]HexN-InsP-Gro-Alkylacyl-C32:0
5	1261.57	1262.00	M+2H	ND	2	195	3.82	-	-	-	-	APTPGD[AEP]Hex5-[AEP]HexN-InsP-Gro-Alkylacyl-C32:0
Total					25	5101.6	100.00	-	-	-	-	

*ND – not determined.

¶ The lines highlighted in blue represent previously described structures.

Table 2.4 Relative quantification of synthetic GPI standards

	MS1			MS2		
Injected ratio sGPI-16:0/s-GPI-18:1	Ratio Average	Standard deviation	Error (%)	Ratio Average	Standard deviation	Error (%)
1	0.9871	0.0637	1.3	0.6968	0.1367	30.3
0.5	0.5408	0.0284	8.2	0.3326	0.1332	33.5
0.2	0.2209	0.0163	10.5	0.1684	0.0193	15.8
0.1	0.1214	0.0041	21.4	0.0905	0.0127	9.5
0.05	0.0668	0.0061	33.6	0.0264	0.0083	47.1
0.02	0.0440	0.0017	119.8	0.0106	0.0026	46.9
0.01	0.0381	0.0030	281.0	0.0052	0.0009	48.2
0.002	0.0285	0.0086	1323.2	0.0028	0.0020	40.0
0.001	0.0352	0.0225	3416.0	0.0021	0.0009	109.6

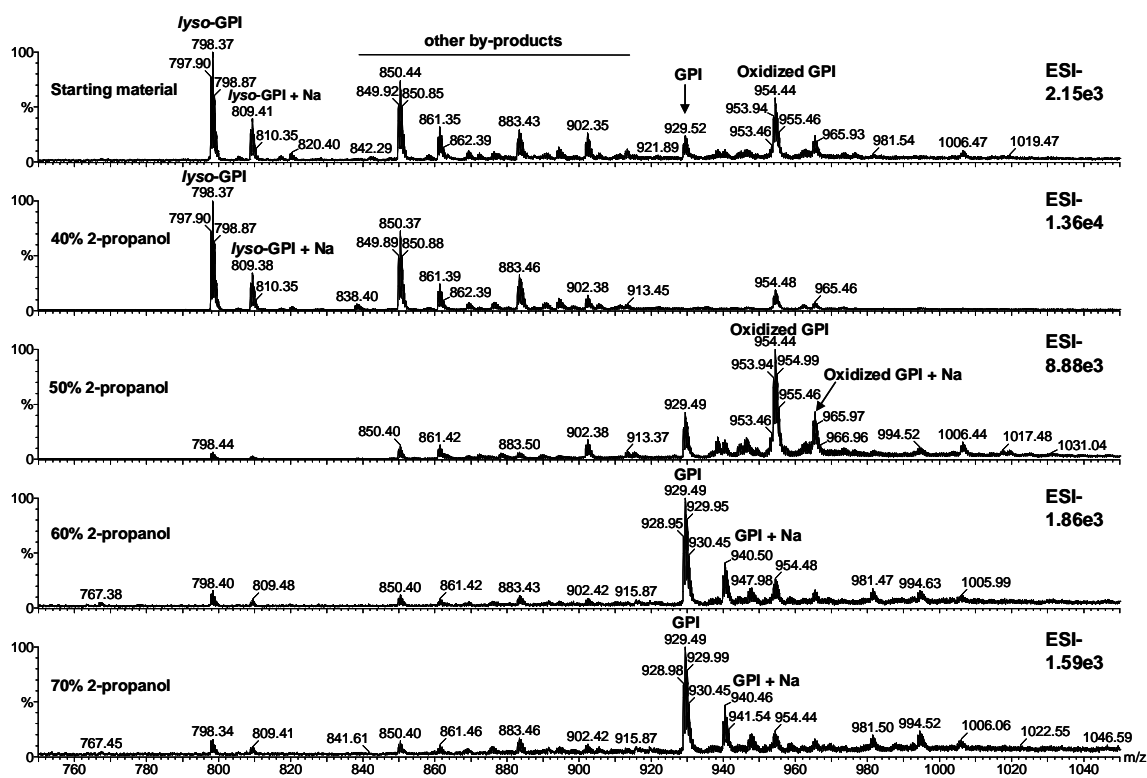


Figure 2.1 Fractionation of the synthetic trypanomastigote mucin GPI and by-products using POROS R1 ziptip. The synthetic mucin GPI (Man-[EtNP]Man-Man₂-[AEP]GlcN-InsP-Gro-1-O-alkyl-C16:0-2-O-acyl-C18:2) was purified through POROS R1 50 resin as described in Material and Methods. The top panel represents 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.

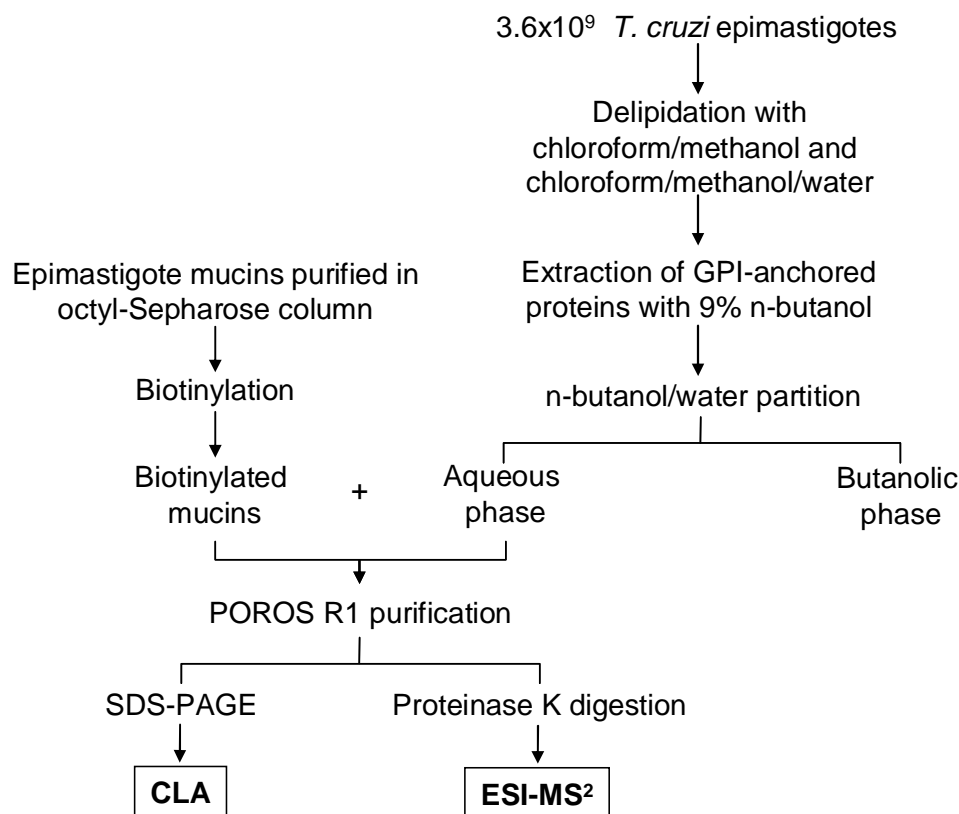


Figure 2.2 (A) Purification of GPI-anchored proteins from *T. cruzi* epimastigotes using POROS R1 50 cartridge. Schematic representation of the methodology used for the purification and analysis of GPI-anchored proteins. Briefly, epimastigote cells were delipidated with chloroform/methanol (1:1, v/v), 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).

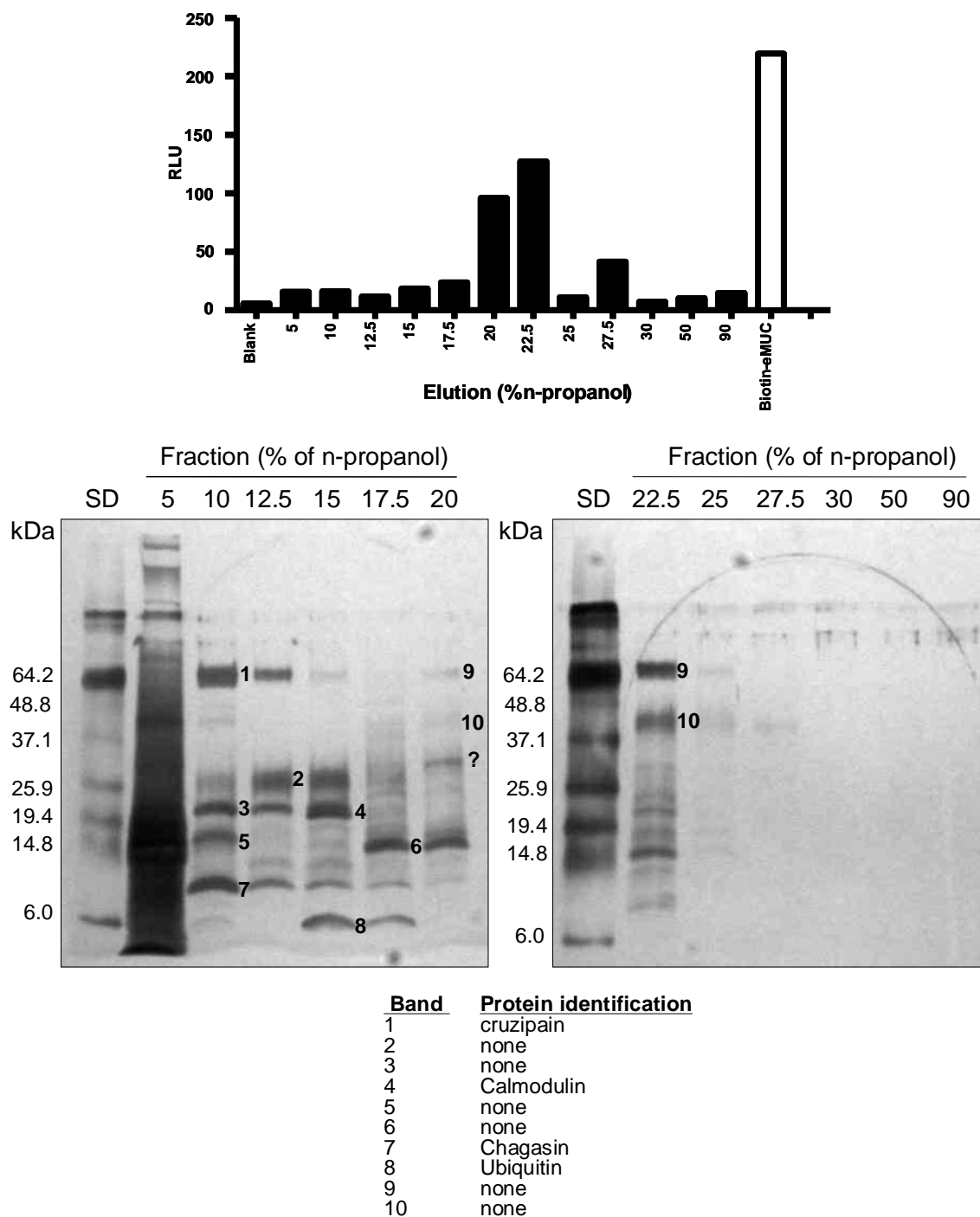


Figure 2.2 (B) CLA (top) and SDS-PAGE/silver stain (bottom) of POROS R1 50 fractions. Numbers on the bottom figure indicate the spots analyzed by LC-MS/MS. Biotin-eMUC, biotinylated epimastigote mucins.

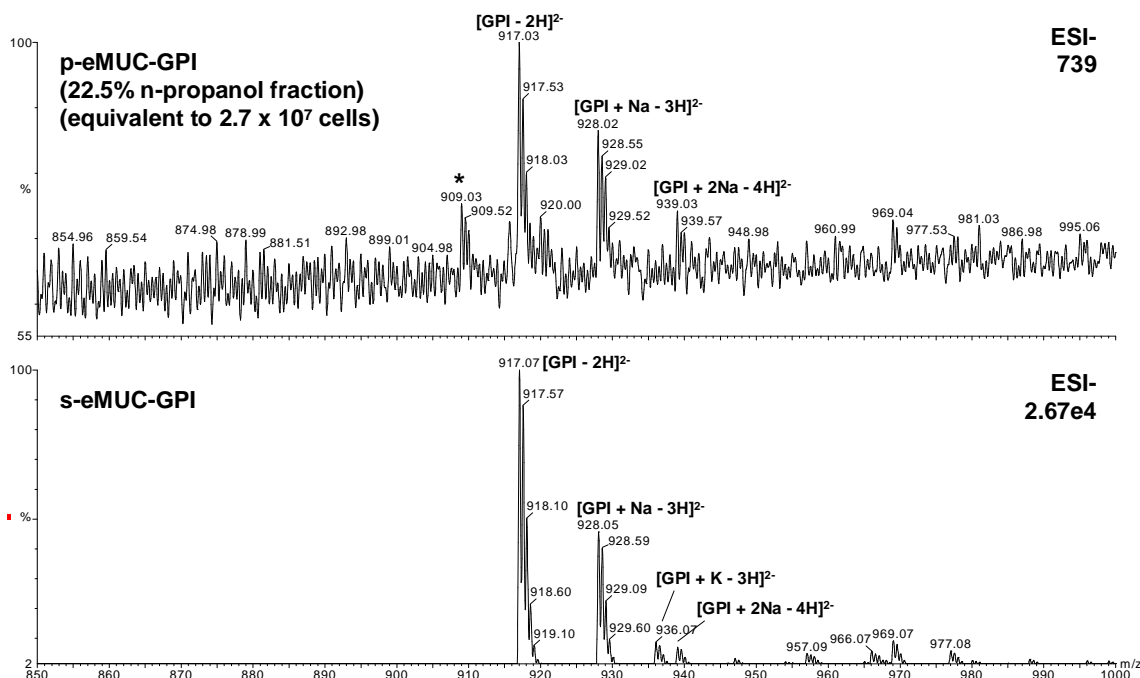


Figure 2.2 (C) The fraction eluted with 22.5% n-propanol was digested with proteinase K, purified in a POROS R1 50 ziptip, and analyzed in negative-ion mode using an ESI-QTOF-MS (Micromass Qtof-1, Waters). The top panel represents the MS spectrum of the purified epimastigote mucin (p-eMUC-GPI) fraction, whereas the bottom panel is the MS spectrum of the synthetic *T. cruzi* epimastigote mucin GPI (s-eMUC-GPI) (Man-[EtNP]Man-Man₂-GlcN[AEP]-InsP-Gro-1-O-C16:0-alkyl-2-O-C16:0-acyl) (Yashunsky *et al*, 2006). The asterisk indicates another species of GPI, previously characterized as [AEP]Man₄-GlcN[AEP]-InsP-Gro-1-O-alkyl-C16:0-2-O-acyl-C16:0 (Almeida *et al*, 2000).

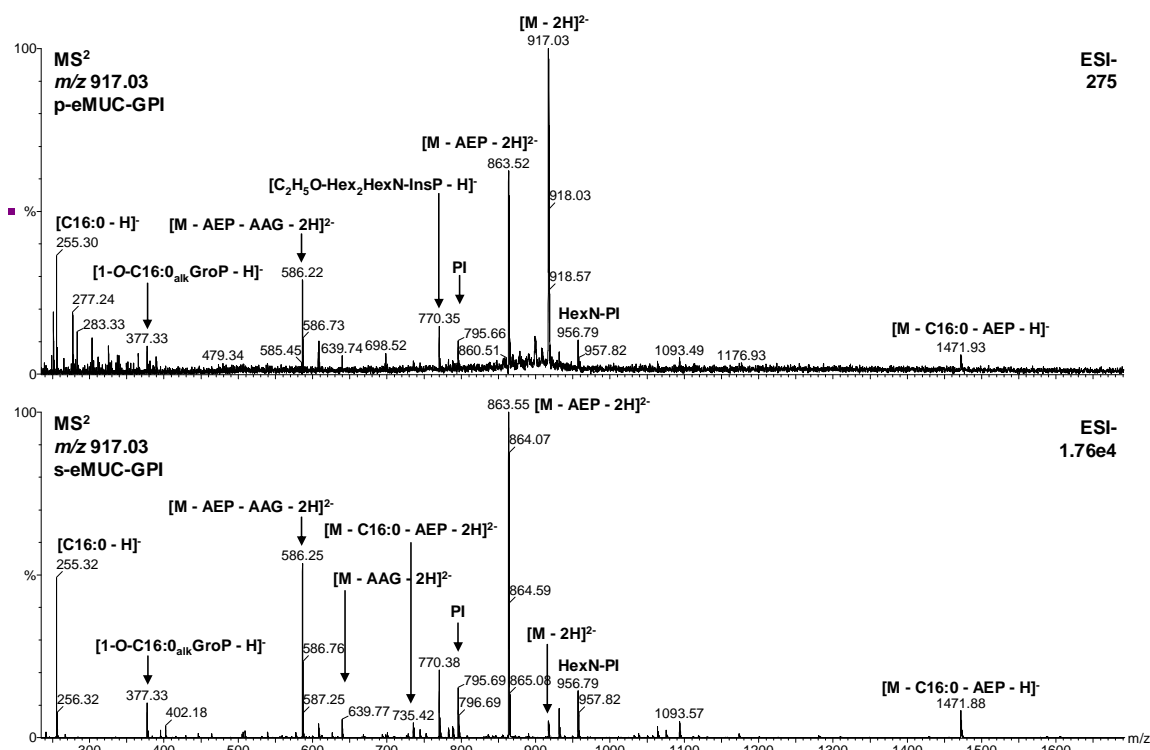


Figure 2.2 (D) Tandem MS spectrum of parent ion at m/z 917.0. The top panel represents the MS² spectrum of the purified *T. cruzi* epimastigote mucin GPIs (p-eMUC-GPIs), whereas the bottom panel represents the MS³ spectrum of the POROS R1-purified synthetic eMUC-GPI (s-eMUC-GPI). Noteworthy, the same 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; GroP, 3-O-phosphoglycerol; Hex, hexose; HexN, hexosamine; Ins, *myo*-inositol; InsP, *myo*-inositolphosphate; M, molecular mass; Man, mannose; m/z , mass-to-charge ratio; PI, phosphatidylinositol.

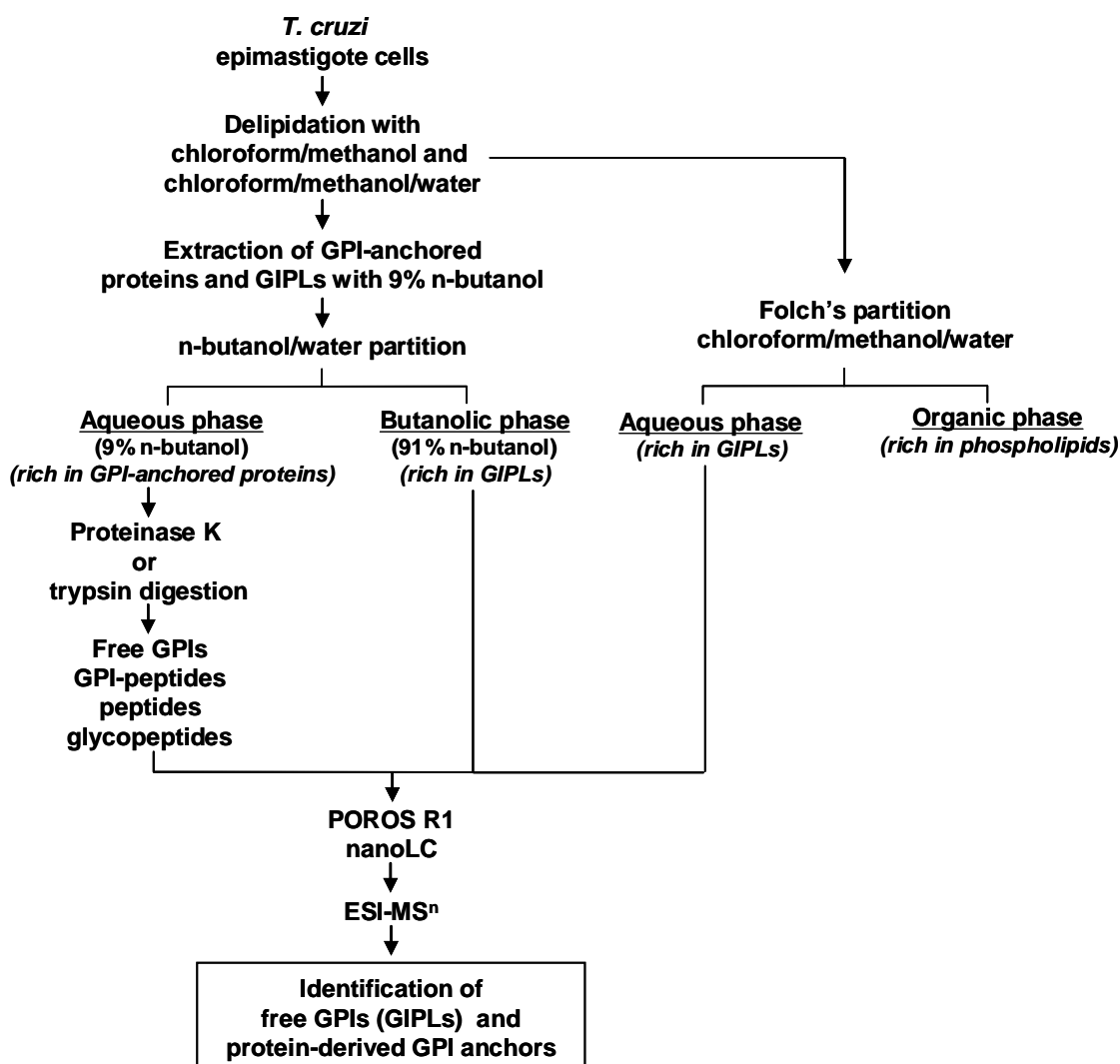
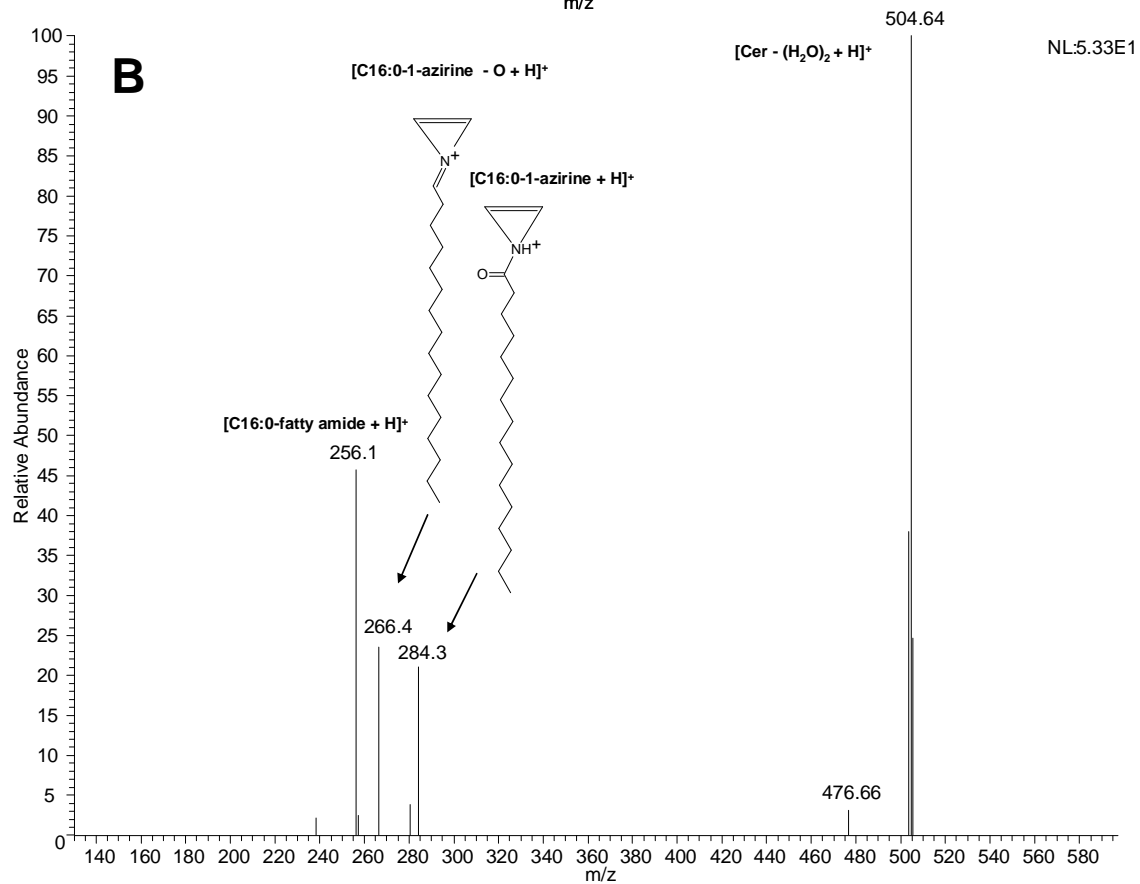
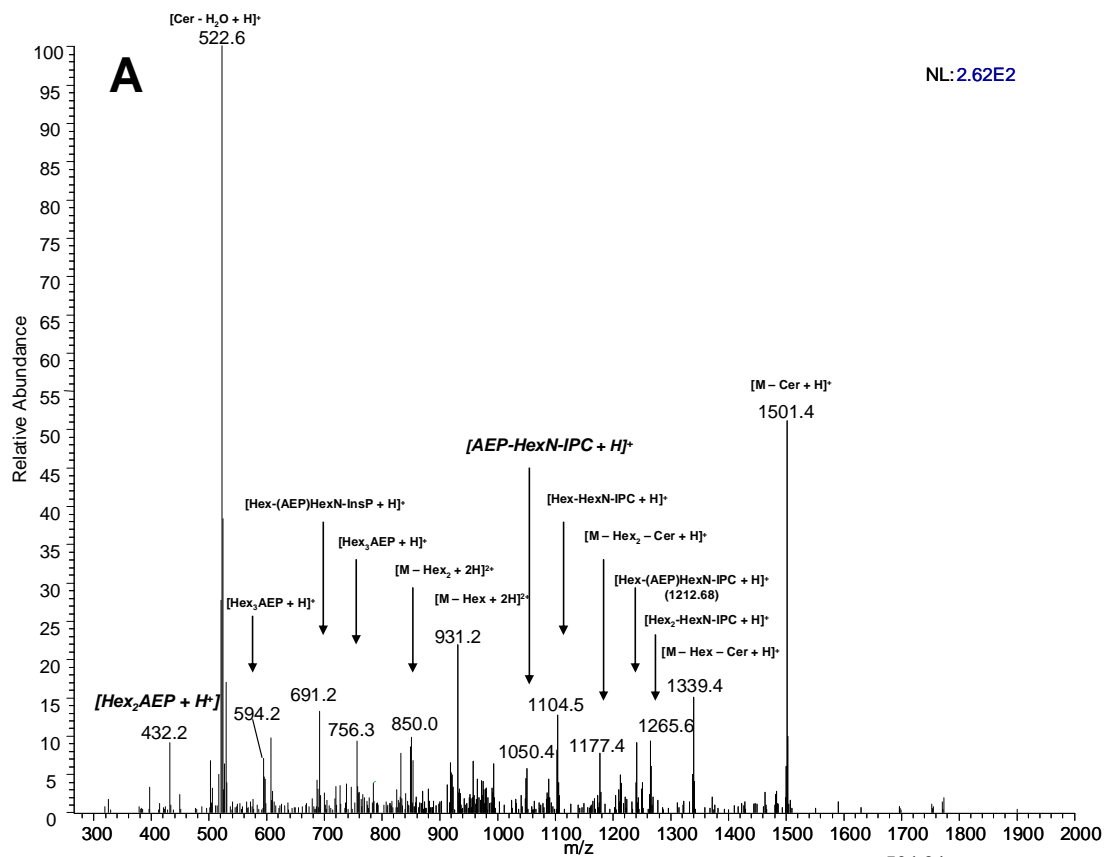


Figure 2.3 General GPIomic approach. Organic extracts rich in GIPLs and GPI-anchored proteins were obtained as described in Materials and Methods. The fraction rich in GIPLs was directly analyzed by LC-MS² and LC-MS³, whereas the fraction rich in GPI-anchored proteins were digested with proteinase K or trypsin prior to LC-MS² and LC-MS³ analyses. In both cases, the LC step was carried out using a POROS R1 10 column (75 μ m x 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.



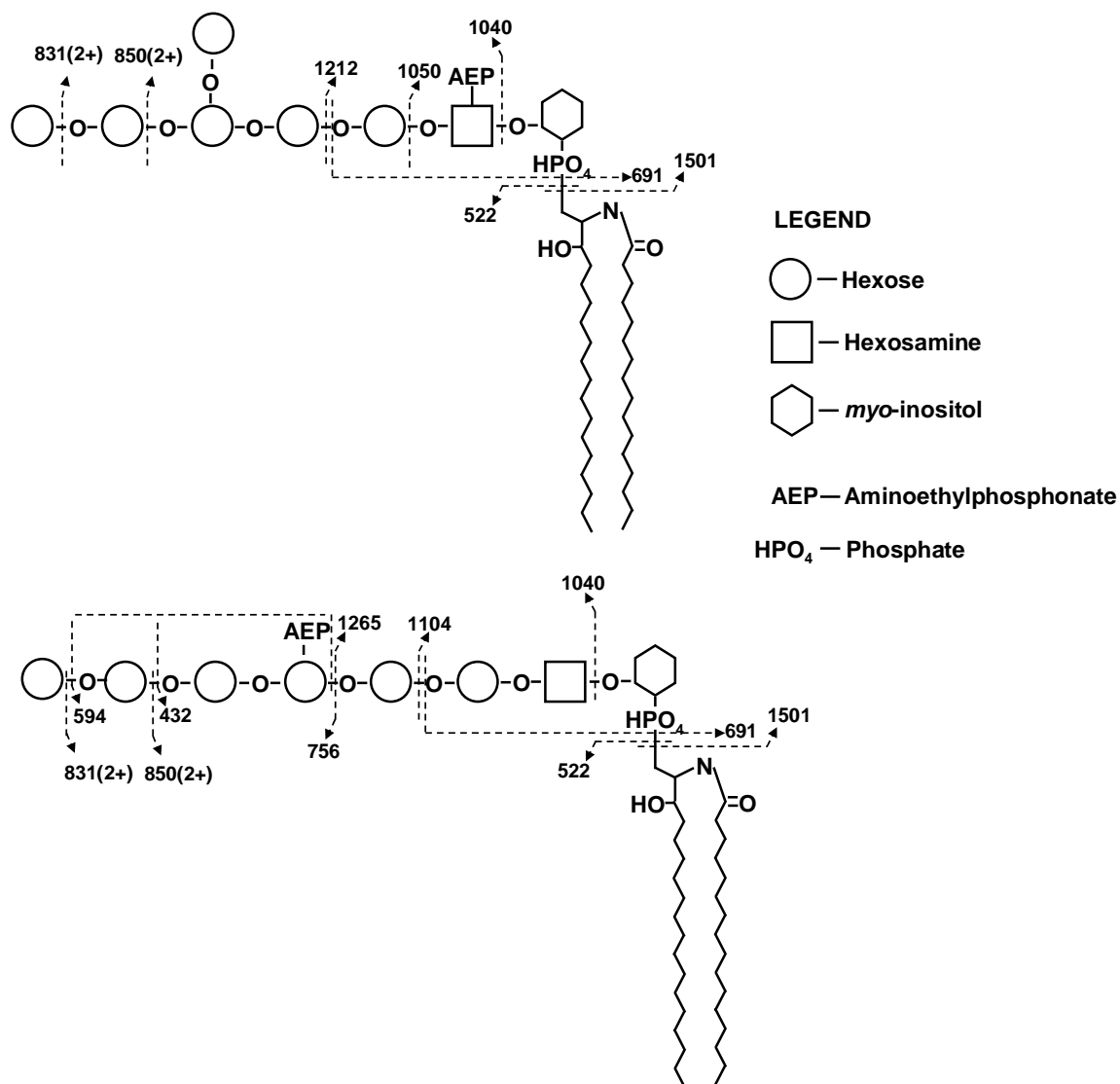


Figure 2.4 (A) Annotated MS² spectrum from GIPL structure at m/z 1012.49 corresponding to AEP-Hex6-HexN-InsP-C16:0/d18:0-Cer. **(B)** MS³ spectrum for the dehydrated ceramide fragment at m/z 522.59. **(C)** Proposed fragmentation and structure of novel GIPL species ([AEP]Hex6-HexN-InsP-C16:0/d18:0-Cer and Hex6-[AEP]HexN-InsP-C16:0/d18:0-Cer). AEP, aminoethylphosphonate; EtNP, ethanolaminephosphate; Gro, glycerol.

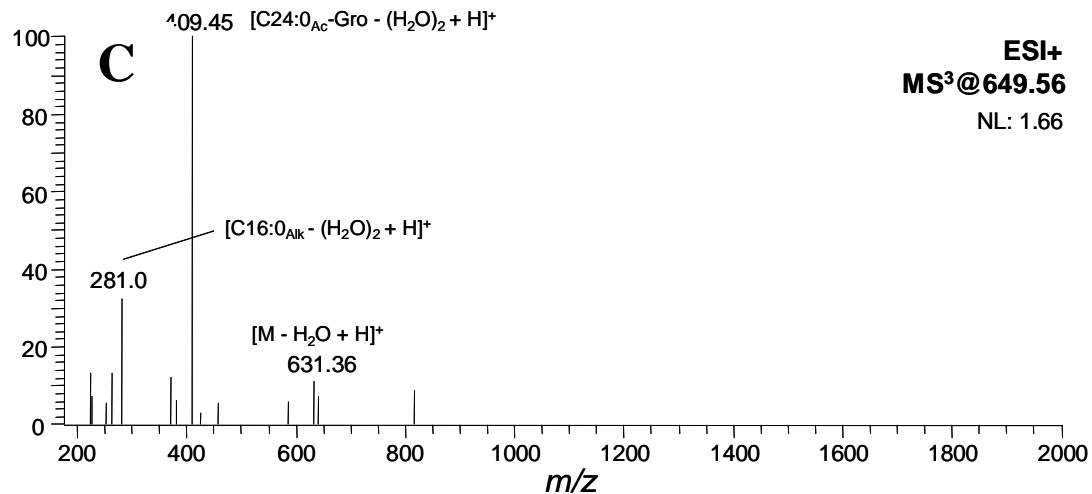
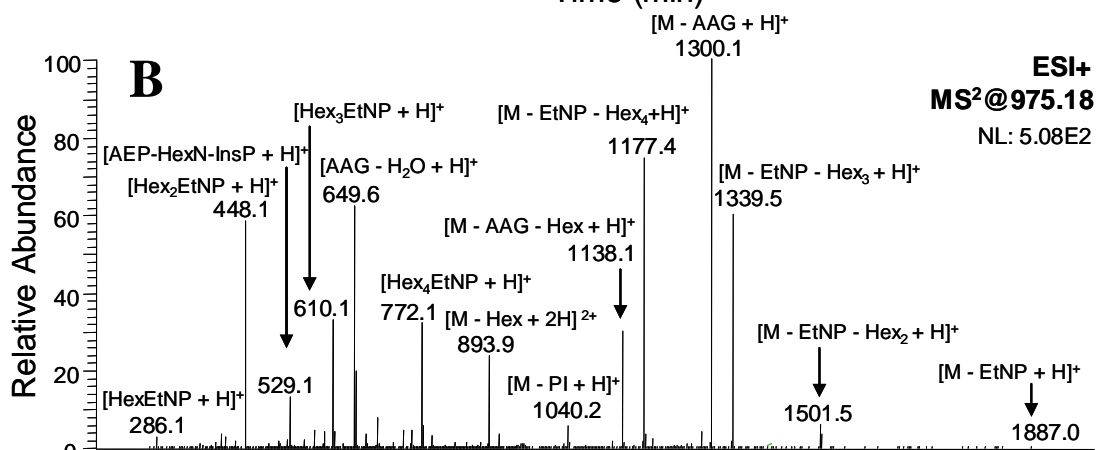
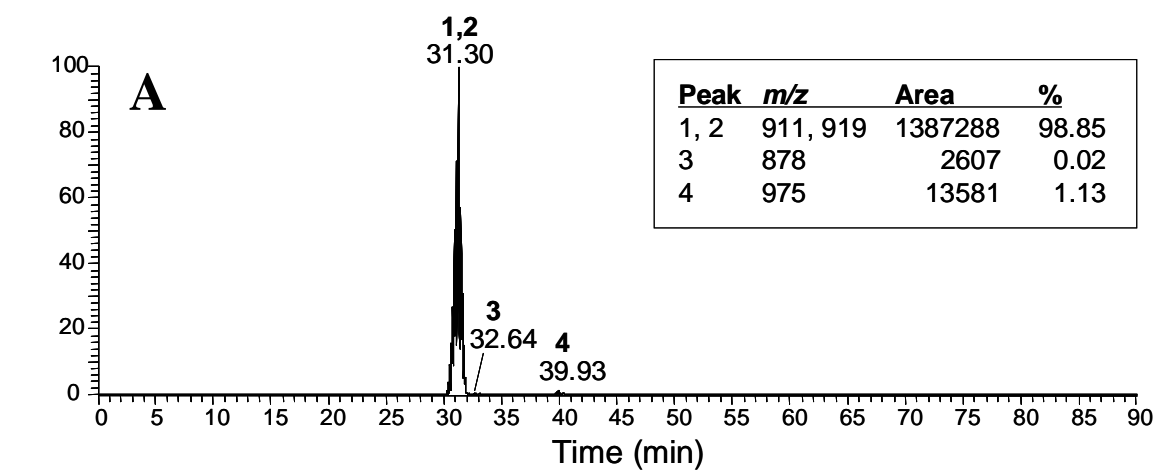


Figure 2.5 (A) Data-dependent acquisition (DDA) (no dynamic exclusion enabled) LC-MS² analysis of epimastigote GPIs. The extracted-ion chromatogram was plotted for 4 major GPI species. The plotted ion species correspond to the dehydrated alkylacylglycerol (AAG - H₂O) and the loss of alkylacylglycerol moiety (M - AAG). The insert shows details about the *m/z* and the area of each peak. (B) Annotated MS² spectrum from GPI structure at *m/z* 975.18 corresponding to EtNP-Hex₄-[AEP]HexN-InsP-Gro-1-O-alkyl-16:0-2-O-acyl-24:0. (C) MS³ spectrum for the dehydrated AAG fragment at *m/z* 649.56. AEP, aminoethylphosphonate; EtNP, ethanolaminephosphate; Gro, glycerol.

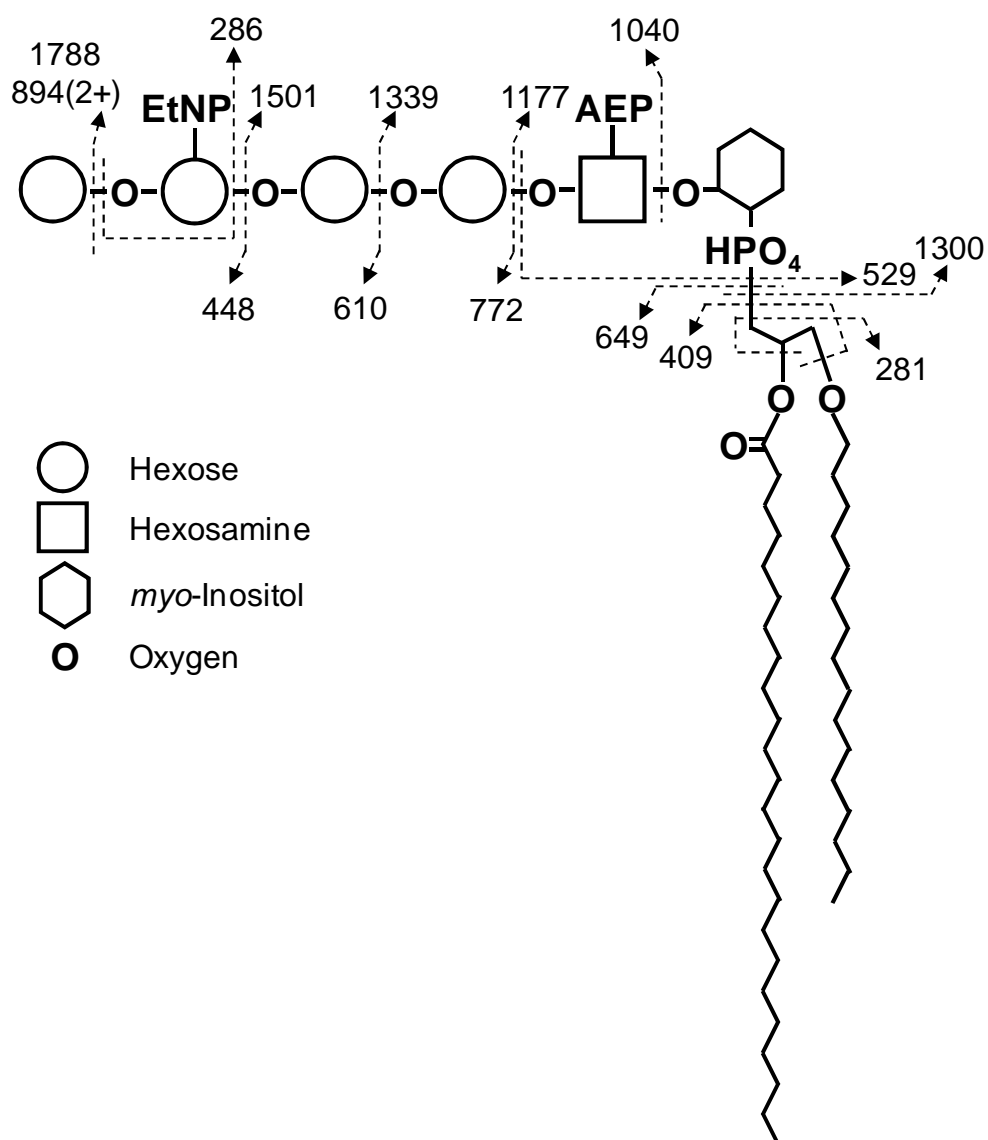


Figure 2.5 (D) Proposed fragmentation and structure of the novel GPI species (EtNP-Hex₄-[AEP]HexN-InsP-Gro-1-O-alkyl-16:0-2-O-acyl-24:0) observed at m/z 975.18.

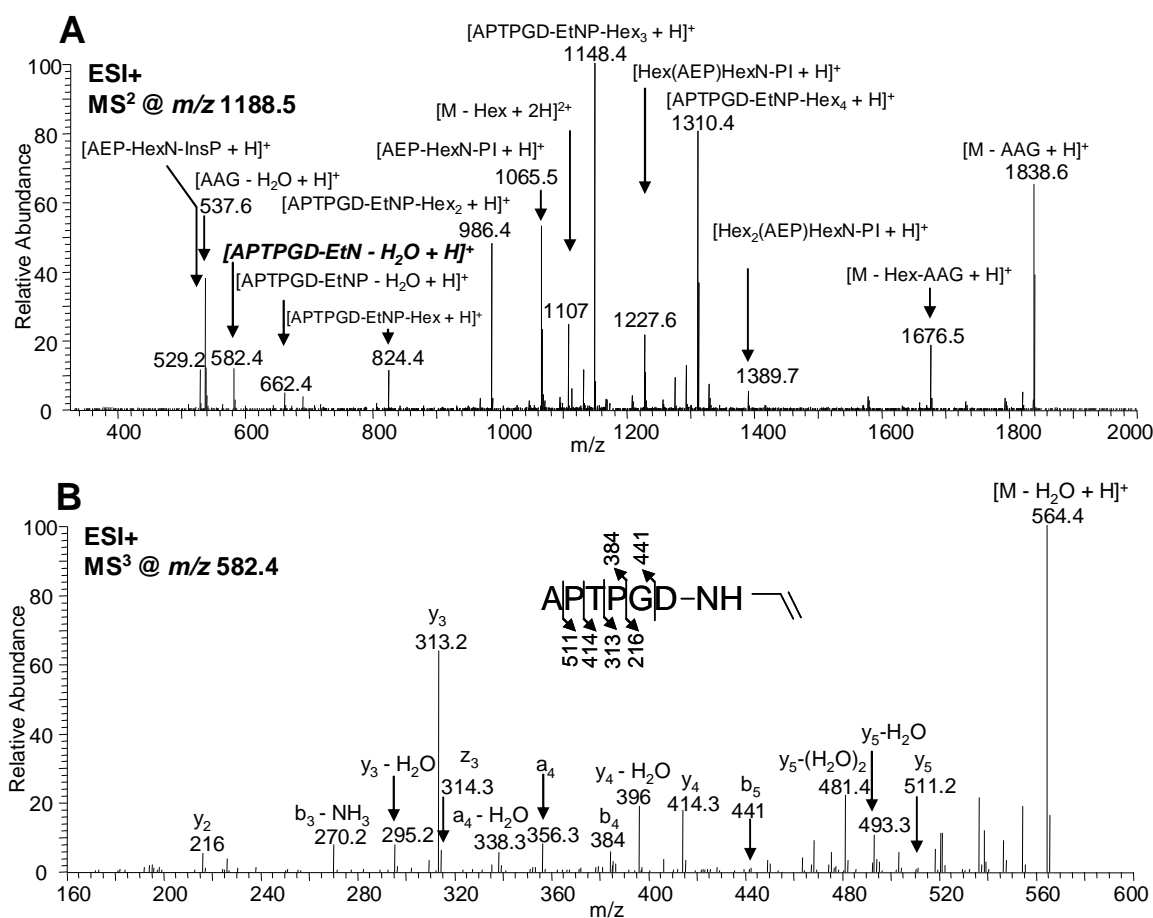


Figure 2.6 Analysis of GPI-peptides and free GPIs released by protease treatment. Analysis by LC-MS² and LC-MS³ of GPI-peptides released by trypsin treatment. **(A)** MS² spectrum of the GPI-peptide species at m/z 1188.5, which corresponds to APTPGD-EtNP-Hex₄-[AEP]HexN-InsP-Gro-1-O-alkyl-16:0-2-O-acyl-16:0. **(B)** MS³ spectrum of the fragment corresponding to the peptide attached to dehydrated EtN at m/z 582.4. The peptide fragments are indicated. AEP, aminoethylphosphonate; EtNP, ethanolaminephosphate; Gro, glycerol.

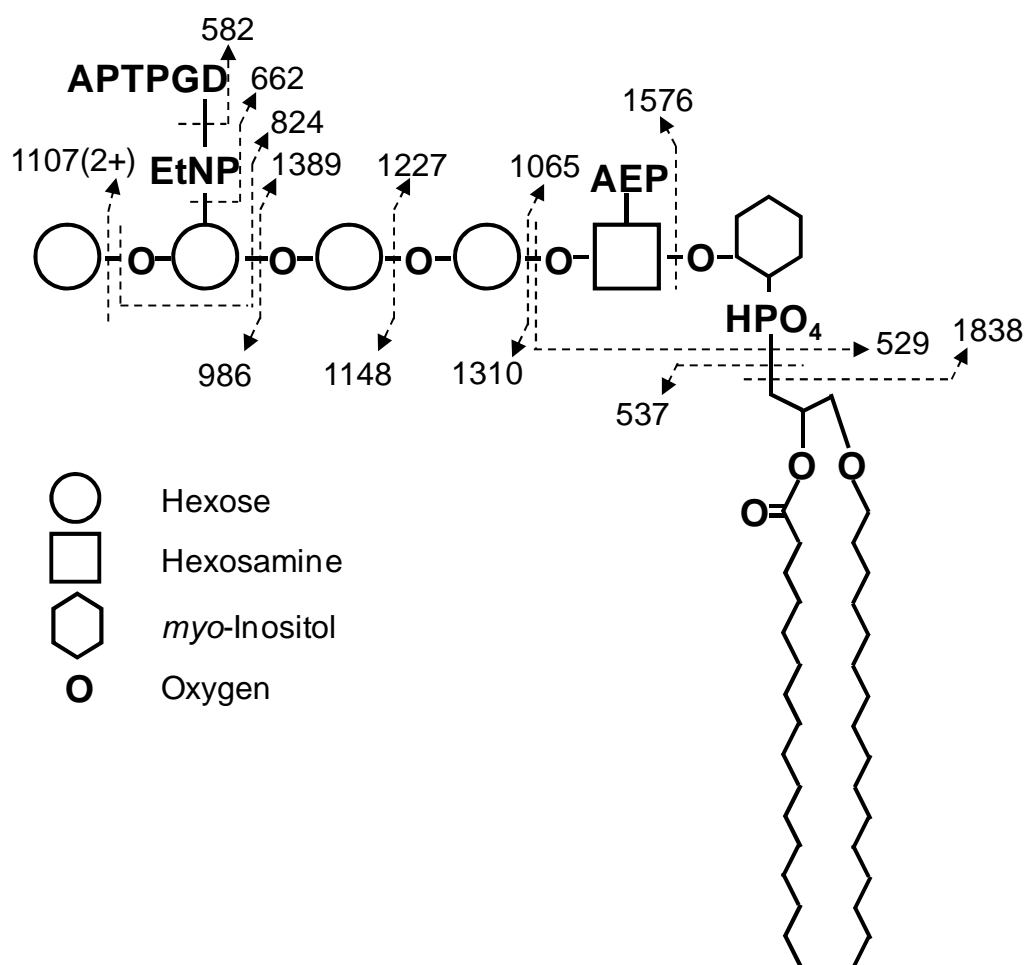


Figure 2.6 (C) Proposed fragmentation and structure of the GPI-peptide APTPGD-EtNP-Hex₄-[AEP]HexN-InsP-Gro-1-O-alkyl-C16:0-2-O-acyl-C16:0.

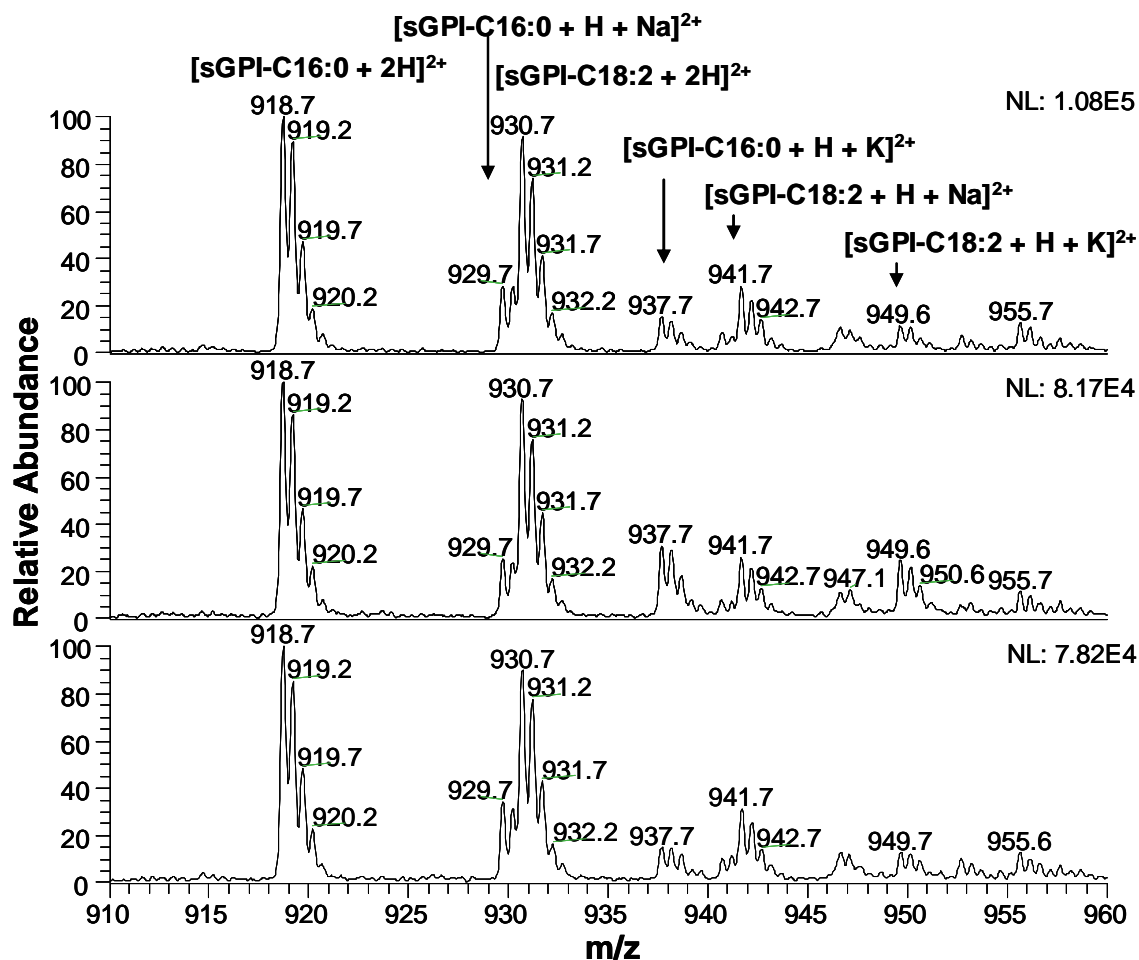


Figure 2.7 (A) Analysis of synthetic GPI ionization efficacy. Synthetic GPIs (Man-[EtNP]Man-Man2-[AEP]GlcN-InsP-Gro-1-O-alkyl-C16:0-2-O-acyl-C16:0 (sGPI-C16:0) and Man-[EtNP]Man-Man2-[AEP]GlcN-InsP-Gro-1-O-alkyl-C16:0-2-O-acyl-C18:2 (sGPI-C18:2) were mixed in the proportion 1:1 (5 μ M each) and analyzed by nanospray linear trap-MS. The triplicate is shown above.

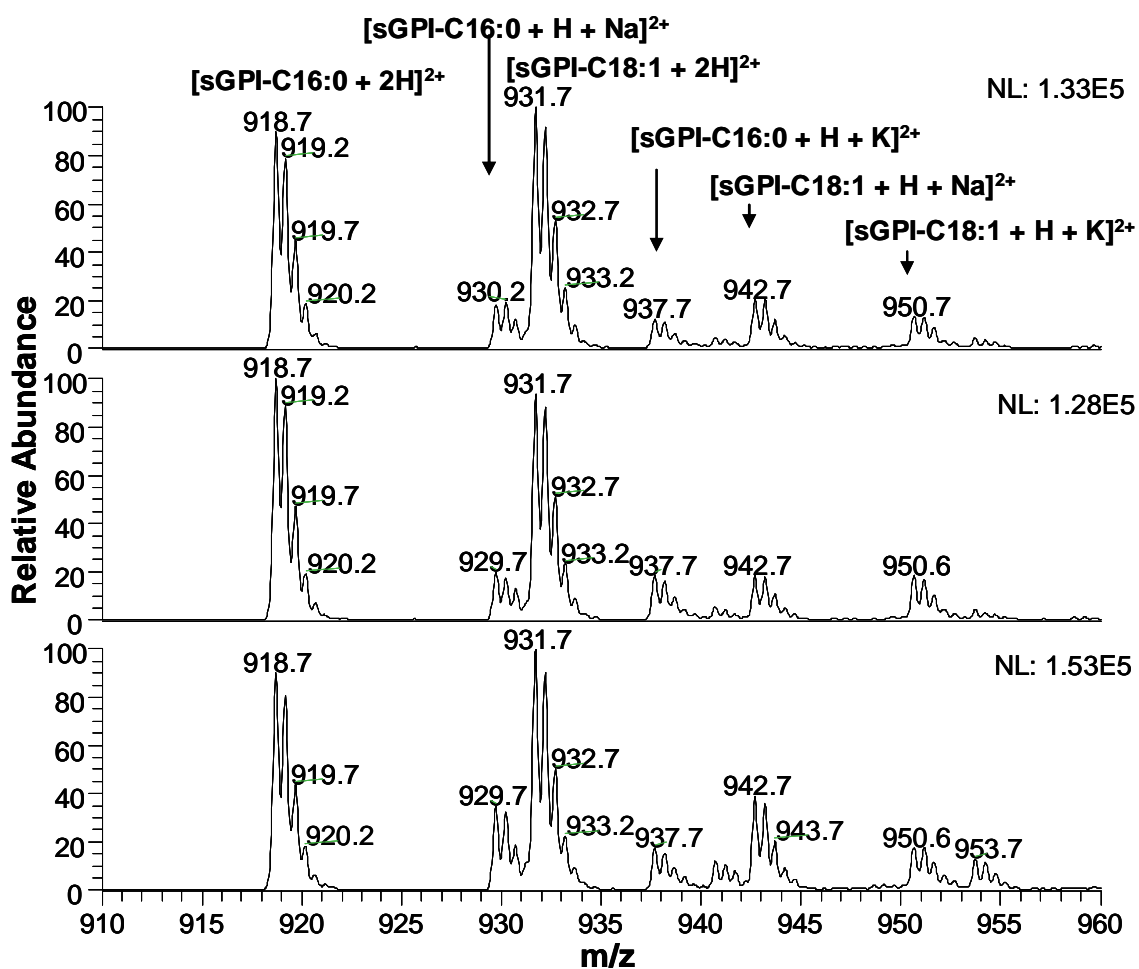


Figure 2.7 (B) Analysis of synthetic GPI ionization efficacy. Synthetic GPIs (Man-[EtNP]Man-Man2-[AEP]GlcN-InsP-Gro-1-O-alkyl-C16:0-2-O-acyl-C16:0 (sGPI-C16:0) and Man-[EtNP]Man-Man2-[AEP]GlcN-InsP-Gro-1-O-alkyl-C16:0-2-O-acyl-C18:1 (sGPI-C18:1) were mixed in the proportion 1:1 (5 μ M each) and analyzed by nanospray linear trap-MS. The triplicate is shown above.

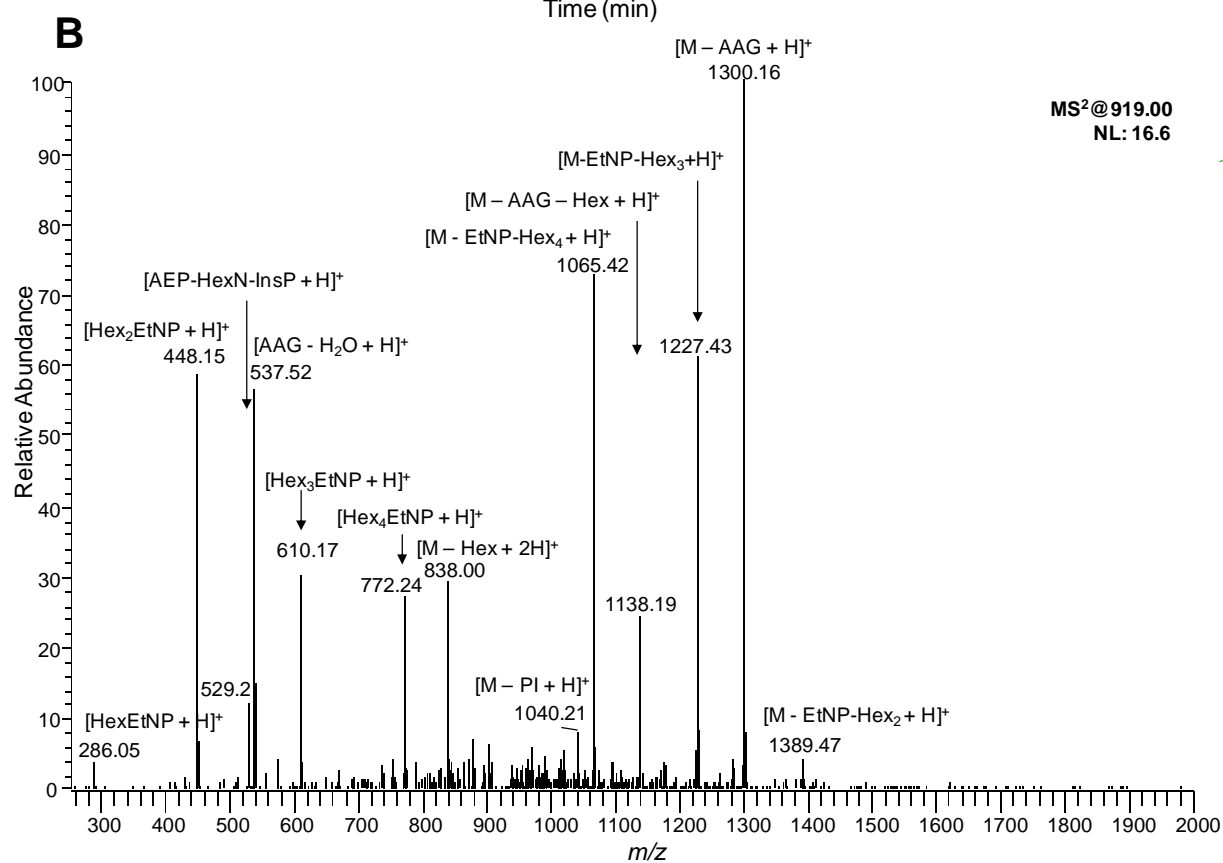
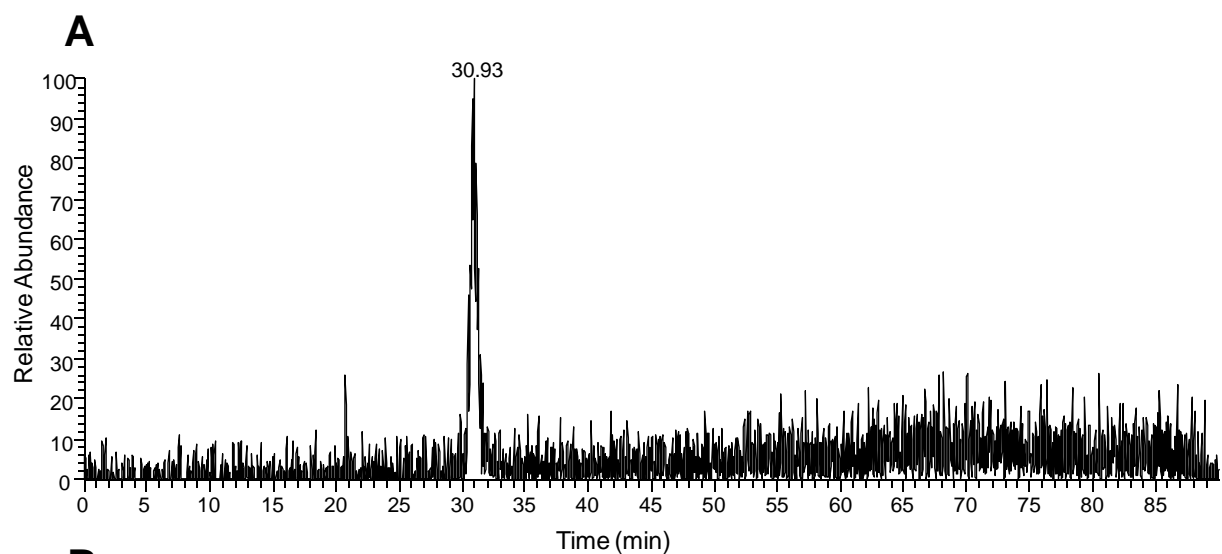


Figure 2.8 Analysis by LC-MRM-MS of proteinase K-released GPIs from *T. cruzi* epimastigotes. The two major GPI species originally observed at m/z 919 and 911 (Supplementary Table I) were subjected to MRM analysis **(A)** Extracted-ion chromatogram of the diagnostic ions corresponding to the dehydrated alkylacylglycerol (AAG - H₂O) ion species (m/z 537.5), and the loss of alkylacylglycerol (M - AAG) (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 10⁵ cells, corresponding to 4.3 femtomoles of total GPIs, was used. **(B)** MS² spectrum of the major GPI species ([EtNP]Hex₄-[AEP]HexN-InsP-Gro-1-O-alkyl-C16:0-2-O-acyl-C16:0) at m/z 919.

Chapter 3: Proteomic Analysis of Detergent-Solubilized Membrane Proteins from Insect-Developmental Forms of *Trypanosoma cruzi*

Esteban M. Cordero,^{†,#} Ernesto S. Nakayasu,^{‡,#} Luciana G. Gentil,[†] Nobuko Yoshida,[†] Igor C. Almeida,^{‡,*} and José Franco da Silveira ^{†,*}

[†] Departamento de Microbiologia, Imunologia e Parasitologia, Escola Paulista de Medicina, UNIFESP, Rua Botucatu, 862, CEP 04023-062, São Paulo, Brazil; [‡] Department of Biological Sciences, The Border Biomedical Research Center, University of Texas at El Paso, El Paso, TX 79912, USA.

[#] These authors contributed equally to the work.

^{*} To whom correspondence should be addressed: Igor C. Almeida, Phone: 915-747-6086; Fax: 915-747-5808; E-mail: icalmeida@utep.edu; José Franco da Silveira, Phone: +55-11-5576-4532, Fax: +55-11-5571-1095, Email: jose.franco@unifesp.br

Keywords: *Trypanosoma cruzi*, insect developmental forms, GPI-anchored proteins, Triton X-114 partitioning, surface glycoproteins, *trans*-sialidase superfamily, adhesins, membrane proteins

3.1 Abstract

The cell surface of *Trypanosoma cruzi*, the etiological agent of Chagas' disease, is covered by a layer of glycosylphosphatidylinositol (GPI)-anchored molecules. These molecules are involved in a variety of interactions between this parasite and its mammalian and insect hosts. Here, we extracted from *T. cruzi* insect developmental forms, by using neutral detergent, the fractions rich in GPI-anchored and other membrane proteins, which were digested with trypsin and analyzed by two-dimensional liquid chromatography coupled to tandem mass spectrometry. Using this approach, 98 proteins from metacyclic trypomastigotes and 280 from epimastigotes were identified. Of these proteins, 64.8% have potential posttranslational modifications or transmembrane domains that could explain their solubility in detergent solution. The identification of some of these modified proteins was also validated by immunoblotting. We also present evidence that the noninfective proliferative epimastigote forms are richer in metabolic enzymes, while the infective nonproliferative metacyclic trypomastigote forms are enriched in surface proteins related to host cell invasion. Metacyclic forms express a large repertoire of surface glycoproteins GP90 and GP82 that are involved in the adhesion of the parasite to the surface of mammalian cells. Taken together, our results show stage-specific protein profiles that appear to be related to the biology of each parasite form.

3.2 Introduction

The major cell surface molecules expressed by insect-transmitted protozoan parasites from the order Kinetoplastida (*Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania spp.*) are glycosylphosphatidylinositol (GPI)-anchored glycoconjugates. *T. brucei* major glycoconjugate is the variant surface glycoprotein, while the *T. cruzi* glycocalix is mainly formed by mucins and members of the *trans*-sialidase (TS) superfamily. The glycocalix coat in trypanosomatids is also rich in complex glycopospholipids, such as lipophosphoglycans (LPG) and glycoinositolphospholipids (GIPLs) (Almeida *et al*, 1999; Colli and Alves, 1999; Ferguson, 1999). Provided that GPI-anchored molecules have been shown to be involved in target cell invasion, signaling and activation of the host proinflammatory response, they are attractive targets for the development of vaccines or drugs against these pathogenic microorganisms.

T. cruzi is the etiological agent of Chagas' disease, or American trypanosomiasis, an infection without any effective drug treatment or immunotherapy that affects around 10 million people in the Americas. *T. cruzi* is naturally transmitted to mammalian hosts through the urine and feces of infected triatomine bugs. When a triatomine feeds on an infected animal, bloodstream trypomastigotes are ingested. These parasites reach the insect's midgut where they differentiate into epimastigotes, which proliferate by asexual reproduction, and later in the insect's rectum, they transform into the infective metacyclic trypomastigotes. Upon release in the insect's feces when it takes another blood meal, metacyclic forms can infect a new mammalian host by penetrating through the wound and invading host cells.

Similarly to the mammalian stages of the parasite, the cell surface of epimastigotes and metacyclic trypomastigotes is covered by a large amount of GPI-anchored mucins and GIPLs, formerly designated lipopeptidophosphoglycan (LPPG). (Almeida *et al*, 1999; Colli *et al*, 1999;

Ferguson, 1999) Metacyclic trypomastigotes which establish the initial parasite-host cell interaction, express two major stage-specific GPI-anchored glycoproteins (GP90 and GP82) with no counterpart in bloodstream trypomastigotes (de Almeida and Heise, 1993; Güther *et al*, 1992; Teixeira and Yoshida, 1986). GP82 is a cell adhesion molecule that induces a bidirectional Ca^{2+} response, an event essential for *T. cruzi* penetration into the host cell (reviewed in Yoshida, 2006). As opposed to GP82, GP90 binds to mammalian cells in a receptor-mediated manner without triggering a Ca^{2+} signal and functions as a negative modulator of cell invasion (Yoshida, 2006). Sequence analysis showed that GP90 and GP82 share similarity with members of the *T. cruzi* *trans*-sialidase (TS) superfamily, which comprises a large number of genes encoding major surface antigens of the parasite infective forms (Araya *et al*, 1994; do Carmo *et al*, 2002; El-Sayed *et al*, 2005).

The plasma membrane of *T. cruzi* is likely to contain proteins that could serve as novel drug targets, diagnostic probes or antigens for vaccines against Chagas' disease. These surface proteins are encoded by several hundred genes, and would be essential to know which genes are expressed. However, as *T. cruzi* regulates gene expression primarily by post-transcriptional mechanisms, such as mRNA turnover and translation control, this limits the use of tools like DNA microarrays to study gene expression in the different developmental forms of the parasite. Thus, sequencing of the proteins by proteomic analysis is particularly useful. In the present study we undertook a comparative proteome analysis of a GPI-anchored membrane protein enriched-fraction from epimastigotes and metacyclic trypomastigotes. Identifying and characterizing these membrane proteins is a special challenge because of their structural complexity and physicochemical properties. By using a simple, fast and sensitive method for the isolation through extraction with the detergent Triton X-114, GPI-anchored proteins were identified by immunoblotting and two-dimensional liquid chromatography

coupled to tandem mass spectrometry. This study highlights the efficiency of an integrative proteomics approach that combines experimental and computational methods to provide the selectivity, specificity and sensitivity required for characterization of posttranslationally modified membrane proteins.

3.3 Materials and Methods

3.3.1 Parasites

T. cruzi strain G was maintained alternately in mice, or at 28 °C in liver infusion tryptose (LIT) medium containing 5 % fetal calf serum (FCS). Metacyclic trypomastigotes were harvested from cultures in the stationary growth phase and purified by anion-exchange chromatography on a DEAE-cellulose column as described elsewhere (Teixeira *et al*, 1986).

3.3.2 Protein extraction and Triton X-114 partitioning

The proteins were extracted according to a protocol published elsewhere, with minimal modifications (Ko and Thompson, 1995). Parasites from cultures were washed three times with PBS to remove the proteins from the medium. A parasite pellet containing 1×10^8 cells was homogenized in 1 mL of lysis buffer (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2% TX-114, 1 mM PMSF) on an ice for 1 h with periodic agitation. The homogenate was clarified by centrifugation at $8,800 \times g$ for 10 min at 0 °C and the supernatant (S1) was stored at -20 °C for 24 h. The pellet was re-extracted with buffer A (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.06% TX-114 and 1 mM PMSF), incubated for 10 min on ice and centrifuged at $8,800 \times g$ at 0 °C to produce the pellet (P1) and supernatant (SBA).

After incubation at -20 °C, the first supernatant was thawed, homogenized and submitted to phase separation by incubation at 37 °C for 10 min. The phases were separated by centrifugation at $3,000 \times g$ for 3 min at room temperature. The upper phase (S2) was collected and the detergent-rich phase (lower phase) re-extracted with 1 mL of buffer A, mixed and incubated at 0 °C for 10 min and then submitted to a new phase extraction under the same conditions as above. The upper phase (S3) was collected, and the detergent-rich phase

was extracted with 1 mL of buffer A, homogenized, incubated for 10 min at 0°C, and clarified by centrifugation at 18,000 x g for 10 min at 0 °C. The pellet (P2) was stored and the upper phase (S4) was submitted to a new phase separation as described above. The upper phase (S5) was then collected and the detergent rich phase (approximately 100 µL) was precipitated with 3 vol of cold acetone at -20 °C for 30 min. The proteins were recovered by centrifugation at 18,000 x g at 0 °C, and the pellet was dried on the bench. This final extract was named GPI-rich fraction.

3.3.3 Protein digestion and peptide purification

GPI-enriched fractions were resuspended in 20 µL of 400 mM NH_4HCO_3 containing 8 M urea and submitted to trypsin digestion (Stone and Williams, 1996). After reducing the disulfide bonds with 5 mM dithiothreitol at 50°C for 15 min, the cysteine residues were alkylated with 10 mM iodoacetamide for 15 min at room temperature and protected from light. The reaction mixture was then diluted 8-fold and incubated overnight at 37°C after 2 µg sequencing-grade trypsin (Promega, Madison, WI) had been added. The reaction was terminated by adding 0.05% trifluoroacetic acid (TFA) (final concentration) and the digested proteins were desalted using reverse phase (RP) ZipTip columns (POROS R2 50, Applied Biosystems, Foster City, CA), as described elsewhere (Jurado *et al*, 2007).

To remove the detergent present in the preparation, the peptides were purified/fractionated in a strong cation-exchange (SCX) ZipTip. Peptides eluted from the RP ZipTip purification were diluted by addition of an equal volume of SCX buffer [25% acetonitrile (ACN)/0.5% formic acid (FA)]. SCX ZipTip columns were manufactured in our laboratory using POROS HS 50 resin (Applied Biosystems) and equilibrated with SCX buffer. After loading the samples, the column was exhaustively washed with SCX buffer (30 times the column volume)

to remove any remaining traces of detergent from the preparation. The peptides were eluted with increasing concentrations of NaCl (25, 50, 100, 200 and 500 mM) dissolved in the same buffer. The fractions were dried in a speed vac (Vacufuge, Eppendorf) to remove the ACN, desalted in reverse phase ZipTips and dried again.

3.3.4 Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis

Each fraction was resuspended in 20 μ L of 0.1% FA and analyzed by LC-MS/MS in a Qtof (Micromass Qtof-1, Waters, Milford, MA) or a linear trap (LTQ XL with ETD, Thermo Fisher Scientific, San Jose, CA). For the Qtof analysis, 1 μ L was injected into a PepMap reversed-phase column (15 cm x 75 μ m, 3- μ m C18, LC Packings, Dionex) coupled to an Ultimate (LC Packings, Dionex) nanoHPLC. The run was performed in a flow rate of 300 nL/min with a gradient from 5 to 35% ACN/0.1% FA over 100 min, and the eluting peptides were directly analyzed in the Qtof mass spectrometer. Spectra were collected in the positive-ion mode in the 400–1800 m/z range for 2 s, and the three most abundant peptides were fragmented with a ramp collision energy (22-60 eV) for 3 seconds in the 50–2050 m/z range. The dynamic exclusion was set to fragment each peptide only once and then to exclude for 2 min.

For linear trap analysis, aliquots of 8 μ L from each sample were loaded into a C18 trap column (0.25 μ L, C18, OPTI-PAK. Oregon City, OR). The separation was performed on a capillary reverse-phase column (Acclaim, 3 μ m C18, 75 μ m x 25 cm, LC Packings. Amsterdam, The Netherlands) connected to a nanoHPLC system (nanoLC 1D plus, Eksigent, Dublin, CA). Peptides were eluted using a linear gradient from 5 to 40% ACN in 0.1% formic acid for 200 min and directly analyzed in a linear ion-trap mass spectrometer equipped with a nanospray source. MS spectra were collected in centroid mode in a range from 400 to 1700

m/z, and the five most abundant ions were submitted twice to CID (35% normalized collision energy) before dynamic exclusion for 120 sec.

3.3.5 Database search and protein/peptide identification

MS/MS spectra collected in the Qtof-MS were converted into peak lists (PKL format) using ProteinLynx 2.0 (Waters). MS peaks were smoothed twice with 5 channels using the Savitzky-Golay method and centered with 4 channels in the top 80%. For the MS/MS peaks a threshold of 5% was set. After smoothing twice with 3 channels using the Savitzky-Golay method, the peaks were centered with 4 channels in the top 80% and converted into monoisotopic ions. MS/MS spectra collected in the linear trap-MS from peptides with masses of 600-3500 Da, at least 100 counts and 15 fragments were converted into DTA files using Bioworks v.3.3.1 (Thermo Fisher Scientific).

Linear trap data were submitted to database search using Sequest (Eng *et al*, 1994) (available in Bioworks) and both Qtof and linear trap data were searched with Phenyx (Colinge *et al*, 2003) (web interface, version 2.5, GeneBio, Switzerland). The database used in the analyses consisted of the forward and reverse *T. cruzi*, bovine, human keratin and porcine trypsin sequences from GenBank (191,762 non-redundant sequences downloaded Mar 17th, 2008 from <http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein&cmd=search&term=>).

Database search parameters included; i) trypsin cleavage in both peptide termini with allowed one missed cleavage site allowed, ii) carbamidomethylation of cysteine residues as a fixed modification, iii) oxidation of methionine residues as a variable modification, iv) peptide mass tolerance of 2.0 Da for linear trap-MS or 500 ppm for Qtof-MS data, and v) 1.0 Da (for both linear trap and Qtof-MS data) for fragment mass tolerance. In addition, Qtof-MS data were submitted to a second round of searching by Phenyx with the same parameters as

above, but with a trypsin cleavage in at least one of the peptide termini with two missed cleavage sites allowed, and 1500 ppm for peptide mass tolerance.

To ensure the quality of protein identification, the false-discovery rate (FDR) was estimated using the following formula:

$$\text{FDR: } \frac{\text{Number of proteins matching the reverse sequences}}{\text{Total number of proteins}}$$

The FDR was calculated after applying the following filters in Bioworks (for Sequest analysis): $\text{DCn} \geq 0.1$, protein probability $\leq 1 \times 10^{-3}$ and $\text{Xcorr} \geq 1.5$, 2.0 and 2.5 for singly-, doubly- and triply-charged peptides, respectively. With these parameters we validated all sequences with an $\text{FDR} < 2\%$. For Phenyx analysis, a minimum peptide Z-score of 5 and p-value of 1×10^{-4} were used. In addition, the protein AC score was set according to each dataset to obtain a maximum of 5% FDR.

3.3.6 Bioinformatics analysis of identified *T. cruzi* protein sequences

All identified *T. cruzi* protein sequences were submitted to GPI-anchor prediction analysis using the FragAnchor algorithm (Poisson *et al*, 2007) (available at <http://navet.ics.hawaii.edu/~fraganchor/NNHMM/NNHMM.html>). The prediction of transmembrane domains and myristoylation was performed using the TMHMM and NMT algorithms available at Expasy server (<http://www.expasy.org>).

3.3.7 Polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting

Polyacrylamide gels (10%) were run for about 4 h at 8 mA/gel, and the proteins separated were transferred onto nitrocellulose membranes. The membranes were then blocked with 5.0 % non-fat milk in PBS (PBS-milk) and probed with monoclonal or polyclonal

antibody in PBS-milk. After washing in PBS-milk containing 0.05% Tween 20, the membranes were incubated with anti-mouse IgG conjugated to horseradish peroxidase diluted in PBS-milk. After extensive washing in PBS containing 0.05 % Tween 20, the bands were revealed by chemiluminescence using the ECL Western blotting detection reagent and Hyper film-MP (Amersham). Monoclonal antibodies (MoAb) 1G7, 3F6 and 10D8 are specific for the *T. cruzi* surface glycoproteins GP90, GP82 and GP35/50 (mucins), respectively (Teixeira *et al*, 1986); MoAb 25 is specific for the flagellar calcium-binding protein (FCBP).

3.4 Results and Discussion

3.4.1 SDS-PAGE and immunoblotting analyses of *T. cruzi* GPI-enriched fractions extracted with Triton X-114

Triton X-114 detergent-based two-phase separation was used for partitioning of hydrophobic and soluble membrane proteins in whole cell fractions of different developmental forms of *T. cruzi*. This targeted modification-specific proteomic strategy can also be applied to the analysis of GPI-anchored membrane proteins in whole parasite extracts. The workflow is outlined in Figure 3.1. This detergent, which is water-soluble at 0°C, fractionates into aqueous and detergent-enriched phases at 37°C. Figure 3.2 shows the analysis of the GPI-enriched fraction from metacyclic trypomastigotes after separation by SDS-PAGE and staining with Coomassie blue (lane 1) or silver (lane 2). This fraction showed three major bands of approximately 90, 24 and 10 kDa and four faint bands of 82, 75, 58 and 14 kDa (Fig. 3.2). Two broad bands of approx 50 and 35 kDa could only be visualized by silver staining (Fig. 3.2, lane 2). Specific monoclonal antibodies (MoAb) were used to identify the proteins present in the GPI-enriched fraction (Fig. 3.2, lanes 3-6). MoAb 1G7, which is specific for GP90, stained the 90 kDa band (lane 3), and MoAb 3F6, directed to GP82, reacted with two bands of 82 and 30 kDa (lane 4). Metacyclic forms from different *T. cruzi* isolates expressed gp30, a surface glycoprotein detectable by MoAb 3F6 (Cortez *et al*, 2003). The reaction with MoAb 10D8, which is specific for *T. cruzi* mucins GP35/50, revealed that these are also major components of the cell surface (lane 5). Mucins were barely detected by silver staining, because of high glycosylation of these molecules. The 24 kDa protein, corresponding to the FCBP, as demonstrated by the reaction with MoAb 25 (lane 6), and the calpain-like cysteine proteases

by reaction with monospecific polyclonal antibodies against the catalytic domain (Cys Pc) of this enzyme (lane 7), appeared as two faint bands of approximately 56 and 60 kDa.

The GPI-enriched fraction from epimastigotes was also analyzed (data not shown). No proteins could be detected by Coomassie blue staining, suggesting that they were present in small amounts. Two major bands of 50 and 35 kDa were detected by silver staining and identified as GP35/50 mucins with MoAb10D8. Our results differ to some extent from those of Añez-Rojas (Añez-Rojas *et al*, 2006), who analyzed the protein profiles of epimastigote detergent extracts from Venezuelan *T. cruzi* isolates and reported the presence of two bands of approximately 30 and 20 kDa. This difference may be due to the fact that different isolates were used in the two studies.

3.4.2 Identification of metacyclic trypomastigote and epimastigote proteins by mass spectrometry

Following cell lysis with Triton X-114, differential centrifugation was performed to obtain fractions enriched in membrane proteins. By following this procedure it is expected that organelle membrane proteins will also be found in the GPI-anchored protein-enriched fraction. Parasite proteins extracted with Triton X-114 were digested with trypsin, fractionated by strong cation-exchange chromatography and subjected to LC-MS/MS. The peak list files were submitted to a database search using Sequest and Phenyx against *T. cruzi* sequences and possible contaminant sequences downloaded from GenBank. The identified protein sequences were further analyzed to predict possible GPI-anchor signals, transmembrane domains and N-terminal myristoylation or prenylation sites (see item 3.3.6 for details).

In total, 378 proteins from epimastigotes and metacyclic trypomastigotes were unambiguously identified from one or more peptide sequences (data not shown). Of these, 245

(64.8%) were predicted to have posttranslational modifications and/or at least one hydrophobic membrane-spanning domain that could account for their presence in the Triton X-114 detergent fraction (Tables 3.1). The remaining proteins represent, at least in part, soluble proteins that were not completely removed by Triton X-114 phase partitioning. Alternatively, some proteins could have physicochemical properties that cause them to partition into the Triton X-114 detergent phase.

We found matching entries in the *T. cruzi* databases for 98 proteins identified in metacyclic trypomastigotes (Tables 3.1). Eighty-two of these proteins (83.7%) had posttranslational modifications and/or hydrophobic membrane-spanning domains that could explain their presence in the Triton X-114 detergent fraction, whereas 16 proteins (16.3%) appeared not to have any potential modification sites. Of the 82 metacyclic proteins predicted to have posttranslational modifications, 21 (21.4%) were GPI-anchored (members of TS superfamily, mucin TcMUC II); 26 (26.5%) contained a typical signal peptide motif or an uncleaved signal-anchor sequence; 15 (15.3%) displayed myristoylation or prenylation targeting sites; 16 (16.3%) contained a mitochondrial targeting sequence; and 4 (4.1%) had transmembrane spanning domain.

The protein content of epimastigotes largely differed from that of metacyclic trypomastigotes (Tables 3.1). Among the 280 proteins identified in epimastigotes (Table 3.1), 163 (58.2%) contained posttranslational modifications and/or hydrophobic membrane-spanning domain, whereas 117 proteins (41.8%) did not show any potential modification sites. Only one protein (Tc85) (0.4%) was predicted to be GPI-anchored; 72 (25.7%) contained a typical signal peptide motif or an uncleaved signal-anchor sequence; 49 (17.5%) contained mitochondrial targeting sequences; 21 (7.5%) displayed myristoylation or prenylation targeting sites; and 20 (7.1%) were transmembrane spanning.

A relatively high proportion of proteins (12.7%) carrying an uncleaved signal-anchor sequence was found. Possibly, many *T. cruzi* surface antigen genes have two or three in-frame ATG initiator codons at their 5' end. For instance, in several TS protein genes, the predicted translation from the first in-frame ATG encodes a sequence which does not conform to any known signal sequences (do Carmo *et al*, 2002; Low and Tarleton, 1997). However, after the second or third methionine residue, there is a typical signal-anchor sequence (see Fig. 3.3). It remains to be determined which is the effective ATG initiator codon used *in vivo* by the parasite.

About 35.2% of the proteins identified here did not display any apparent posttranslational modification and/or transmembrane domain (Tables 3.1). However, some of these proteins, such as cytoskeletal proteins (17 identifications in epimastigotes and 4 metacyclic trypomastigotes), translation elongation factor 1-beta (eEF1B) (2 in metacyclic trypomastigotes and one in epimastigotes), enolase and acidic ribosomal proteins (14 identifications), had physicochemical properties that allowed them to be extracted with the detergent. The cytoskeleton of trypanosomes contains a membrane-associated cytoskeletal structure (membrane skeleton), which consists of a dense array of subpellicular microtubules that are intimately associated to the overlying cell membrane. *T. cruzi* cytoskeleton proteins are tightly associated to the plasma membrane and are extracted with detergent (Pereira *et al*, 1978). Analysis of *T. brucei* plasma membrane subproteome showed the presence of 52 proteins that were common to both purified plasma membrane and cytoskeleton fractions (Bridges *et al*, 2008). Cytoskeleton components (paraflagellar rod proteins, α - and β -tubulins) are very abundant in the organelle-enriched fraction from *T. cruzi* tissue-culture trypomastigotes (Atwood *et al*, 2006). The eEF1B from several trypanosomatids may, in addition to its role in protein synthesis, protect the parasite from lipid peroxidation. For

instance, *L. major* eEF1B shows peroxidase activity against a variety of organic hydroperoxides and may be localized to the surface of the endoplasmic reticulum (Vickers *et al*, 2004). what may have facilitated the extraction of eEF1B, along with other hydrophobic proteins. We found one enolase (EAN97849.1), which, besides participating in the glycolytic pathway, has also been shown to possess antigenic properties and to be present in the cell membrane of certain invasive microorganisms such as *Plasmodium falciparum*, *Trichomonas vaginalis* and *Candida albicans* (Pal-Bhowmick *et al*, 2007).

In epimastigotes, we found 14 members of the acidic ribosomal protein family of *T. cruzi*, which are believed to complex with ribosomes and other complexes mainly by hydrophobic forces via the hydrophobic zipper or hydrophobic hook-and-eye motif (Ayub *et al*, 2005). The tertiary structure of TcP0, a *T. cruzi* acidic ribosomal protein, allows the formation of an unusual hydrophobic zipper with the ribosomal P1/P2 proteins to form the P0/P1/P2 complex (Ayub *et al*, 2005). Acidic ribosomal proteins have been identified in whole tissue-culture trypomastigote proteomic studies (Paba *et al*, 2004).

3.4.3 Functional Classification of Identified Proteins

Functional classification of proteins was based on information from the published literature and the Swiss-Prot Protein Knowledgebase (www.expasy.org/sprot/). The metacyclic and epimastigote proteins were grouped into functional categories (Tables 3.2). Although several categorization schemes could have been used, we grouped proteins that share similar functions rather than following a strict biochemical classification. Functions were assigned according to the known or putative involvement of the protein in a cellular process or pathway. Almost 71.4% of all identifications in epimastigote and metacyclic forms could be associated with five broad functional categories: 1) host-parasite relationships (virulence, adhesion to and

invasion of host cells and escape from the host immune system); 2) intermediate metabolism and transport (amino acid, lipid and carbohydrate metabolisms and bioenergetics); 3) cellular communication and signal transduction; 4) protein metabolism (folding and stabilization; modification; targeting; sorting and translocation; and proteolysis); 5) structural function and cellular dynamics (biogenesis, organization and structure of the cell); 6) stress response and cell defense (antioxidation and detoxification); and 7) transcription, translation and DNA metabolism.

Seventy-five (76.5%) of the 98 proteins in the metacyclic forms (i.e., all except the 23 hypothetical proteins) could be assigned to a cellular function (Table 3.2). The majority of metacyclic proteins falls into the host-parasite category and are GPI-anchored (39.8%) (see also Table 3.3). These are followed by proteins in the categories: intermediate metabolism and bioenergetics (11.2%); signal transduction and cellular communication (11.2%); protein metabolism (peptidolysis and proteolysis, protein folding and chaperons) (8.2%); structural function and cellular dynamics (4.1%); and cellular defense (antioxidation and detoxification) (2.0%).

The epimastigote proteins grouped into functional categories are shown in Table 3.2. One hundred and ninety-five (69.6%) of the 280 proteins identified (i.e., all except the 85 hypothetical proteins) could be assigned to a cellular function. Seventy-one proteins (25.4%) participate in intermediate metabolism: 19 in energy metabolism and oxidation-reduction pathways, 18 in amino acid metabolism, 11 in lipid metabolism, 8 in carbohydrate metabolism, and 15 are membrane transporters. Proteins involved in proteolysis, protein folding and cellular trafficking were in second place, with 48 identifications (17.1%). The other categories were ranked as follows: transcription, translation and DNA metabolism (23 proteins, 8.2%); signal transduction (20 proteins, 7.1%); structural function and cellular dynamics (17 proteins, 6.1%)

cellular defense (antioxidation and detoxification) (10 proteins, (3.6%); and host-parasite relationship (6 proteins, 2.1%).

3.4.4 Metacyclic Proteins

Thirty-nine identifications (39.8%) in metacyclic forms were of surface proteins (35 members of the TS superfamily, 2 procyclic-form surface glycoproteins, 1 mucin, 1 gp63 protease) that are involved in the host-parasite relationship (evasion from the host immune response, invasion of mammalian cells and glycosylation of parasite surface molecules) (Table 3.3). From the results of Western blot analysis, it was evident that the members of the TS superfamily (GP90 and GP82 glycoproteins) and mucins were major components of the GPI-enriched fraction (Fig. 3.2). Although mucins are highly abundant being easily detected by immunoblotting and silver staining in detergent fractions from metacyclic and epimastigote forms (data not shown), only one sequence from these glycoconjugates was identified by proteomic analysis, possibly due to extensive posttranslational modifications that may make their detection by mass spectrometry more difficult.

Analysis of metacyclic fractions resulted in the identification of 390 peptides that mapped to 35 proteins from the TS superfamily (Table 3.3), which members can be classified into four groups according to sequence similarity and function (Colli, 1993; Frasch, 2000). All the TS sequences (GP90, GP82, ASP-1, ASP-2, GP85 and Tc85) found in our study belong to group II which comprises genes encoding a number of related surface glycoproteins expressed in trypomastigotes and intracellular amastigotes (Colli, 1993; Frasch, 2000). In the clone CL Brener (El-Sayed *et al*, 2005), it was found 1430 sequences (737 genes and 693 pseudogenes) that were annotated as putative *trans*-sialidases. Twenty two proteins identified in metacyclic forms were annotated as putative *trans*-sialidases and here were further

classified as GP82, GP90, GP85, Tc85, ASP-1 and ASP-2 using sequence similarity (Azuaje *et al*, 2007). GP82, GP90 and Tc85 are adhesins that bind to specific mammalian cell receptors and/or to extracellular proteins. In metacyclic forms, among the top ten identifications were GP90 and GP82. TS glycoproteins, previously reported as specific for amastigotes (ASP-1 and ASP-2) (Low *et al*, 1997; Santos *et al*, 1997) or bloodstream trypomastigotes (Tc85-11 and GP85) (Magdesian *et al*, 2001), were present in metacyclic forms. It is possible that in metacyclic forms these proteins are synthesized in very small quantities or undergo posttranslational modifications that make their detection by conventional methods difficult.

While metacyclic forms appear to express subsets of the TS superfamily, only one TS protein was detected in epimastigotes (Table 3.3). A large number of variants of GP90 and GP82 were identified in metacyclic trypomastigotes (Tables 3.2 and 3.3), confirming that this developmental form expresses a large repertoire of surface glycoproteins involved in the host cell invasion. In contrast to African trypanosomes, *T. cruzi* needs to enter cells in order to survive, and maybe for this reason it expresses simultaneously many variant surface antigens. The presence of many variants of surface proteins could allow the parasite to adhere to different molecules on the host cell membrane and the extracellular matrix. Although GP90 and GP82 are very abundant surface proteins in metacyclic trypomastigotes, until now, only one GP90 (AAM47176) and one GP82 (AAF02477) were identified in whole metacyclic proteomic studies (Atwood *et al*, 2005; Parodi-Talice *et al*, 2007).

The peptide mapping coverage of GP90 and GP82 metacyclic proteins with the fitting of tryptic peptides from 3 independent samples each that match with the deduced GP90 and GP82 proteins is shown in Figure 3.3. The peptide mapping coverage of GP90 and GP82 was 40% and 22%, respectively (Fig. 3.3). We scanned GP82, looking for domains related to the binding of the parasite to mammalian cells or extracellular matrix proteins. The cell binding site

of GP82 is formed by the juxtaposition of two sequences, namely, P4 and P8, which contain several charged residues and are separated by a more hydrophobic stretch (Manque *et al*, 2000). We identified the peptides (LVGLLSNSASGDAWIDDYR) that covers 65% of peptide P8 (NSASGDAWIDDYRSVNAKVM) (Fig. 3.3). The peptide (NVFLYNRPL), identified in GP90 and GP82, accounts for 82% of the FLY domain (Fig. 3.3). This domain (VTVXNVFLYNR) is found at the C-terminal domain of TS proteins from group II and reported to bind to cytokeratin 18 on the surface of LLC-MK2 epithelial cells and to increase tissue culture trypomastigote entry (Magdesian *et al*, 2001).

We present the first expression evidence for the procyclic form surface glycoprotein in *T. cruzi*. This gene family was identified in *T. cruzi* genome sequencing project (El-Sayed *et al*, 2005) and their members annotated as putative procyclic form surface glycoprotein. We found five proteins (Table 3.3) that share approximately 40% identity with the procyclic form surface glycoproteins from *T. brucei* (Jackson *et al*, 1993). *T. brucei* procyclic form surface glycoprotein is a stage-specific antigen with the features of a typical transmembrane glycoprotein but with unusual cytoplasmic tail composed of a proline-rich tandem repeats (Jackson *et al*, 1993). The putative procyclic form surface glycoproteins identified in our work have 2-3 transmembrane helices and three of them contain an uncleaved signal-anchor sequence, suggesting that they are located at the cell surface.

Metacyclic trypomastigotes experience a shift in environmental conditions during their life cycle when they are transferred from triatomines to the vertebrate host, where they enter into macrophages and transform into amastigotes. Metacyclic trypomastigotes may form a complex cell envelope consisting of different types of glycoproteins as a form of protection. Many surface proteins that are directly exposed to these extreme conditions are predicted to be retained by GPI-anchors in the cell membrane. This would prevent their extraction by

hydrolytic enzymes present in the insect gut and vertebrate host. The presence of surface proteins could also allow the parasite to adhere to different molecules on the host cell membrane and the extracellular matrix.

About 11.2% of the identifications showed similarities to proteins involved in signal transduction, including multiple rab proteins, phosphatases, ras-related proteins and calcium-binding proteins (Table 3.4). The peptide mapping coverage of calcium binding protein, protein tyrosine-phosphatase and small GTP binding Rab1 protein was 46%, 35% and 27%, respectively (Table 3.4). Most of these proteins identified in epimastigote and metacyclic proteomes can be directed to the cell surface through the signal peptides and transmembrane domains or after posttranslational modifications (myristoylation, palmitoylation and prenylation). Calcium-binding proteins act specifically as calcium sensor proteins and are located in the flagellum of trypanosomes, a unique organelle that has many functions, including motility, chemotaxis, and cell signaling (Buchanan *et al*, 2005). Typical tyrosine kinase genes were not identified in the *T. cruzi* genome (El-Sayed *et al*, 2005). Protein-tyrosine phosphatases control cell differentiation in trypanosomes (Cuevas *et al*, 2005; Szöör *et al*, 2006). Two protein-tyrosine phosphatases were present in the metacyclic and epimastigote forms in our study and regulatory subunit of protein kinase was found in epimastigotes. The importance of protein phosphorylation/dephosphorylation events was corroborated by the presence of different protein phosphatases.

Eleven proteins (11.2%) identified in the metacyclic forms were implicated in the intermediate metabolism, made up as follows: 7 in energy metabolism and oxidation-reduction pathways; 2 in lipid metabolism, one in amino acid metabolism and one hexose transporter (Table 3.2). Two proton transporters were also identified in metacyclic forms, one located in the plasma membrane (proton motive P-type ATPase 2) and another in the mitochondrial inner

membrane (ATP synthase) (Brown *et al*, 2001; Luo *et al*, 2002). Both were included in the category of bioenergetic/biological oxidation proteins (Table 3.2). The presence of citrate synthase and succinate dehydrogenase is consistent with the high mitochondrial metabolism in the metacyclic forms indicated by the extraordinarily enhanced activities of metacyclic citrate synthase, isocitrate dehydrogenase, and succinate dehydrogenase (Adroher *et al*, 1988).

Metacyclic trypomastigotes produce a number of enzymes and substrates involved in antioxidant defense against the potential respiratory burst of phagocytic cells in the mammalian host. In addition to its role in protein synthesis, the translation elongation factor eEF1B identified in metacyclic forms may protect the parasite from lipid peroxidation (Vickers *et al*, 2004).

Recently, Atwood *et al*. reported a glycoproteomic analysis of plasma membrane and organelle-enriched fractions from *T. cruzi* tissue-culture trypomastigotes (Atwood *et al*, 2006). Nineteen proteins identified in this analysis were also found in our study. They were 5 structural proteins, 4 heat shock proteins, 3 hypothetical proteins (two of them having a signal peptide), 2 mitochondrial proteins (ATPase beta subunit and phosphate transporter), 2 vesicular trafficking proteins, 1 cysteine proteinase, 1 glycosomal phosphoenolpyruvate carboxykinase and 1 TS protein (EAN90192). It is of interest to note that only one TS protein was common to both studies confirming that metacyclic and tissue culture trypomastigotes express distinct repertoires of TS gene superfamily.

3.4.5 Epimastigote Proteins

There are qualitative differences between the protein profiles of epimastigotes and those of metacyclic trypomastigotes. Epimastigotes undergo biochemical and morphological changes in the the midgut of the insect vector. They multiply by binary fission and differentiate

into metacyclic trypomastigotes upon migration to the hindgut. Enzymes from intermediate metabolism pathways, including membrane transporters, are abundant in epimastigotes. Seventy-one (25.4%) of the identifications in epimastigotes were of enzymes related to energy metabolism, including those involved in the uptake and degradation of glucose via glycolysis, the pentose-phosphate shunt or the tricarboxylic acid (TCA) cycle. Several glycosomal and mitochondrial proteins were detected, including a hexose transporter. A glycosomal malate dehydrogenase and pyruvate phosphate dikinase were only detected in epimastigotes. The following members of the respiratory chain complexes were also detected in epimastigotes: succinate dehydrogenase, isocitrate dehydrogenase [NADP], Rieske iron-sulfur protein and cytochrome b5.

Epimastigotes catabolize proteins and amino acids to produce NH_3 , and glucose to produce reduced catabolites (mainly succinate and L-alanine) (Cazzulo, 1994). In this study we identified nineteen enzymes and one membrane transporter involved in the metabolism of amino acids, confirming that epimastigotes are particularly adapted to take advantage of free amino acids available in the gut of their insect vector (Atwood *et al*, 2005). Among the enzymes involved in the catabolism of amino acids were four glutamate dehydrogenases (one NADP-linked) and 7 aminotransferases, including four tyrosine aminotransferases involved in the catabolism of aromatic amino acids. Three glutamate dehydrogenase isoforms and one aminotransferase have been identified in epimastigotes submitted to nutritional stress and metacyclic trypomastigotes (Parodi-Talice *et al*, 2007). Glutamate dehydrogenase provides a link between carbohydrate and amino acid metabolism in *T. cruzi* (Barderi *et al*, 1998). This enzyme may play an important role during the transit of parasites in the insect gut, where they are deprived of carbohydrates and free amino acids derived from the hemoglobin degradation are available. The presence of urocanate hydratase, an enzyme that participates in the

transformation of histidine to glutamate, is noteworthy. Genes encoding the enzymes of this pathway, which are abundant in the *T. cruzi* insect stages but nearly undetectable in the mammalian stages (Atwood *et al*, 2005), were not found in *T. brucei* or *L. major*, suggesting that they are species-specific for *T. cruzi*.

Eleven enzymes involved in lipid metabolism were present in the epimastigote fractions. Among them were seven from the fatty acid beta-oxidation pathway and three from the ergosterol biosynthesis pathway. Two enzymes (3,2-trans-enoyl-CoA isomerase and acyl-CoA dehydrogenase) were located in the mitochondria. In trypanosomes and *Leishmania*, fatty acid beta-oxidation can occur in two separate cellular compartments: glycosomes and mitochondria.

As expected, proteins involved in stress response and cell defense were detected in epimastigotes: the cytosol-localized trypanothione peroxidase; trypanothione, the substrate for trypanothione peroxidase; thiol transferase Tc52; the GPR1/FUN34/yaaH family; carbonic anhydrase-like protein and a dehydrogenase named TcOYE (old yellow enzyme). The expression of these proteins by epimastigotes is consistent with a preadaptation of metacyclic forms to resist the potential respiratory burst of phagocytic cells in the mammalian host. Thiol transferase Tc52 and trypanothione peroxidase have been detected in nutritionally stressed and differentiating (adherent) epimastigotes and tissue culture trypomastigotes (Paba *et al*, 2004; Parodi-Talice *et al*, 2007). It has been suggested that in vitro-induced benznidazole resistance in *T. cruzi* is correlated with deletion of copies of the TcOYE gene (Andrade *et al*, 2008; Murta *et al*, 2006).

Forty-eight (17.1%) epimastigote identifications were associated with protein metabolism, including proteolysis, folding and stabilization, modification, targeting, sorting and translocation (Table 3.2). Several heat shock proteins (HSPs 85, 70 and 60; glucose regulated

78 protein; mitochondrial chaperonin HSP60) were found in epimastigotes. They were also identified in whole extract and organelle-enriched fractions from tissue-culture trypomastigotes (Atwood *et al*, 2006; Paba *et al*, 2004). Proteinases and peptidases represented 7.5% of the proteins identified in epimastigotes, while chaperones and proteins involved in transporting, targeting, sorting and translocation represented 7.1% and 2.5%, respectively. We have confirmed the presence of a large family of calpain-related proteins, including a calpain-like cytoskeleton-associated protein (CAP5.5). The presence of calpain-like proteins in trypanosomes was first suggested by analysis of a cytoskeleton-associated protein in *T. brucei* (CAP5.5) that showed similarity to the catalytic region of calpain-type proteases (Hertz-Fowler *et al*, 2001). CAP5.5 has been shown to be both myristoylated and palmitoylated, suggesting a stable interaction with the cell membrane (Hertz-Fowler *et al*, 2001). Bioinformatics analysis of the *T. cruzi* genome database confirmed the presence of calpain-like proteins in this parasite (Ersfeld *et al*, 2005). Most of them contain a well-conserved protease domain and a novel N-terminal sequence motif unique to kinetoplastids. Furthermore, they contain N-terminal fatty acid acylation motifs, indicating association of these proteins with cellular membranes. Proteins involved in the transport machinery and cellular trafficking were identified and can be expected to facilitate studies of these less-defined processes in *T. cruzi*.

As discussed above, we were not surprised by the presence of cytoskeletal proteins in this survey since these proteins have been reported to be associated to the plasma membrane of trypanosomes (Atwood *et al*, 2006; Bridges *et al*, 2008; Pereira *et al*, 1978). We identified a lysosomal membrane protein (p67, EAN81944.1) similar to the mammalian LAMP (Kelley *et al*, 1999) and a Golgi- and lysosome-associated transmembrane glycoprotein previously described in African trypanosomes (Lingnau *et al*, 1999). Gim5A is one of the most abundant glycosomal membrane proteins in *T. brucei* and is essential for the survival of bloodstream

forms (Maier *et al*, 2001). We tentatively assigned *T. cruzi* GIM5 to the category of structural proteins and cellular dynamics (biogenesis, organization and structure of the cell). Gim5A and vesicle-associated protein (EAN93210) were also detected in the organelle-enriched fraction from tissue-culture trypomastigotes (Atwood *et al*, 2006).

Also noticed is the high proportion of ribosomal and RNA binding proteins and transcription and translation factors in the epimastigote proteome, which is consistent with the replication status of this developmental form. None of these proteins were detected in the nonreplicating metacyclic trypomastigotes.

3.4.6 Hypothetical proteins

Of the 378 *T. cruzi* proteins identified in our analysis, 108 are from genes annotated as hypothetical, confirming these as bona fide genes (Table 3.2). On the other hand, genes annotated as hypothetical in *T. cruzi* were difficult to assess as no functional insights could be gained from our findings. However, most hypothetical genes have homologues in *Leishmania major* and *T. brucei*, suggesting a common function for a given gene in closely related trypanosomatids.

3.5 Conclusions

In summary, we have shown in this paper that each of the insect stages of *T. cruzi* reveals distinct protein datasets when solubilized with neutral detergents. While metacyclic trypomastigotes have a higher number of potential GPI-anchored proteins that might be involved in the process of attachment and invasion of host cells, the proteins found in epimastigote samples were related to metabolism, a finding that is in agreement with the high energy requirements of this proliferative stage of the parasite. Our data show that metacyclic trypomastigotes express a large repertoire of surface glycoproteins involved in the host cell invasion. In addition, we speculate that the latter proteins are present in metabolism-related organelles, such as mitochondria, glycosomes and reservosomes. We hope that in future the results of this study will contribute to the design of more rational therapies against Chagas' disease.

3.6 Acknowledgments

This work was supported by grants from FAPESP and CNPq (Brazil) to JFS and from NIH grants 5G12RR008124 (to BBRC/Biology/UTEP), and 5S06GM08012-37 and RO1AI070655 (to ICA). ECV and LGG were awarded postdoctoral and doctoral fellowships, respectively, by CNPq. ESN was partially supported by the Georges A. Krutilek memorial scholarship from the Graduate School (UTEP). We would like to thank A. Galetovic for the generous gift of the polyclonal antiserum against *T. cruzi* calpain and Dr. Sergio Schenkman and Dr. David Engman for the monoclonal antibody to the *T. cruzi* flagellar calcium-binding protein.

3.7 References

Adroher FJ, Osuna A, Lupianez JA (1988) Differential energetic metabolism during *Trypanosoma cruzi* differentiation. I. Citrate synthase, NADP-isocitrate dehydrogenase, and succinate dehydrogenase. *Arch Biochem Biophys* **267**: 252-261.

Almeida IC, Gazzinelli R, Ferguson MA, Travassos LR (1999) *Trypanosoma cruzi* mucins: potential functions of a complex structure. *Mem Inst Oswaldo Cruz* **94 Suppl 1**: 173-176.

Andrade HM, Murta SM, Chapeaurouge A, Perales J, Nirde P, Romanha AJ (2008) Proteomic analysis of *Trypanosoma cruzi* resistance to Benznidazole. *J Proteome Res* **7**: 2357-2367.

Añez-Rojas N, García-Lugo P, Crisante G, Rojas A, Añez N (2006) Isolation, purification and characterization of GPI-anchored membrane proteins from *Trypanosoma rangeli* and *Trypanosoma cruzi*. *Acta Trop* **97**: 140-145.

Araya JE, Cano MI, Yoshida N, da Silveira JF (1994) Cloning and characterization of a gene for the stage-specific 82-kDa surface antigen of metacyclic trypomastigotes of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **65**: 161-169.

Atwood JA, 3rd, Minning T, Ludolf F, Nuccio A, Weatherly DB, Alvarez-Manilla G, Tarleton R, Orlando R (2006) Glycoproteomics of *Trypanosoma cruzi* trypomastigotes using subcellular fractionation, lectin affinity, and stable isotope labeling. *J Proteome Res* **5**: 3376-3384.

Atwood JA, 3rd, Weatherly DB, Minning TA, Bundy B, Cavola C, Oppenheimer FR, Orlando R, Tarleton RL (2005) The *Trypanosoma cruzi* proteome. *Science* **309**: 473-476.

Ayub MJ, Barroso JA, Levin MJ, Aguilar CF (2005) Preliminary structural studies of the hydrophobic ribosomal P0 protein from *Trypanosoma cruzi*, a part of the P0/P1/P2 complex. *Protein Pept Lett* **12**: 521-525.

Azuaje F, Ramirez JL, Da Silveira JF (2007) An exploration of the genetic robustness landscape of surface protein families in the human protozoan parasite *Trypanosoma cruzi*. *IEEE Trans Nanobioscience* **6**: 223-228.

Barderis P, Campetella O, Frasch AC, Santome JA, Hellman U, Pettersson U, Cazzulo JJ (1998) The NADP⁺-linked glutamate dehydrogenase from *Trypanosoma cruzi*: sequence, genomic organization and expression. *Biochem J* **330 (Pt 2)**: 951-958.

Bridges DJ, Pitt AR, Hanrahan O, Brennan K, Voorheis HP, Herzyk P, de Koning HP, Burchmore RJ (2008) Characterisation of the plasma membrane subproteome of bloodstream form *Trypanosoma brucei*. *Proteomics* **8**: 83-99.

Brown BS, Stanislawski A, Perry QL, Williams N (2001) Cloning and characterization of the subunits comprising the catalytic core of the *Trypanosoma brucei* mitochondrial ATP synthase. *Mol Biochem Parasitol* **113**: 289-301.

Buchanan KT, Ames JB, Asfaw SH, Wingard JN, Olson CL, Campana PT, Araujo AP, Engman DM (2005) A flagellum-specific calcium sensor. *J Biol Chem* **280**: 40104-40111.

Cazzulo JJ (1994) Intermediate metabolism in *Trypanosoma cruzi*. *J Bioenerg Biomembr* **26**: 157-165.

Colinge J, Masselot A, Giron M, Dessingy T, Magnin J (2003) OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics* **3**: 1454-1463.

Colli W (1993) Trans-sialidase: a unique enzyme activity discovered in the protozoan *Trypanosoma cruzi*. *Faseb J* **7**: 1257-1264.

Colli W, Alves MJ (1999) Relevant glycoconjugates on the surface of *Trypanosoma cruzi*. *Mem Inst Oswaldo Cruz* **94 Suppl 1**: 37-49.

Cortez M, Neira I, Ferreira D, Luquetti AO, Rassi A, Atayde VD, Yoshida N (2003) Infection by *Trypanosoma cruzi* metacyclic forms deficient in gp82 but expressing a related surface molecule, gp30. *Infect Immun* **71**: 6184-6191.

Cuevas IC, Rohloff P, Sanchez DO, Docampo R (2005) Characterization of farnesylated protein tyrosine phosphatase TcPRL-1 from *Trypanosoma cruzi*. *Eukaryot Cell* **4**: 1550-1561.

de Almeida ML, Heise N (1993) Proteins anchored via glycosylphosphatidylinositol and solubilizing phospholipases in *Trypanosoma cruzi*. *Biol Res* **26**: 285-312.

do Carmo MS, dos Santos MR, Cano MI, Araya JE, Yoshida N, da Silveira JF (2002) Expression and genome-wide distribution of the gene family encoding a 90 kDa surface glycoprotein of metacyclic trypomastigotes of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **125**: 201-206.

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Wortley 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, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**: 409-415.

Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**: 976-989.

Ersfeld K, Barraclough H, Gull K (2005) Evolutionary relationships and protein domain architecture in an expanded calpain superfamily in kinetoplastid parasites. *J Mol Evol* **61**: 742-757.

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

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

Güther ML, de Almeida ML, Yoshida N, Ferguson MA (1992) Structural studies on the glycosylphosphatidylinositol membrane anchor of *Trypanosoma cruzi* 1G7-antigen. The structure of the glycan core. *J Biol Chem* **267**: 6820-6828.

Hertz-Fowler C, Ersfeld K, Gull K (2001) CAP5.5, a life-cycle-regulated, cytoskeleton-associated protein is a member of a novel family of calpain-related proteins in *Trypanosoma brucei*. *Mol Biochem Parasitol* **116**: 25-34.

Jackson DG, Smith DK, Luo C, Elliott JF (1993) Cloning of a novel surface antigen from the insect stages of *Trypanosoma brucei* by expression in COS cells. *J Biol Chem* **268**: 1894-1900.

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.

Kelley RJ, Alexander DL, Cowan C, Balber AE, Bangs JD (1999) Molecular cloning of p67, a lysosomal membrane glycoprotein from *Trypanosoma brucei*. *Mol Biochem Parasitol* **98**: 17-28.

Ko YG, Thompson GA, Jr. (1995) Purification of glycosylphosphatidylinositol-anchored proteins by modified triton X-114 partitioning and preparative gel electrophoresis. *Anal Biochem* **224**: 166-172.

Lingnau A, Zufferey R, Lingnau M, Russell DG (1999) Characterization of tGLP-1, a Golgi and lysosome-associated, transmembrane glycoprotein of African trypanosomes. *J Cell Sci* **112 Pt 18**: 3061-3070.

Low HP, Tarleton RL (1997) Molecular cloning of the gene encoding the 83 kDa amastigote surface protein and its identification as a member of the *Trypanosoma cruzi* sialidase superfamily. *Mol Biochem Parasitol* **88**: 137-149.

Luo S, Scott DA, Docampo R (2002) *Trypanosoma cruzi* H⁺-ATPase 1 (*TcHA1*) and 2 (*TcHA2*) genes complement yeast mutants defective in H⁺ pumps and encode plasma membrane P-type H⁺-ATPases with different enzymatic properties. *J Biol Chem* **277**: 44497-44506.

Magdesian MH, Giordano R, Ulrich H, Juliano MA, Juliano L, Schumacher RI, Colli W, Alves MJ (2001) Infection by *Trypanosoma cruzi*. Identification of a parasite ligand and its host cell receptor. *J Biol Chem* **276**: 19382-19389.

Maier A, Lorenz P, Voncken F, Clayton C (2001) An essential dimeric membrane protein of trypanosome glycosomes. *Mol Microbiol* **39**: 1443-1451.

Manque PM, Eichinger D, Juliano MA, Juliano L, Araya JE, Yoshida N (2000) Characterization of the cell adhesion site of *Trypanosoma cruzi* metacyclic stage surface glycoprotein gp82. *Infect Immun* **68**: 478-484.

Murta SM, Krieger MA, Montenegro LR, Campos FF, Probst CM, Avila AR, Muto NH, de Oliveira RC, Nunes LR, Nirde P, Bruna-Romero O, Goldenberg S, Romanha AJ (2006) Deletion of copies of the gene encoding old yellow enzyme (TcOYE), a NAD(P)H flavin oxidoreductase, associates with in vitro-induced benznidazole resistance in *Trypanosoma cruzi*. *Mol Biochem Parasitol* **146**: 151-162.

Paba J, Ricart CA, Fontes W, Santana JM, Teixeira AR, Marchese J, Williamson B, Hunt T, Karger BL, Sousa MV (2004) Proteomic analysis of *Trypanosoma cruzi* developmental stages using isotope-coded affinity tag reagents. *J Proteome Res* **3**: 517-524.

Pal-Bhowmick I, Mehta M, Coppens I, Sharma S, Jarori GK (2007) Protective properties and surface localization of *Plasmodium falciparum* enolase. *Infect Immun* **75**: 5500-5508.

Parodi-Talice A, Monteiro-Goes V, Arrambide N, Avila AR, Duran R, Correa A, Dallagiovanna B, Cayota A, Krieger M, Goldenberg S, Robello C (2007) Proteomic analysis of metacyclic trypomastigotes undergoing *Trypanosoma cruzi* metacyclogenesis. *J Mass Spectrom* **42**: 1422-1432.

Pereira NM, Timm SL, da Costa SC, Rebello MA, de Souza W (1978) *Trypanosoma cruzi*: isolation and characterization of membrane and flagellar fractions. *Exp Parasitol* **46**: 225-234.

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.

Santos MA, Garg N, Tarleton RL (1997) The identification and molecular characterization of *Trypanosoma cruzi* amastigote surface protein-1, a member of the trans-sialidase gene super-family. *Mol Biochem Parasitol* **86**: 1-11.

Stone KL, Williams KR (1996) Enzymatic digestion of proteins in solution and in SDS polyacrilamide gel. In *The protein protocol handbook.*, Walker JM (ed). Totowa, NJ: Humana Press Inc.

Szöör B, Wilson J, McElhinney H, Tabernero L, Matthews KR (2006) Protein tyrosine phosphatase *TbPTP1*: A molecular switch controlling life cycle differentiation in trypanosomes. *J Cell Biol* **175**: 293-303.

Teixeira MM, Yoshida N (1986) Stage-specific surface antigens of metacyclic trypomastigotes of *Trypanosoma cruzi* identified by monoclonal antibodies. *Mol Biochem Parasitol* **18**: 271-282.

Vickers TJ, Wyllie S, Fairlamb AH (2004) *Leishmania major* elongation factor 1B complex has trypanothione S-transferase and peroxidase activity. *J Biol Chem* **279**: 49003-49009.

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

Table 3.1 Predicted Posttranslational Modifications of *T. cruzi* Proteins Extracted with Triton X-114

Modification	Metacyclic Trypomastigote N (%)^a	Epimastigote N (%)^a
GPI Anchor	21 (21.4%)	1 (0.4%)
Signal Peptide ^b	26 (26.5%)	72 (25.7%)
Myristoylation and Prenylation	15 (15.3%)	21 (7.5%)
Mitochondrial Targeting	16 (16.3%)	49 (17.5%)
Trans-Membranic Domain	4 (4.1%)	20 (7.1)
None ^c	16 (16.3%)	117 (41.8)
Total	98 (100%)	280 (100%)

^a Number and percentage of proteins.

^b Proteins contained a typical signal peptide motif or an uncleaved signal-anchor sequence were included.

^c Proteins with no apparent posttranslational modification.

Table 3.2 Functional Classification of *T. cruzi* Proteins Extracted with Triton X-114

Functional Category	Metacyclic Trypomastigote N (%) ^a	Epimastigote N (%) ^a
Host-Parasite Relationship (surface proteins)	39 (39.8%)	6 (2.1%)
Cellular Communication and Signal Transduction	11 (11.2%)	20 (7.1%)
Intermediate Metabolism	11 (11.2%)	71 (25.4%)
Bioenergetic and Biological Oxidations	7	19
Lipid Metabolism	2	11
Amino Acid Metabolism	1	18
Carbohydrate Metabolism	0	8
Transporters	1	15
Protein Metabolism	8 (8.2%)	48 (17.1%)
Peptidolysis	6	21
Protein Folding	2	20
Vesicular Trafficking	0	7
Structural Proteins and Cellular Dynamics	4 (4.1%)	17 (6.1%)
Stress Response and Cell Defense	2 (2.0%)	10 (3.6%)
Transcription, Translation and DNA Repair	0	23 (8.2%)
Hypothetical Proteins ^b	23 (23.5%)	85 (30.4%)
Total	98 (100%)	280 (100%)

^a Number and percentage of proteins.

^b Hypothetical proteins conserved in TriTryps.

Table 3.3 Surface Proteins Identified in Metacyclic Trypomastigotes and Epimastigotes

Accession Number ^a	Protein Name ^b	Coverage	Number of Identified Peptides	Developmental Stage ^c
AAM47176.1	surface glycoprotein GP90	40%	82	MT
EAN87246.1	trans-sialidase, putative (GP90)	23%	14	MT
ABR19835.1	stage-specific surface protein GP82	22%	31	MT
BAF74647.1	glycoprotein 82 kDa	18%	20	MT
BAF74641.1	glycoprotein 82 kDa	14%	18	MT
EAN81650.1	trans-sialidase, putative (GP90)	13%	24	MT
EAN91385.1	trans-sialidase, putative (GP90)	10%	10	MT
EAN99311.1	trans-sialidase, putative (GP90)	10%	3	MT
ABR19836.1	stage-specific surface protein GP82	9%	2	MT
EAN95004.1	trans-sialidase, putative (GP90)	9%	20	MT
AAP22091.1	GP63 group I member b protein	9%	3	MT
EAN98599.1	trans-sialidase, putative (ASP-2)	9%	2	MT
EAN82035.1	procyclic form surface glycoprotein, putative	9%	3	E
EAN95001.1	trans-sialidase, putative (GP90)	8%	14	MT
AAA18827.1	surface glycoprotein (GP85)	8%	6	MT
EAN82322.1	trans-sialidase, putative (Tc-85/16)	8%	15	MT
EAN93322.1	trans-sialidase, putative (GP85-TSA-E2)	8%	5	MT
EAN95642.1	trans-sialidase, putative (GP85-TSA-E3)	8%	16	MT
EAN96370.1	trans-sialidase, putative (Tc-85/16)	8%	16	MT
EAN98005.1	procyclic form surface glycoprotein, putative	7%	2	MT, E
EAN93266.1	trans-sialidase, putative (Tc85-11)	6%	14	MT

AAB18265.1	surface protein-1, ASP-1	6%	6	MT
EAN82837.1	trans-sialidase, putative (GP82)	6%	4	MT
AAD10620.1	surface glycoprotein, GP85	6%	5	MT
AAM47178.1	surface glycoprotein GP90	6%	1	MT
EAN88221.1	trans-sialidase, putative (GP90)	6%	8	MT
EAN98004.1	procyclic form surface glycoprotein, putative	5%	1	E
EAN93978.1	mucin TcMUCII, putative	5%	1	MT
EAN88154.1	trans-sialidase, putative (GP82)	5%	6	MT
ABQ53588.1	surface protein-2, ASP-2	5%	3	MT
EAN82390.1	trans-sialidase, putative (ASP-2)	5%	5	MT
EAN90192.1	trans-sialidase, putative (GP90)	5%	10	MT
BAF74652.1	glycoprotein 82 kDa	5%	3	MT
A48458	gp85/sialidase homolog	4%	1	MT
EAN84247.1	extracellular receptor, putative	4%	1	E
EAN86738.1	trans-sialidase, putative (GP85-SA85-1.2)	4%	4	MT
AAC47720.1	amastigote surface protein-2, ASP-2	3%	2	MT
AAD13347.1	surface glycoprotein Tc85-11	3%	1	E
EAN91734.1	trans-sialidase, putative (GP90)	3%	1	MT
EAN93068.1	trans-sialidase, putative (GP85-TSA-1)	3%	3	MT
EAN98006.1	procyclic form surface glycoprotein, putative	2%	1	MT
EAN82034.1	procyclic form surface glycoprotein, putative	2%	1	E
ABO28970.1	trans-sialidase-like protein (GP82)	2%	2	MT
EAN85322.1	trans-sialidase, putative (ASP-2)	2%	1	MT

^a GenBank accession number.

^b Sequences annotated as putative trans-sialidases (TS) in the *T. cruzi* genome-sequencing project were assigned to the families (indicated in the parentheses) of TS superfamily according to Azuaje²⁹.

^c MT, metacyclic trypomastigote; E, epimastigote

Table 3.4 Cellular Communication and Signal Transduction Proteins Identified in Metacyclic Trypomastigotes and Epimastigotes

Accession Number ^a	Protein Name	Coverage	Number of Identified Peptides	Developmental Stage ^b
CAA90898.1	F29 (flagellar calcium-binding protein)	46%	34	MT, E
EAN94827.1	protein tyrosine phosphatase, putative	35%	9	MT, E
AAA99985.1	calcium-binding protein	34%	26	MT, E
EAN92885.1	membrane-bound acid phosphatase, putative	29%	6	E
AAB08762.1	24 kDa flagellar calcium-binding protein	28%	20	MT
BAA13411.1	calcium-binding protein	28%	8	MT, E
EAN85763.1	small GTP-binding protein Rab1, putative	27%	9	MT, E
AAT77755.1	14-3-3 protein	17%	3	E
AAT77756.1	14-3-3 protein; Tcf2p	16%	4	MT, E
EAN99708.1	protein phosphatase, putative	15%	6	E
EAN94492.1	protein tyrosine phosphatase, putative	14%	3	MT, E
EAN95123.1	14-3-3 protein, putative	13%	2	E
EAN85068.1	membrane-bound acid phosphatase, putative	10%	2	E
EAN94164.1	14-3-3 protein, putative	10%	3	E
EAN99768.1	rab7 GTP binding protein, putative	10%	3	M
EAN86478.1	membrane-bound acid phosphatase, putative	9%	2	E
EAN86455.1	calcium-binding protein, putative	8%	30	M
EAN88612.1	small GTP-binding protein Rab11, putative	7%	2	E
EAN92627.1	14-3-3 protein, putative	7%	5	E
EAN92863.1	phosphatidic acid phosphatase protein, putative	7%	2	E

AAM91818.1	flagellar calcium binding protein 3	7%	1	MT
EAN87884.1	ras-related protein rab-2a, putative	5%	1	E
EAN93719.1	membrane-bound acid phosphatase 2, putative	4%	3	E
EAN88194.1	regulatory subunit of protein kinase a-like protein	2%	1	E

^a GenBank accession number.

^b MT, metacyclic trypomastigote; E, epimastigote.

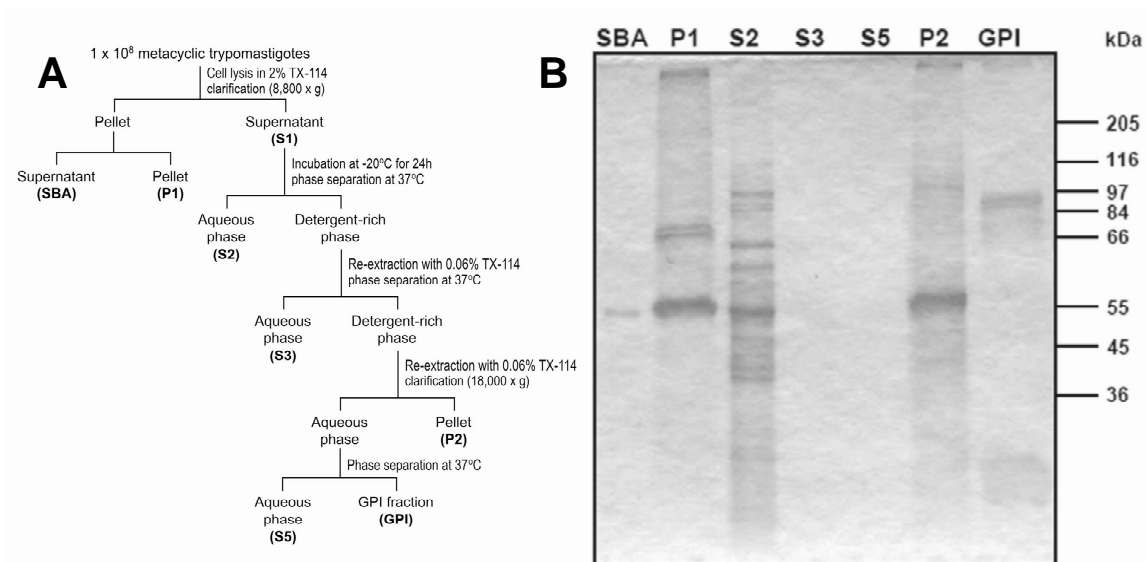


Figure 3.1 (A) Schematic representation of the extraction protocol for membrane proteins. Metacyclic trypomastigotes (1×10^8) were lysed in TBS containing 2% Triton X-114 (TX-114). The homogenate was clarified by centrifugation and the supernatant stored at 20 °C for 24 h. The supernatant was then submitted to phase separation with TX-114 at 37 °C. The final detergent-rich extract corresponds to the GPI-anchored protein-enriched fraction. **(B)** Analysis by SDS-PAGE and Coomassie staining of molecules obtained in the purification steps of GPI-anchored proteins. SBA, supernatant from extraction with buffer A; P1, Pellet 1; S2, supernatant 2; S3, supernatant 3; S5, supernatant 5; P2, pellet 2; GPI, GPI-fraction. The molecular masses are indicated on the right.

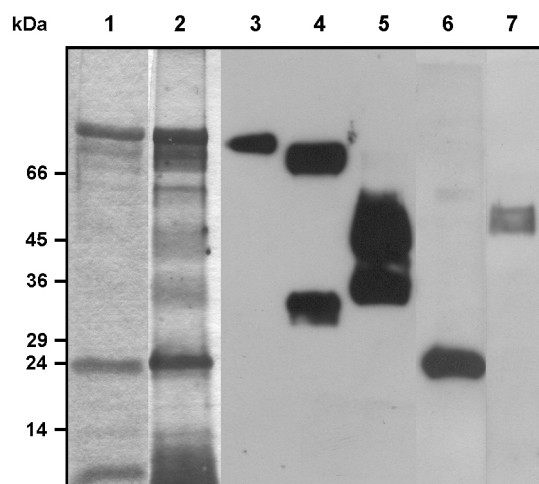


Figure 3.2 Gel electrophoresis and immunoblotting of Triton X-114-soluble membrane proteins from metacyclic trypomastigotes. Lanes 1 and 2, SDS-PAGE. Proteins were separated by SDS-PAGE on 10%-Laemmli gels and stained with Coomassie blue (lane 1) or silver (lane 2). Lanes 3 to 7, immunoblotting with monoclonal antibodies (MoAb) against *T. cruzi* GPI-anchored surface proteins and flagellar calcium-binding protein (FCBP) (lanes 3 to 6) and with a monospecific polyclonal antiserum against the catalytic domain of *T. cruzi* calpain-like cysteine proteinase (lane 7). The MoAbs used were: MoAb 1G7 against GP90 surface glycoprotein (lane 3); MoAb 3F6 against GP82 surface glycoprotein (lane 4); MoAb 10D8 against the GP35/35 kDa mucin complex (lane 5); MoAb 25 against the *T. cruzi* FCBP (lane 6).

```

>gi|21321389|gb|AAM47176.1|AF426132_1 surface glycoprotein GP90 [Trypanosoma cruzi]
- coverage - 279/726 aa = 38.42%
YIDKLGEKSANNHCALWKDQQQHTHRPHTHIYMLSRVASVKALRTHNRRRVTLFSGRRREGRESERQRPNMSRRVFASAVLLL
LFVMMCCSGGTSNAVKSSNGNAQLPHAVDLFLPNQTLVVPKGGTSQETTREAFTSPPLVSAGGVIAAFPEGRVYVEYVVG
SIETNSSDVVAEYIDATWDWSALVGKVSSESKWKAYTVLGPTDGTDNRVGFFYHPTTTTKGNKVFLLVGSLGELKESGGRRTD
NLGLKLVVGDVREPTDSEPTGGITWGEIKSPLKQSTIDAHEVKLTTELFAAGGSSILMEDGTLVFPLMANSGNGDFYSMIYISK
EDGENCLLSTGVSPAKCYHPRIPWKGSLLMIVDCEDEQNVYESRDMGTAWTEAIGTLPGVVVKSSQFFSDLKLRVDALITPT
IEGRKVMLYTQRGNSSEKNKPNPLYLWVTDNNSLSVGPVGMNAGKGELSSSLYSDGNLQLLQQRGNGECSAISLSRLTEE
LKTIKSVLKTWAQKDAFFSNLSIPTAGLVAVLSDAANKDTWNDEYLCVNATVKNKVEDGLKLTESDSEVMWPVNTRVNNVR
HVSLSHNFTLVASVTIEEAPSADAPLMGAMLGDTNSQYTMGVLYTADKEWVTIFNGKKTTESGTWEPGKEYQVALMLQGNKSS
VYVDGKSLGEEELPLQSERPLEYLSFCFGGCGIKNFPVTVKNVFLYNRPLNPTEMTAIKYKG

>gi|148943290|gb|ABR19835.1| stage-specific surface protein GP82 [Trypanosoma
cruzi] - coverage - 161/726 aa = 21.18%
MLSRVAAMAPRTHNRRRVGTSSSGRRREGGESERQRPNMSRRVFDSTILLLLVTMCCDTCGAAAANKENDGKSDLRSAEELQW
VNLFPVQTTTPVLPEGGTTPGTRDAFVSPSLVSAGGVLAFAFARGEIDAQYAVDGLIKPTSSAVVAEYIDS*SWDWF*TLVEKVS
ESTWKAYTVLSKAEGKGNLDVVLSP*TTMKGNKVFLLVGSYDMLNESGIWKRDPDLKLVVGEVTKPSAGGEPGSGWITWGTP
SLNQTTLKIPKAGLKDFYSSGGSGVMEDGTIVFPVIAFNAGNAGFSTTIYSTDDGANWMLSNGTPPAECLEPRITEWEGSLP
MIVDCVDGQ*RVYESRDMGTWTEAVGTLSGVWAKSSQFFRDLNLRVDALIAATIEGRKVMLYTQ*RGYASGEKRVNPLYLWVTD
NNRSFYFGPIAMGNAANSFMFVSSLLYSDGSLHLLQRRANDKGSVISLARLTEELKTIKSVLSTWSKLDASFSASSTPTAGLVG
LLSNSASGDWIDDYRSVNAKVMNAVKVHDGFKFTGFGSGAIWPVNNRESNGPHTFVNYNFTLVATVIVHKVPKNSTTLLGAV
LAEPIS*TLFIGLSYGTDTGTWETVFNGETTTSGSTWMPGKEYQVALMLQDGNKGSVYVDGMSVGS*LATLPTPEVRGA*EIADFYF
VGGEDEEDKKSSSVTVKNVFLYNRPLGADEL*RMVKKIDGSMHGGVSRALLLLGLCGFAALY

```

Figure 3.3 Peptide mapping coverage of GP90 and GP82. Representative fitting of tryptic peptides (shaded in gray) from 3 independent samples of GPI extracts from metacyclic trypomastigotes that match with the deduced proteins GP90 (AAM47176) and GP82 (ABR19835) are shown. Potential initiator methionines are indicated by asterisks. Underlining indicates the following motifs: signal anchor or uncleaved signal peptide (positions 39-63 in GP90 and GP82); P4 and P8 cell binding sites of GP82 (positions 462-481 and 502-521, respectively); FLY domain (positions 670-680 and 678-688 in GP90 and GP82, respectively). The hydrophobic C-terminal region of GP82, characteristic of GPI-anchored proteins, is indicated in italic (positions 702-726). The coverage percentage for GP90 and GP82 is 40% and 22%, respectively.

Chapter 4: Improved proteomic approach for the discovery of potential vaccine targets in *Trypanosoma cruzi*

Ernesto S. Nakayasu¹, Tiago J. Sobreira², Rafael Torres Jr.¹, Luciane Ganiko¹, Paulo S.L. Oliveira², Alexandre F. Marques¹, and Igor C. Almeida¹

¹Department of Biological Sciences, University of Texas at El Paso, El Paso TX, USA.

²Instituto do Coração, Universidade de São Paulo, São Paulo SP, Brazil.

Correspondence: Dr. Igor C. Almeida. Department of Biological Sciences, University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968 USA. Telephone: +1 (915) 747-6086. Fax: +1 (915) 747-5808. E-mail: icalmeida@utep.edu

Abbreviations: **FA**, formic acid; **GPI**, glycosylphosphatidylinositol; **SCX**, strong-cation exchange;

Key words: antigens / Chagas disease/ multidimensional separation / surface proteins / *Trypanosoma cruzi*

4.1 Abstract

Chagas disease caused by *Trypanosoma cruzi* is a devastating infectious disease with millions of cases in Latin America, and recently became a public health concern in United States and Europe. Although many efforts have been made for the development of an effective immunotherapy, currently there is no human vaccine for Chagas disease. One restriction for the rational approach to develop a *T. cruzi* vaccine is the limited information about the proteins expressed by different phylogenetic lineages, strains, and stages of the parasite. Here, we developed a new approach to perform two-dimensional liquid chromatography coupled to tandem mass spectrometry. The application of this technique to the analysis of the proteome of the infective trypomastigote stage of *T. cruzi* led to the identification of 1448 non-redundant proteins (6154 individual proteins). Furthermore, about 14% of the identified sequences comprise surface proteins, most of them related to parasite pathogenesis. Bioinformatic analysis of these predicted surface proteins revealed more than one hundred potential sequences for major histocompatibility (MHC) class I, and several thousands for MHC class II. We discuss the potential use of these epitopes for vaccine development and possible mechanisms of immune evasion by *T. cruzi* trypomastigotes.

4.2 Introduction

Chagas disease or American trypanosomiasis is considered a neglected infectious disease with an estimate of 11 million of cases in Latin America, and about 50,000 deaths annually (Barrett *et al*, 2003; Dias *et al*, 2002; Stuart *et al*, 2008). Chagas disease has become a public health concern in the U.S. and Europe due to the number of asymptomatic infected people migrating from endemic areas (Bern *et al*, 2007; Piron *et al*, 2008). In addition, autochthonous cases have already been reported in the U.S. (Bern *et al*, 2007; Tarleton *et al*, 2007). The etiologic agent of Chagas disease is the protozoan parasite *Trypanosoma cruzi*, which has four stages, namely epimastigotes, metacyclic trypomastigotes, amastigotes, and bloodstream trypomastigotes in two distinct hosts, namely the triatomine insect and mammal (Barrett *et al*, 2003; Stuart *et al*, 2008). In the natural cycle, bloodstream trypomastigotes of the parasite are ingested by the triatomine insect vector (kissing bug) during a bloodmeal. In the midgut of the insect, trypomastigotes differentiate into epimastigotes, which reproduce by binary fission. In the posterior regions of the gut, upon nutritional stress epimastigotes are transformed into metacyclic trypomastigotes, which are then excreted with the feces during a bloodmeal. The parasite can penetrate into mammalian host bloodstream through the bite wound or exposed mucosas and infect a variety of nucleated cells. Inside the cells, the metacyclic trypomastigote escapes the parasitophorous vacuole and transforms into amastigote form, which replicates by asexual binary division. After several divisions, amastigotes differentiate into trypomastigotes, which are then released to the extracellular milieu and therefrom to the bloodstream to infect new cells or be uptaken by the kissing bug, completing the life cycle (Barrett *et al*, 2003; Stuart *et al*, 2008).

Current therapy against Chagas disease is based only in two drugs: benznidazole (Rochagan, Roche) and nifurtimox (Lampit, Bayer). Both these compounds have variable

efficacy and severe side effects (Urbina and Docampo, 2003; Wilkinson *et al*, 2008). On the other hand, there is no available vaccine for the treatment or prevention of Chagas disease (Dumonteil, 2007; Garg and Bhatia, 2005). Several studies have focused in the host immune response against *T. cruzi* (Tarleton, 2007). Although some promising progress have been made, the difficulty of using a more rational approach to develop a vaccine against *T. cruzi* is the limited information about proteins expressed on the cell surface of different parasite stages, strains, and phylogenetic lineages.

T. cruzi is covered by a thick later of glycosylphosphatidylinositol (GPI)-anchored molecules, which are encoded by multigene families, such as *trans*-sialidase (TS)/gp85, mucins, mucin-associated surface proteins (MASP), and gp63, each one of them with hundreds to more than 1,000 members (Acosta-Serrano *et al*, 2007; El-Sayed *et al*, 2005). However, the lack of positive correlation between the levels of transcripts and proteins makes difficult to use approaches such as microarray and cDNA library to map the expressed sequences (Jager *et al*, 2007). Thus, the use of proteomic approaches is an interesting and reliable alternative to analyze gene expression in *T. cruzi*. A large-scale proteomic study from Brazil strain of *T. cruzi* identified 1168 non-redundant proteins (2784 individual proteins) combining the data from all four stages of the parasite (Atwood *et al*, 2005).

Multidimensional liquid chromatography (MDLC) coupled to tandem mass spectrometry (MS) for proteomic analysis was first introduced by John Yates' group and became one of the most popular approaches to analyze large-scale proteomic samples (Washburn *et al*, 2001). In this approach, proteins from complex mixture samples are chemically or proteolytically digested, resulting in thousands of peptides, which are analyzed by MDLC-MS. In the first dimension, the separation is generally done in a strong-cation exchange (SCX), strong-anion exchange (SAX), reverse phase (RP), hydrophilic interaction (HILIC) or even mixed bead (SCX

and SAX) chromatography capillary column, whereas in the second dimension RP chromatography is widely used for its compatibility with MS solvents (reviewed in (Motoyama and Yates, 2008)).

Here, we have developed a simple and robust method to perform 2-D LC-MS/MS analysis, using autosampler injections to elute the peptides from the SCX capillary column. We applied this new method for the analysis of the proteome of infective blood trypomastigote forms of a highly virulent strain of *T. cruzi* (Y strain). Our results showed a substantial (about 2.5-fold) increase in the trypomastigote proteome coverage compared to the previous study (Atwood *et al*, 2005). Moreover, using bioinformatic analysis of putative surface proteins we have predicted thousands of potential epitope sequences for MHC class I and II with high binding affinity.

4.3 Materials and methods

4.3.1 Cell cultures and protein extraction

Tissue culture cell-derived trypomastigote forms of *T. cruzi* (Y strain) were obtained 5 to 9 days after infection of monolayers of green monkey kidney LLC-MK₂ epithelial cells (ATCC, Manassas, VA), maintained at 37°C, under 5% CO₂ atmosphere, in DMEM supplemented with 10% FBS, as described by Andrews and Colli (Andrews and Colli, 1982). Trypomastigotes were collected by centrifugation (3,000xg for 10 min at 4°C), washed 3 times with phosphate-buffered saline solution (PBS) and resuspended in PBS containing protease inhibitor cocktail (Sigma). The lysis was performed by sonicating the sample for 3x15 s with 30% amplitude (Sonics, Vibra Cell). After the lysis, the sample was centrifuged at 16,000xg for 20 min at 4°C. The soluble protein fraction was separated from the insoluble proteins and quantified with BCA kit (Pierce), according to the manufacturer's protocol.

4.3.2 Protein digestion

After quantification, 1 mg of soluble proteins and the equivalent number of parasites from the pellet (it was not possible to quantify protein in this fraction because its insolubility) was precipitated with 10% trichloroacetic acid for 20 min in an ice bath, centrifuged at 16,000xg for 20 min at 4°C, washed with ice-cold acetone, and dried in a vacuum centrifuge (Vacufuge, Eppendorf). Proteins were redissolved in 200 µL 0.4 M NH₄HCO₃ containing 8 M urea, and disulfide bounds were reduced with 5 mM dithiothreitol for 15 min at 50°C. Then, cysteine residues were alkylated with 10 mM iodoacetamide for 30 min at room temperature, protected from the light. Proteins were digested with 20 µg sequencing grade trypsin (Promega), for 24 h at 37°C. The reaction was stopped by adding 10 µL formic acid (FA) and

the peptides were desalted in a C18 solid phase extraction cartridge (1 mL, Discovery DSC-18, Supelco). The cartridge was activated with 4 mL methanol and equilibrated with 4 mL 0.05% TFA. After loading and washing with 4 mL 0.05% TFA, the sample was eluted with 2 mL 80% ACN/0.05% TFA and dried in a vacuum centrifuge.

4.3.3 Offline peptide fractionation

An equivalent of 50 µg of each sample (dissolved in 25%ACN/0.5% FA) was loaded onto a strong-cation exchange (SCX) ziptip manufactured with 25 µL POROS 50 HS resin (Applied Biosystems) in 200-µL micropipette tip, previously equilibrated with 200 µL 25% ACN/0.5% FA. The elution was done with 100 µL of increasing NaCl concentrations (0, 10, 20, 40, 60, 80, 100, 150, 200, and 500 mM NaCl in 25% ACN/0.5% FA, followed by 500 mM NaCl in 20% ACN/10 mM ammonium acetate, pH 7.0). The fractions were then dried in a vacuum centrifuge and desalted in reverse phase ziptips. The ziptips were fabricated with 20 µL POROS 50 R2 resin (40 mg/mL suspension in isopropanol) (Applied Biosystems) in 200-µL micropipette tip, and washed with 100 µL methanol. After equilibrating the ziptip with 200 µL 0.05% TFA, the sample was loaded and washed with 200 µL of the sample solution. The peptides were then eluted with 100 µL 80% ACN/0.05% TFA and dried in a vacuum centrifuge.

4.3.4 Two-dimensional liquid chromatography-tandem mass spectrometry (2-D LC-MS/MS)

Online and offline 2-D LC-MS/MS were performed in a nanoLC-1Dplus system (Eksigent, Dublin, CA) connected to an LTQ XL/ETD (Thermo Fisher Scientific, San Jose, CA). The system comprises one 20-µL loop, one SCX trap column (5 µL, Opti-Pak, Optimize Technologies), two C18 trap columns (0.25 µL, Opti-Pak) and one capillary C18 column

(Proteopep II, 10 cm x 75 μ m, 3 μ m, New Objective) (Fig. 4.1). For the online SCX fractionation, in the first dimension the peptides were loaded into the SCX trap column and the peptides were eluted using the autosampler by injecting 20 μ L of increasing salt concentrations (0, 10, 20, 40, 60, 80, 100, 150, 200, and 500 mM NaCl in 5% ACN/0.5% FA, followed by one injection of 500 mM NaCl in 20% ACN/10mM ammonium acetate, pH 7.0). Eluting peptides were automatically loaded into the C18-trap column, which was then washed with solvent A (solvent A = 2% ACN/0.1% FA) at 2.5 μ L/min flow rate for 130 min. The offline fractionated peptides were loaded directly onto the C18-trap column, which was then washed with solvent A at 2.5 μ L/min flow rate for 130 min. In the second dimension the peptides were separated in the capillary C18 column at 300 nL/min flow rate, using a linear gradient from 5-40% solvent B over 100 min (solvent B = 80% ACN/0.1% FA), followed by 10 min wash with 100% solvent B and 20 min with 5% solvent B. The electrospray was set to 1.9 kV, and the ion trap was set for a maximum injection time of 100 ms for full scan and 150 ms for MS/MS scan. The 10 most abundant ions were selected for CID with an isolation width of 3.0 a.m.u. and normalized collision energy of 35%. Dynamic exclusion was set for fragmenting each ion twice and then excluding it for 1 min.

4.3.5 Bioinformatics analysis

The MS/MS spectra from peptides from 600 to 3500 Da, at least 15 ions and a minimum of 10 counts were converted to DTA files using Bioworks (v3.3.1, Thermo Fisher Scientific). Then database search was done with TurboSequest algorithm against collected sequences from *T. cruzi*, bovine, human keratin, and porcine trypsin (downloaded on March 17th, 2008, from GenBank), in the forward and reverse orientations, forming a dataset of 191,762 sequences. The parameters for database search were: fully trypsin digestion; 1 missed

cleavage site; 2 Da for peptide mass tolerance; 1 Da for fragment mass tolerance; and cysteine carbamidomethylation as fixed modification and methionine oxidation as variable modification. The datasets were filtered with $DCn \geq 0.085$, protein probability $\leq 1e-3$, and Xcorr of 1.5, 2.2, and 2.7 for singly-, doubly- and triply-charged peptides, respectively. The datasets were then converted to XML files.

XML files were combined by an in-house written Perl script. The algorithm basically assembles the peptides into protein sequences by searching the peptide sequences using Blast and considering only those with 100% identity. The assembly was performed in an Intel Core 2 duo/2.53 MHz computer with 2 GB RAM memory and installed with Debian/Linux and data bank Postgres. The redundant protein hits were assembled into protein groups, since it is impossible to determine which from those proteins are actually expressed.

The valid proteins were annotated by gene ontology using GOblet algorithm (<http://goblet.molgen.mpg.de/cgi-bin/goblet/webapp-goblet.cgi>) (Hennig *et al*, 2003). GO analysis was carried out on October 15th, 2008, against the invertebrate taxonomy and with an e-value threshold $\leq 1e-10$. The GPI prediction analysis was performed using FragAnchor (<http://navet.ics.hawaii.edu/~fraganchor/NNHMM/NNHMM.html>) (Poisson *et al*, 2007) and only the sequences with high probability were accepted as potential GPI-anchored. The predicted surface proteins were predicted for MHC class I (H-2 Db allele) and II (HLA DRB1*0101 allele) epitopes using average prediction binding mode using Immune Epitope Database and Analysis Resource (IEDB) prediction tool (<http://tools.immuneepitope.org/>) (Zhang *et al*, 2008). Only epitopes with high binding affinity (inhibition concentration (IC) ≤ 10 nM) were accepted in our analysis.

4.4 Results and Discussion

4.4.1 Implementation of autosampler-based gradient in a two-dimensional liquid chromatography-tandem mass spectrometry system

Online 2-D LC-MS/MS is widely used for proteomic applications (Motoyama *et al*, 2008). In this case, the HPLC is composed of at least 2 binary pumps. One pump is set for the first dimension gradient (i.e. RP, SCX, HILIC), whereas the other pump is set for the gradient of the second dimension (RP-LC-MS/MS). When the separation is done in three dimensions, an HPLC with a quaternary pump is required for the first two dimensions of separation (Wei *et al*, 2005). Here we describe a simple way to perform online 2-D LC-MS/MS separations, without the need of having an HPLC with multiple pumps. In our design, the first dimension is driven by a primary HPLC pump set to a 2.5 $\mu\text{L}/\text{min}$ constant flow of 0.1% FA/ 2% ACN, which basically loads and washes the sample. The peptide mixture is loaded into a SCX trap column (5 μL) and the salt-step gradient is performed by injecting 20 μL of different concentrations of NaCl solution (4x column volume) using an autosampler. The eluted peptides are concentrated into one of the two RP trap columns (0.25 μL). In this setup, while the salt gradient is running into the first RP trap column, the RP gradient is running in the second RP trap column (Fig. 4.1, Table 4.1). In the next run, the RP gradient runs through the first RP trap column, while the SCX is eluting in the second RP trap column, thus eliminating time gaps between runs (Fig. 4.1, Table 4.1). Moreover, in our approach the number of different elution solvents is not limited by the number of pumps in the HPLC, thus it could be easily adapted to run 3-D LC-MS/MS. To test the possibility of using different elution solvents, the last step of SCX was performed with 20% ACN instead of 5% ACN, and 10 mM ammonium acetate (pH 7.0) instead 0.5% FA (pH 2.5), without noticeable changes in column pressure. Next, to estimate the

repeatability of the system, we ran the same sample twice (Table 4.2, sample C and D) and the difference in the number of detected peptides or protein groups were about 10%, which is similar to previous reported by other groups for MDLC-MS (Faca *et al*, 2007; Wei *et al*, 2005). We can conclude that the autosampler-based gradient is a simple, robust and inexpensive alternative to perform online 2-D separations. It is also worth to point out that our approach requires only minimum modifications in the 1-D LC-MS/MS system.

4.4.2 Comparison between offline and online 2-D LC-MS/MS

In order to evaluate the performance of our autosampler-based gradient, we compared this methodology with an offline separation, using of *T. cruzi* trypomastigote tryptic digests as the testing sample. Briefly, trypomastigotes were lysed by sonication and centrifuged. The resulting pellet (insoluble proteins) and supernatant (soluble proteins) were digested with trypsin and the equivalent of 20 µg was used for online or offline separations. The offline separation was carried out in SCX ziptips, whereas the online was performed as described in item 4.4.1. We could notice that the online separation led to the identification of more protein groups compared to the offline separation (Table 4.2). In the sample of insoluble proteins, online separation increased by more than 30% the number of identified protein groups, whereas in the soluble protein samples this number increased more than 70% (Table 4.2).

To investigate why the online separation has a better performance, we summed the number of peptides from all fractions (which includes the peptides eluted in more than fraction). We obtained similar numbers from the analysis of insoluble proteins (3970 and 4098 peptides for online and offline separations, respectively), but a lower number of peptides were detected in the offline separation of the soluble proteins (3347 peptides against 4514 and 4811 peptides for the online separation) (Table 4.2). However, the number of unique peptides (which

does not count the repeated peptides eluted in multiple fractions) was significantly different between distinct fractionation methods. By comparing the number of unique peptides and the sum of peptides from all fractions, we can estimate the efficacy of each separation method. The online separations showed that only 13.0-16.7% of the peptides were eluted in more than one fraction, whereas in the offline separations this number was 46.6-58.3% (Table 4.2). Taken together these results clearly show the superior performance of the autosampler-based online separation in comparison with the offline separation.

4.4.3 Proteomic analysis of trypomastigote forms of *T. cruzi*

Next, to study the proteins expressed by trypomastigotes, we combined all five 2-D LC-MS/MS runs and assembled the peptides into proteins and the redundant proteins into protein groups. The dataset was further filtered with protein probability ($<5e-4$) and sum of cross-correlation scores from distinct peptides (≥ 3.0), resulting in a false-discovery rate (FDR) of only 1.9% at the protein level. A total of 7370 individual proteins were identified, which correspond to 1616 protein groups. Of this total, 6154 proteins and 1448 protein groups were from *T. cruzi*, whereas 1216 proteins and 168 protein groups were contaminants from the medium and sample preparation (Table 4.3).

To functionally categorize the identified protein groups, we performed a gene ontology annotation using GOblet tool. Seventy-one percent of the sequences were annotated for molecular function. The most abundant activity was the enzymatic/catalytic activity (43.9%), with emphasis in hydrolases (22.4%), transferases (10.3%), and oxidoreductases (6.3%). Other well-represented categories were transport (7.5%) and binding activities (37.7%). From the cellular component categories (28.5%), we noticed that most of the proteins were intracellular (24.1%), associated to membranes (8.1%) and organelles (16.7%) (Fig. 4.2, Table

4.4). Among the biological process categories (55.9%), about half of the proteins were involved in metabolism (Fig. 4.2, Table 4.4). Interestingly, 147 (10.2%) protein groups were annotated to be involved in the “interaction between organisms” category, particularly in pathogenesis (Fig. 4.2, Table 4.4). Most of these proteins were from TS/gp85 superfamily (141 proteins), four were hypothetical proteins and one, 140/116 kDa antigen. TS/gp85 members have been shown to be essential to the parasite infection. TS is an enzyme that transfers sialic acid from host to parasite glycoconjugates, which are involved in the cell invasion and protection against the host immune response (Acosta-Serrano *et al*, 2007; Frasch, 2000). Other members have been shown to be involved in the adhesion to the host cells and protection against the complement system (Acosta-Serrano *et al*, 2007; Frasch, 2000).

Although not detected by GO annotation, other superfamilies of surface proteins, such as mucins and gp63, are known to be involved in the pathogenesis; thus, increasing the number of protein groups involved in the pathogenesis process to 170 (11.7%). Mucins and their GPI anchors have been shown to be strong proinflammatory agents, which activate Toll-like receptor 2 (TLR2) (Almeida *et al*, 2000; Campos *et al*, 2001). Recent data from our lab, have been shown that the activation of TLR2 in fact increases the host cell invasion by the parasite (Almeida *et al*, unpublished data). On the other hand, gp63 is a GPI-anchored metalloproteinase that was also shown to be involved in the host cell invasion (Cuevas *et al*, 2003). It is noticeable the high number of identified surface proteins (TS/gp85, mucins, and gp63) involved in the host cell invasion and virulence, make them a very attractive target to the development of immunotherapies.

4.4.4 Analysis of surface proteins and epitope mapping

The surface proteins are generally used as vaccine targets, since they are exposed and in direct contact with the immune system. Thus, we next focused in the analysis of *T. cruzi* surface proteins. Although the centrifugation is not a high performance separation technique, as expected we noticed a two- to three-fold enrichment of the surface proteins in the insoluble protein fraction (Table 4.2). A total of 208 (14.4%) protein groups (1087 individual proteins, 17.7%) were annotated as surface proteins. The identified surface protein families were TS/gp85 (140 protein groups/789 proteins), MASP (37 protein groups/136 proteins), mucin TcMUCII (12 protein groups/42 proteins), gp63 (11 protein groups/83 proteins), TolT (4 protein groups/16 proteins), mucin-like (3 protein groups/12 proteins), and YASP (1 protein group/5 proteins) (Table 4.5). Interestingly, we found a high number of MASP and mucin TcMUCII sequences. These proteins are difficult to be studied by proteomic analysis, since there are highly glycosylated and several of genes are expressed at the same time, decreasing the stoichiometry of individual proteins (Buscaglia *et al*, 2004). Thus, it is also possible that the number of expressed MASPs and TcMUCII mucins is still underestimated.

To determine the number of proteins that might be GPI-anchored, we performed a prediction analysis using FragAnchor algorithm (Poisson *et al*, 2007). We found 178 protein groups (12.3%) and 890 individual proteins (14.5%) with high probability to be GPI-anchored (Table 4.5). Of these entries, only 4 protein groups and 15 proteins were not previously annotated as surface proteins. We recently performed a genome-wide GPI-anchoring prediction analysis and found that about 12.9% of the *T. cruzi* genes encodes possible GPI-anchored proteins (Nakayasu *et al.*, submitted to publication). These numbers are much higher compared to other eukaryotes that have from 0.5 to 2% predicted GPI-anchored genes in their genomes (Poisson *et al*, 2007).

Next, we performed the prediction analysis of putative epitopes for MHC class I and II presentation using annotated membrane and predicted GPI-anchored proteins. For MHC class I prediction, 28 and 168 epitopes with high binding affinities ($IC \leq 10$ nM) were found in protein groups and proteins, respectively (Table 4.5). On the other hand, the prediction of putative MHC class II epitopes led to the identification of 6,999 and 36,206 epitopes with high binding affinity ($IC \leq 10$ nM) in protein groups and proteins, respectively (Table 4.5).

A prediction for MHC class I epitopes of TS sequences was previously carried out and demonstrated that immunodominant epitopes vary according to the parasite strain (Martin *et al*, 2006). Thus, Martin *et al*. proposed that *T. cruzi* evades cytotoxic T cell response by having a different subset of TS proteins expressed on the cell surface for each parasite cell or strain (Martin *et al*, 2006). Of those epitopes we predicted, only one sequence fully matched with the prediction made for Brazil and CL Brenner strains (Alvarez *et al*, 2008; Martin *et al*, 2006), supporting the idea that different strains of parasite express unique subsets of surface proteins, or even more, the genes from different strains might also present significant variations.

In the humoral immune response against *T. cruzi*, patients generally present very high titers of lytic anti- α -galactosyl (anti- α Gal) antibodies (Almeida *et al*, 1994; Almeida *et al*, 1991; Avila *et al*, 1989; Gazzinelli *et al*, 1991). It has been shown that α -Gal epitopes are very abundant on the cell surface of the trypomastigotes, and since humans normally do not express these epitopes, they are highly immunogenic (Almeida *et al*, 1994; Macher and Galili, 2008). Anti- α Gal antibodies are able to lyse the parasite plasma membrane even in the absence of complement system and control the infection (Almeida *et al*, 1994; Almeida *et al*, 1991; Gazzinelli *et al*, 1991). Here we show that parasite may express from hundreds to more than one thousand proteins on the cell surface. Furthermore, the prediction of MHC class

epitopes led to the identification of many thousands of sequences. The very high number of proteins expressed in the cell surface may lead to a “dilution” of the epitopes present to immune response. A recent study also showed that antibodies against one specific TS sequence failed to inhibit the enzymatic activity of the whole cell lysate, supporting the idea that many sequences are expressed at the same time (Ratier *et al*, 2008). Moreover, antibodies against gp63 or Tc85 only partially inhibit host cell invasion, probably because they recognize only a small subset of these molecules (Alves *et al*, 1986; Cuevas *et al*, 2003). Therefore, the dominant humoral immune response during Chagas disease is against the more conserved epitope, the α -Gal, abundantly expressed on trypomastigote mucins and other major glycoconjugates (Almeida *et al*, 1994; Almeida *et al*, 1993; Couto *et al*, 1990). In the evolutionary point of view, it might seem paradoxical that the parasite expresses large amounts of this highly immunogenic glycosidic epitope. However, it is worth to point out that only humans and Old World primates lack α -Gal epitopes (Macher *et al*, 2008), and the contact of humans with *T. cruzi* may have occurred from 50,000 to 10,000 years ago with the arrival of humans in America (Teixeira *et al*, 2006), thus a short period of time in the evolution of *T. cruzi*.

4.5 Conclusions

We have developed a simple and robust approach to perform 2-D LC-MS/MS analysis, using autosampler injections to elute peptides from the first dimension column. The application of this methodology to analyze the whole cell lysate of *T. cruzi* trypomastigotes led to the identification of 1448 non-redundant proteins from *T. cruzi*. Our results show that the trypomastigote stage of *T. cruzi* is rich in cell surface virulence factors, such TS/gp85 and mucins. Furthermore, a bioinformatic analysis for putative MHC epitopes predicted over one hundred high-affinity binding peptides for MHC class I and thousands of sequences for MHC class II. These findings may have many implications for the design of immunotherapeutics for the prevention or treatment of Chagas disease.

4.6 Acknowledgements

This work was funded by grants from the National Institutes of Health (5S06GM08012-37, 1R01AI070655, and 5G12RR008124). E.S.N. was partially supported by the George A. Krutilek memorial graduate scholarship from Graduate School, UTEP. 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. We are also grateful to Dr. Daniel M. Lorenzini (Petrobras, Brazil) for the critical reading of the manuscript.

4.6 References

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.

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, Ferguson MA, Schenkman S, Travassos LR (1994) Lytic anti-alpha-galactosyl antibodies from patients with chronic Chagas' disease recognize novel O-linked oligosaccharides on mucin-like glycosyl-phosphatidylinositol-anchored glycoproteins of *Trypanosoma cruzi*. *Biochem J* **304 (Pt 3)**: 793-802.

Almeida IC, Krautz GM, Krettli AU, Travassos LR (1993) Glycoconjugates of *Trypanosoma cruzi*: a 74 kD antigen of trypomastigotes specifically reacts with lytic anti-alpha-galactosyl antibodies from patients with chronic Chagas disease. *J Clin Lab Anal* **7**: 307-316.

Almeida IC, Milani SR, Gorin PA, Travassos LR (1991) Complement-mediated lysis of *Trypanosoma cruzi* trypomastigotes by human anti-alpha-galactosyl antibodies. *J Immunol* **146**: 2394-2400.

Alvarez MG, Postan M, Weatherly DB, Albareda MC, Sidney J, Sette A, Olivera C, Armenti AH, Tarleton RL, Laucella SA (2008) HLA Class I-T Cell Epitopes from trans-Sialidase Proteins Reveal Functionally Distinct Subsets of CD8 T Cells in Chronic Chagas Disease. *PLoS Negl Trop Dis* **2**: e288.

Alves MJ, Abuin G, Kuwajima VY, Colli W (1986) Partial inhibition of trypomastigote entry into cultured mammalian cells by monoclonal antibodies against a surface glycoprotein of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **21**: 75-82.

Andrews NW, Colli W (1982) Adhesion and interiorization of *Trypanosoma cruzi* in mammalian cells. *J Protozool* **29**: 264-269.

Atwood JA, 3rd, Weatherly DB, Minning TA, Bundy B, Cavola C, Opperdoes FR, Orlando R, Tarleton RL (2005) The *Trypanosoma cruzi* proteome. *Science* **309**: 473-476.

Avila JL, Rojas M, Galili U (1989) Immunogenic Gal alpha 1----3Gal carbohydrate epitopes are present on pathogenic American *Trypanosoma* and *Leishmania*. *J Immunol* **142**: 2828-2834.

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 A, Jr., 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.

Campos MA, Almeida IC, Takeuchi O, Akira S, Valente EP, Procopio DO, Travassos LR, Smith JA, Golenbock DT, Gazzinelli RT (2001) Activation of Toll-like receptor-2 by glycosylphosphatidylinositol anchors from a protozoan parasite. *J Immunol* **167**: 416-423.

Couto AS, Goncalves MF, Colli W, de Lederkremer RM (1990) The N-linked carbohydrate chain of the 85-kilodalton glycoprotein from *Trypanosoma cruzi* trypomastigotes contains sialyl, fucosyl and galactosyl (alpha 1-3)galactose units. *Mol Biochem Parasitol* **39**: 101-107.

Cuevas IC, Cazzulo JJ, Sanchez DO (2003) gp63 homologues in *Trypanosoma cruzi*: surface antigens with metalloprotease activity and a possible role in host cell infection. *Infect Immun* **71**: 5739-5749.

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, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**: 409-415.

Faca V, Pitteri SJ, Newcomb L, Glukhova V, Phanstiel D, Krasnoselsky A, Zhang Q, Struthers J, Wang H, Eng J, Fitzgibbon M, McIntosh M, Hanash S (2007) Contribution of protein fractionation to depth of analysis of the serum and plasma proteomes. *J Proteome Res* **6**: 3558-3565.

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.

Gazzinelli RT, Pereira ME, Romanha A, Gazzinelli G, Brener Z (1991) Direct lysis of *Trypanosoma cruzi*: a novel effector mechanism of protection mediated by human anti-gal antibodies. *Parasite Immunol* **13**: 345-356.

Hennig S, Groth D, Lehrach H (2003) Automated Gene Ontology annotation for anonymous sequence data. *Nucleic Acids Res* **31**: 3712-3715.

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 U S A* **104**: 2035-2042.

Macher BA, Galili U (2008) The Gal α 1,3Gal β 1,4GlcNAc-R (alpha-Gal) epitope: a carbohydrate of unique evolution and clinical relevance. *Biochim Biophys Acta* **1780**: 75-88.

Martin DL, Weatherly DB, Laucella SA, Cabinian MA, Crim MT, Sullivan S, Heiges M, Craven SH, Rosenberg CS, Collins MH, Sette A, Postan M, Tarleton RL (2006) CD8⁺ T-Cell responses to *Trypanosoma cruzi* are highly focused on strain-variant trans-sialidase epitopes. *PLoS Pathog* **2**: e77.

Motoyama A, Yates JR, 3rd (2008) Multidimensional LC separations in shotgun proteomics. *Anal Chem* **80**: 7187-7193.

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.

Ratier L, Urrutia M, Paris G, Zarebski L, Frasch AC, Goldbaum FA (2008) Relevance of the diversity among members of the *Trypanosoma cruzi* trans-sialidase family analyzed with camelids single-domain antibodies. *PLoS ONE* **3**: e3524.

Stuart K, Brun R, Croft S, Fairlamb A, Gurtler RE, McKerrow J, Reed S, Tarleton R (2008) Kinetoplastids: related protozoan pathogens, different diseases. *J Clin Invest* **118**: 1301-1310.

Tarleton RL (2007) Immune system recognition of *Trypanosoma cruzi*. *Curr Opin Immunol* **19**: 430-434.

Tarleton RL, Reithinger R, Urbina JA, Kitron U, Gurtler RE (2007) The challenges of Chagas Disease-- grim outlook or glimmer of hope. *PLoS Med* **4**: e332.

Teixeira AR, Nascimento RJ, Sturm NR (2006) Evolution and pathology in chagas disease--a review. *Mem Inst Oswaldo Cruz* **101**: 463-491.

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

Washburn MP, Wolters D, Yates JR, 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**: 242-247.

Wei J, Sun J, Yu W, Jones A, Oeller P, Keller M, Woodnutt G, Short JM (2005) Global proteome discovery using an online three-dimensional LC-MS/MS. *J Proteome Res* **4**: 801-808.

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 U S A* **105**: 5022-5027.

Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, Bui HH, Buus S, Frankild S, Greenbaum J, Lund O, Lundegaard C, Nielsen M, Ponomarenko J, Sette A, Zhu Z, Peters B (2008) Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res* **36**: W513-518.

Table 4.1 Fraction number and autosampler injection.

Run	Valve position	Online		Offline	
		Injected sample	RP gradient	Injected sample	RP gradient
1	1	20 µg tryptic digest	Blank	Fraction 1	Blank
2	2	Fraction 2	Fraction 1	Fraction 2	Fraction 1
3	1	Fraction 3	Fraction 2	Fraction 3	Fraction 2
4	2	Fraction 4	Fraction 3	Fraction 4	Fraction 3
5	1	Fraction 5	Fraction 4	Fraction 5	Fraction 4
6	2	Fraction 6	Fraction 5	Fraction 6	Fraction 5
7	1	Fraction 7	Fraction 6	Fraction 7	Fraction 6
8	2	Fraction 8	Fraction 7	Fraction 8	Fraction 7
9	1	Fraction 9	Fraction 8	Fraction 9	Fraction 8
10	2	Fraction 10	Fraction 9	Fraction 10	Fraction 9
11	1	Fraction 11	Fraction 10	Fraction 11	Fraction 10
12	2	Blank	Fraction 11	Blank	Fraction 11

Fraction	Elution solvent	ACN %	
		Offline	Online
1	0.5% FA	25	5
2	10 mM NaCl/0.5% FA	25	5
3	20 mM NaCl/0.5% FA	25	5
4	40 mM NaCl/0.5% FA	25	5
5	60 mM NaCl/0.5% FA	25	5
6	80 mM NaCl/0.5% FA	25	5
7	100 mM NaCl/0.5% FA	25	5
8	150 mM NaCl/0.5% FA	25	5
9	200 mM NaCl/0.5% FA	25	5
10	500 mM NaCl/0.5% FA	25	5
11	500 mM NaCl/10mM H3COONH4	20	20

Table 4.2 Statistics of number of identified peptides and protein groups

Fraction #	Sample A	Sample B	Sample C	Sample D	Sample E
1	12	21	16	22	18
2	18	108	33	43	59
3	13	911	15	22	772
4	602	919	834	910	820
5	935	518	772	839	434
6	562	703	591	641	549
7	544	378	508	471	323
8	299	327	614	687	245
9	425	152	530	548	93
10	506	57	517	554	31
11	54	4	84	74	3
Sum of number peptides in all fraction	3970	4098	4514	4811	3347
Number of unique peptides	3456	2796	3867	4259	2115
% of peptides eluted in more than one fraction	14.9	46.6	16.7	13.0	58.3
Peptide - FDR (%)	2.6	2.7	1.4	2.8	2.7
Number of protein groups	1258	940	1008	1097	608
Surface proteins	149	129	85	87	41

Sample	Description
A	Peptides derived from insoluble proteins - online fractionation - run 1
B	Peptides derived from insoluble proteins - offline fractionation - run 1
C	Peptides derived from soluble proteins - online fractionation - run 1
D	Peptides derived from soluble proteins - online fractionation - run 2
E	Peptides derived from soluble proteins - offline fractionation - run 1

Table 4.3 Identified *T. cruzi*-specific protein groups

	groups	proteins
<i>T. cruzi</i>	1448	6154
Contaminants	168	1216
Total	1616	7370

Group	# proteins	Accession number protein description	Unique peptides	Spectra	\sum Xcorr	Probability
1	2	XP_821225.1 pyruvate phosphate dikinase	60	1819	228.99	6.36E-06
2	7	AAT35592.1 elongation factor 2	53	1861	224.18	1.64E-06
3	4	XP_809762.1 elongation factor 2	50	1884	216.79	1.64E-06
4	3	EAN98352.1 pyruvate phosphate dikinase	57	1658	215.30	6.36E-06
5	4	XP_819508.1 chaperonin HSP60, mitochondrial precursor	46	3114	209.92	2.33E-07
6	7	Q95046 Chaperonin HSP60, mitochondrial precursor	46	3090	209.17	3.36E-07
7	1	BAA09433.1 elongation factor 2	47	1614	203.14	1.64E-06
8	1	AAG12986.1 pyruvate phosphate dikinase 2	54	1604	201.15	6.36E-06
9	14	EAN93041.1 heat shock protein 85	45	2630	182.09	1.85E-06
10	9	AAL75957.1 beta tubulin 2.3	38	8557	182.03	6.01E-08
11	2	EAN95887.1 heat shock protein 70	45	2198	180.39	2.65E-05
12	2	EAN95886.1 heat shock protein 70	45	2342	179.63	2.65E-05
13	3	AAA30205.1 heat shock protein 70	44	2260	178.81	2.65E-05
14	3	CAA47952.1 heat shock protein 70	44	2023	169.59	2.65E-05
15	2	EAO00045.1 aconitase	41	956	160.82	7.96E-05
16	10	EAN87966.1 glucose-regulated protein 78	39	1295	158.20	0.000247
17	7	AAA30221.1 major paraflagellar rod protein	39	1186	157.81	0.000468
18	4	EAN92219.1 heat shock protein	38	622	142.62	7.53E-07
19	4	EAN81053.1 alpha tubulin	29	8013	141.99	0.000126
20	4	Q27352 Tubulin alpha chain	28	7326	136.62	0.000126
21	10	EAN81600.1 NADH-dependent fumarate reductase	38	301	134.97	0.000491
22	2	EAN85366.1 hypothetical protein	36	442	131.11	0.000141
		EAN88964.1 glycosomal phosphoenolpyruvate				
23	8	carboxykinase	32	1379	129.59	0.000324
24	2	XP_809814.1 cytoskeleton-associated protein CAP5.5	34	538	127.41	5.02E-05

25	2	XP_803757.1 cytoskeleton-associated protein CAP5.5	33	563	125.81	5.02E-05
26	4	EAN83233.1 trans-sialidase	32	1068	123.80	1.62E-05
27	7	EAN87927.1 clathrin heavy chain	33	203	117.98	3.43E-05
28	5	EAN97849.1 enolase	29	1978	116.62	7.38E-05
29	5	EAN86294.1 phosphoglycerate kinase	27	578	112.68	0.000151
30	2	EAN96827.1 transitional endoplasmic reticulum ATPase EAN84370.1 heat shock 70 kDa protein, mitochondrial precursor	28	344	110.48	6.69E-05
31	12	P05456 heat shock 70 kDa protein	27	711	110.38	0.000481
32	2	XP_820347.1 phosphoglycerate mutase	27	1237	107.16	2.47E-06
33	2	EAN82858.1 elongation factor 1-alpha (EF-1-alpha)	26	459	106.02	7.37E-05
34	47	EAN87519.1 fructose-bisphosphate aldolase, glycosomal	24	2717	105.98	5.69E-06
35	8	XP_814639.1 cytoskeleton-associated protein CAP5.5	24	603	102.53	9.35E-06
36	2	EAN98870.1 hypothetical protein	27	440	100.77	2.90E-05
37	6	XP_818396.1 trans-sialidase	24	410	96.73	3.81E-05
38	2	EAN87141.1 hexokinase	26	843	95.55	1.12E-05
39	6	XP_817834.1 ATPase beta subunit	24	602	94.42	5.37E-06
40	4	EAN83529.1 hypothetical protein	23	391	94.28	4.23E-06
41	2	EAN95608.1 protein kinase A regulatory subunit	26	393	92.56	0.000141
42	3	EAN86295.1 phosphoglycerate kinase	23	531	91.69	2.39E-05
43	5	EAN86005.1 chaperonin alpha subunit	22	378	90.21	1.71E-05
44	4	XP_821599.1 isocitrate dehydrogenase	24	162	89.84	3.90E-06
45	7	EAN98776.1 pyruvate kinase 2	24	482	89.38	0.000339
46	2	XP_818259.1 malic enzyme	25	212	89.29	4.29E-05
47	4	EAN85937.1 aspartate aminotransferase	21	959	89.25	2.14E-05
48	5	EAN94513.1 threonyl-tRNA synthetase	22	334	86.58	0.000426
49	2	EAN90123.1 seryl-tRNA synthetase	22	160	85.72	8.88E-05
50	7	EAN84558.1 arginine kinase	22	165	83.90	9.21E-05
51	2	XP_812195.1 hypothetical protein	21	1381	83.52	9.24E-08
52	2	EAN88376.1 hypothetical protein	22	96	82.72	3.64E-05
53	4	EAN89351.1 pyruvate kinase 2	21	317	80.41	6.23E-05
54	2	O96507 Arginine kinase	22	182	79.70	4.29E-05
55	3	EAN91216.1 protein disulfide isomerase	20	1263	77.98	9.24E-08
56	4	XP_809276.1 hypothetical protein	20	550	77.76	0.000255
57	4	EAN89698.1 heat shock 70 kDa protein	20	196	77.62	4.77E-08
58	4	XP_818742.1 paraflagellar rod component	21	499	75.35	1.31E-05
59	6		19	175	75.07	0.000353

60	2	XP_819282.1 hypothetical protein	22	115	74.29	0.000364
61	6	XP_818651.1 lipophosphoglycan biosynthetic protein	21	342	74.20	1.06E-05
62	2	EAN95952.1 hypothetical protein	18	568	73.04	6.59E-08
63	2	EAN89173.1 threonyl-tRNA synthetase	19	130	72.44	8.88E-05
64	30	AAX40296.1 actin	17	366	72.29	0.000204
65	4	EAN96732.1 succinyl-CoA ligase [GDP-forming] beta-chain EAN89502.1 D-isomer specific 2-hydroxyacid	18	400	72.25	9.62E-08
66	7	dehydrogenase-protein	19	259	71.66	6.60E-05
67	2	XP_814131.1 ATP-dependent DEAD/H RNA helicase	18	266	68.77	5.14E-05
68	13	XP_812138.1 glyceraldehyde 3-phosphate dehydrogenase	17	1634	68.03	3.08E-06
69	2	EAN99782.1 t-complex protein 1, eta subunit	17	102	66.47	2.12E-05
70	5	EAN87302.1 S-adenosylhomocysteine hydrolase	18	308	65.55	4.99E-05
71	2	EAN86961.1 pyridoxal kinase	16	190	65.38	8.86E-07
72	4	XP_808914.1 hypothetical protein	16	1055	65.28	6.45E-05
73	5	EAN98527.1 eukaryotic initiation factor 4a	16	325	63.99	3.35E-06
74	2	EAN91517.1 isocitrate dehydrogenase [NADP]	15	243	61.69	1.77E-05
75	6	XP_821173.1 protein disulfide isomerase	17	217	61.06	0.000123
76	2	XP_816963.1 hypothetical protein	16	116	60.74	2.12E-07
77	7	AAR22882.1 cyclic nucleotide phosphodiesterase	15	94	59.98	0.000408
78	4	XP_819492.1 ubiquitin-activating enzyme E1	14	80	59.54	4.90E-05
79	2	EAN98468.1 chaperonin	17	124	59.45	0.000133
80	6	EAN92377.1 calcium-translocating P-type ATPase	16	91	59.12	8.78E-06
81	4	EAN86474.1 vacuolar ATP synthase subunit B	16	74	58.97	1.55E-05
82	2	EAN92158.1 2-amino-3-ketobutyrate coenzyme A ligase	14	138	58.30	8.76E-05
83	2	EAN98983.1 succinyl-CoA synthetase alpha subunit	13	379	58.24	5.47E-08
84	5	EAN90953.1 asparagine synthetase A	14	458	57.51	0.000125
85	4	XP_807185.1 hypothetical protein EAN97607.1 ADP-ATP carrier protein 1, mitochondrial	16	75	57.37	1.69E-05
86	23	precursor	16	386	57.06	8.30E-06
87	18	XP_819811.1 aminopeptidase	15	85	57.04	4.01E-07
88	2	XP_809675.1 hypothetical protein	13	495	56.64	6.59E-08
89	6	XP_818370.1 hypothetical protein	16	138	56.43	0.000377
90	4	XP_818003.1 hypothetical protein	17	68	56.22	0.000355
91	2	EAN95883.1 activated protein kinase C receptor	14	234	56.09	1.43E-07
92	2	EAN95823.1 heat shock protein 78	16	82	55.95	0.000464
93	2	XP_818369.1 hypothetical protein	16	147	55.70	0.000377
94	3	XP_810535.1 heat shock protein DnaJ	13	201	55.42	0.000233

95	2	XP_809575.1 glutamate dehydrogenase	15	257	55.16	0.000144
96	41	AAA66352.1 trans-sialidase	13	121	54.94	1.54E-05
97	4	XP_810375.1 3-ketoacyl-CoA thiolase	13	138	54.30	7.02E-05
		XP_808700.1 UTP-glucose-1-phosphate uridylyltransferase				
98	4	2	15	96	54.17	0.000166
99	4	XP_807578.1 peptidylprolyl isomerase-like	14	198	54.03	0.000212
100	2	EAN89180.1 trans-sialidase	14	173	53.44	0.000158
101	4	EAN89878.1 hypothetical protein	13	91	53.18	5.88E-07
102	2	EAN80791.1 trans-sialidase	13	329	53.16	1.62E-05
103	4	EAN83442.1 2-amino-3-ketobutyrate coenzyme A ligase	13	137	53.06	4.20E-09
104	11	AAX40410.1 trypanothione reductase	14	106	52.98	0.0001
105	3	AAC06213.1 putative glutamate dehydrogenase	14	261	52.55	0.000144
106	2	EAN82648.1 NADH-dependent fumarate reductase	15	106	51.94	0.000112
		EAN94510.1 mitochondrial processing peptidase, beta				
107	6	subunit	14	91	51.78	6.42E-05
108	4	EAN87846.1 aminopeptidase	14	95	51.60	0.000413
109	2	XP_813169.1 I/6 autoantigen	13	272	51.57	1.88E-05
110	6	AAC32076.1 paraflagellar rod component Par4	16	64	51.47	0.000169
111	2	XP_814739.1 acyl-CoA dehydrogenase	12	115	51.47	0.000454
112	3	BAA84681.1 succinate dehydrogenase	13	116	51.13	1.10E-06
113	6	XP_820270.1 cAMP specific phosphodiesterase	14	81	51.09	0.000408
114	4	XP_814674.1 ribosomal protein L3	12	215	51.03	2.44E-05
115	8	P26796 60S acidic ribosomal protein P0	12	211	51.00	1.60E-05
116	1	AAC97378.1 TcSTI1	13	113	50.66	0.000134
117	2	XP_811342.1 nucleoside phosphorylase	13	227	49.84	2.82E-06
118	2	AAT07417.1 85 kDa surface glycoprotein	14	240	49.11	0.000149
119	2	XP_821793.1 asparaginyl-tRNA synthetase	15	68	49.04	0.000204
120	9	1MS4 trans-sialidase	12	93	49.02	1.17E-05
121	8	EAN82665.1 60S ribosomal protein L4	12	338	48.99	4.23E-06
122	4	XP_805359.1 trans-sialidase	14	287	48.82	0.000333
123	9	EAN87849.1 40S ribosomal protein S4	15	258	48.66	0.000494
124	4	XP_812989.1 hypothetical protein	14	60	48.37	0.0003
125	2	EAN89721.1 60S ribosomal protein L10a	13	213	47.80	0.000414
126	2	EAN99332.1 RAB GDP dissociation inhibitor alpha	14	87	47.76	0.000391
127	2	AAP32151.1 surface glycoprotein Tc-85/32	14	164	47.72	0.000293
128	16	EAN80947.1 trypanothione peroxidase	12	645	47.71	3.37E-07
129	2	XP_821559.1 protein phosphatase	13	111	47.25	0.000148

130	2	XP_804015.1 chaperonin TCP20	13	116	46.81	4.07E-05
131	4	XP_805897.1 trans-sialidase	13	107	46.63	0.000299
132	4	XP_805772.1 inosine-5'-monophosphate dehydrogenase	13	45	46.46	0.000285
133	4	EAN93464.1 trans-sialidase	13	89	45.76	2.29E-05
134	4	EAO00033.1 60S ribosomal protein L10a	13	212	45.76	0.000414
135	4	EAO00186.1 aminopeptidase	14	44	45.66	0.000254
136	4	XP_808069.1 pyrroline-5-carboxylate synthetase-like protein	13	115	45.48	1.37E-05
137	2	EAN96072.1 hypothetical protein	11	134	45.43	9.48E-06
138	5	AAP80764.1 unknown	11	108	45.16	1.54E-05
139	1	AAD08694.1 SERCA-type calcium-ATPase	12	65	44.89	7.81E-07
140	2	EAN84211.1 hypothetical protein	12	114	44.87	4.15E-05
141	4	XP_806053.1 thimet oligopeptidase	13	108	44.67	0.000138
142	26	AAK71216.1 trypanothione reductase	12	93	44.64	0.0001
143	6	EAN85586.1 axoneme central apparatus protein	13	58	44.46	7.48E-05
144	4	XP_804821.1 trans-sialidase	12	85	44.22	0.000293
145	5	EAN99758.1 poly(A)-binding protein	11	227	44.13	7.18E-05
146	37	XP_821969.1 glucose-6-phosphate isomerase, glycosomal	12	86	44.00	1.08E-05
147	2	EAN85115.1 trans-sialidase	12	85	43.96	0.000293
148	9	AAP32150.1 surface glycoprotein Tc-85/20	13	227	43.92	0.000454
149	2	XP_805740.1 transaldolase	12	218	43.85	0.0003
150	2	XP_815489.1 inosine-5'-monophosphate dehydrogenase	12	37	43.83	0.000285
151	2	XP_819104.1 malate dehydrogenase	9	581	43.51	6.22E-09
152	4	XP_820902.1 hypothetical protein	12	33	43.33	0.000423
153	4	XP_814980.1 nucleoside phosphorylase	11	141	43.08	2.80E-05
154	3	EAN96884.1 chaperonin containing T-complex protein	12	97	43.06	0.000148
155	2	XP_812243.1 trans-sialidase	12	61	43.01	2.29E-05
156	2	XP_820557.1 hypothetical protein	11	121	42.95	2.78E-05
157	8	XP_821649.1 Gim5A protein	10	171	42.68	9.27E-06
158	10	AAP32144.1 surface glycoprotein Tc-85/35 XP_814868.1 vacuolar-type proton translocating	12	259	42.58	0.000293
159	6	pyrophosphatase 1	10	95	42.50	4.79E-07
160	4	EAN84299.1 citrate synthase	11	140	42.38	1.63E-05
161	6	EAN87308.1 glycyl-tRNA synthetase	11	56	42.27	0.000304
162	5	XP_810144.1 adenylate kinase	10	100	42.16	1.27E-05
163	4	EAN99395.1 hypothetical protein	13	27	42.00	0.000176
164	4	EAN85268.1 kynureninase	12	52	41.64	0.00013

165	6	XP_820100.1 calpain-like cysteine peptidase	11	163	41.51	0.000331
166	4	XP_820498.1 trans-sialidase	10	158	41.15	3.89E-06
167	4	EAN88318.1 guanine deaminase	11	79	41.08	0.000289
168	5	XP_816015.1 14-3-3 protein	10	273	40.95	5.24E-06
169	2	EAN86462.1 aspartate aminotransferase, mitochondrial	11	70	40.92	7.71E-06
170	2	EAN94997.1 retrotransposon hot spot (RHS) protein	12	60	40.79	0.000331
171	9	AAB48404.1 29 kDa proteasome subunit TCPR29	10	77	40.65	0.000166
172	2	EAN89048.1 pyridoxal kinase	10	151	40.59	8.86E-07
173	6	EAN97794.1 40S ribosomal protein S3a	11	231	40.58	2.01E-06
174	4	EAN92871.1 hypothetical protein XP_804829.1 inosine-adenosine-guanosine-nucleoside	11	131	40.42	0.00018
175	4	hydrolase	9	74	40.30	0.000154
176	2	XP_820921.1 hypothetical protein	11	170	40.18	0.000221
177	4	XP_814911.1 carnitine/choline acetyltransferase	11	78	40.15	0.000473
178	6	EAN82966.1 trans-sialidase	11	107	40.06	8.58E-06
179	2	EAO00202.1 6-phospho-1-fructokinase	11	67	39.56	6.03E-05
180	2	EAN96878.1 dynein	11	41	38.94	0.000254
181	2	XP_804332.1 trans-sialidase	10	146	38.78	4.79E-05
182	10	XP_819899.1 glyceraldehyde 3-phosphate dehydrogenase	10	139	38.46	4.77E-05
183	4	XP_821530.1 NIMA-related kinase	9	49	38.34	3.57E-06
184	12	EAN94705.1 trans-sialidase	11	98	38.28	0.000114
185	6	EAN85931.1 proteasome activator protein PA26	10	111	38.14	1.26E-05
186	4	XP_818124.1 hypothetical protein	10	67	38.11	0.000187
187	5	XP_816016.1 phosphoenolpyruvate mutase	10	68	37.92	0.000263
188	3	EAN99727.1 cyclophilin A	9	805	37.79	0.000162
189	2	AAK71205.1 trypanothione reductase	11	89	37.70	6.49E-05
190	4	EAN87716.1 microtubule-associated protein	11	1167	37.66	0.00031
191	4	XP_808405.1 fatty acyl CoA synthetase	12	35	37.55	6.47E-05
192	4	EAN90895.1 V-type ATPase, A subunit	10	50	37.46	6.49E-07
193	5	EAN99691.1 40 kDa cyclophilin	9	111	37.03	8.36E-08
194	8	EAN81441.1 fatty acyl CoA synthetase 2	10	28	36.71	0.000263
195	4	EAN98106.1 hypothetical protein	10	54	36.49	2.60E-05
196	8	EAN81270.1 sterol 24-c-methyltransferase	8	41	36.45	3.00E-05
197	2	EAN99779.1 calmodulin	10	68	36.41	5.18E-05
198	2	EAN99796.1 hypothetical protein	9	60	36.22	0.000409
199	4	EAN81864.1 surface protein TolT	8	194	36.16	3.51E-06

200	2	EAN81585.1 microtubule-associated protein	10	1305	36.13	7.55E-05
201	2	EAN94994.1 electron transfer flavoprotein	9	47	35.98	1.00E-05
202	6	XP_802286.1 trans-sialidase	10	82	35.95	0.000148
203	5	EAN90450.1 nascent polypeptide associated complex	9	129	35.82	0.0001
204	2	XP_816558.1 trans-sialidase	10	89	35.78	5.70E-05
205	2	XP_807183.1 2-oxoglutarate dehydrogenase subunit	10	74	35.77	0.00036
206	4	EAN88342.1 alanyl-tRNA synthetase	9	38	35.65	0.000395
207	6	XP_808230.1 proteasome alpha 5 subunit	9	69	35.45	0.000301
208	6	XP_804506.1 trans-sialidase	11	155	35.31	0.000417
209	2	EAN81368.1 hypothetical protein	9	59	35.26	0.000251
210	2	EAN92382.1 hypothetical protein	9	61	35.20	0.000489
211	4	EAN94544.1 hypothetical protein	7	423	35.16	5.71E-06
212	4	EAN81927.1 thioredoxin	10	48	35.03	1.01E-05
213	8	EAN81557.1 serine carboxypeptidase (CBP1)	8	152	34.86	1.63E-08
214	11	XP_821294.1 trans-sialidase	11	52	34.79	3.67E-05
215	4	EAN92411.1 hypothetical protein	9	579	34.72	8.43E-06
216	6	AAP69224.1 SR RNA-binding protein; TcSR62	8	202	34.55	0.000124
217	11	XP_811932.1 ATP-dependent RNA helicase	10	58	34.49	0.000203
218	2	EAN92689.1 mitochondrial phosphate transporter	8	98	34.43	0.000101
219	4	EAN91257.1 inosine-5'-monophosphate dehydrogenase	10	34	34.38	0.00033
220	4	EAN84570.1 trifunctional enzyme alpha subunit	10	38	34.23	8.54E-06
221	6	EAN84223.1 mitochondrial RNA binding protein	7	66	34.11	5.90E-09
222	4	EAN83480.1 fatty acyl CoA synthetase 2	10	37	33.58	2.65E-05
223	2	XP_807499.1 trans-sialidase	8	77	33.57	1.17E-05
224	22	AAX84939.1 iron superoxide dismutase A	7	145	33.50	3.73E-06
225	4	EAN99686.1 proteasome alpha 1 subunit	9	62	33.47	4.98E-07
226	2	EAN97412.1 elongation factor TU	9	61	33.42	2.50E-05
227	5	EAN95097.1 glycosomal malate dehydrogenase	9	195	33.37	8.11E-05
228	4	XP_802208.1 trans-sialidase	8	156	33.34	2.66E-05
229	3	EAN94309.1 hypothetical protein	9	26	33.30	9.61E-06
230	4	EAN84039.1 hypothetical protein	10	34	33.20	0.000283
231	8	1CI1 Triosephosphate Isomerase	9	163	33.20	6.87E-05
232	4	XP_806359.1 IgE-dependent histamine-releasing factor XP_805653.1 D-isomer specific 2-hydroxyacid	8	118	33.20	9.62E-05
233	2	dehydrogenase-protein	8	114	33.12	0.000112
234	14	EAN82322.1 trans-sialidase	8	203	33.07	3.28E-05

235	4	XP_816802.1 hypothetical protein	11	35	32.90	0.000465
236	4	EAN87689.1 hypothetical protein	8	178	32.30	2.15E-05
237	6	EAN96507.1 proteasome regulatory ATPase subunit 5	9	53	32.22	6.02E-06
238	4	XP_811282.1 hypothetical protein	9	37	32.07	9.38E-05
239	2	XP_820714.1 nucleoside diphosphate kinase	8	271	32.03	0.000233
240	6	XP_813787.1 14-3-3 protein	8	105	32.01	5.54E-05
241	1	AAC14084.1 TcC31.28	8	101	31.88	0.000146
242	2	XP_810932.1 trans-sialidase	9	35	31.76	0.000277
243	4	XP_813166.1 centrin	8	63	31.74	0.0004
244	2	EAN91745.1 hypothetical protein	8	31	31.47	2.28E-05
245	5	EAN99433.1 60S ribosomal protein L7a	7	252	31.45	4.97E-07
246	2	XP_809984.1 hypothetical protein	10	80	31.42	5.06E-05
247	4	XP_817446.1 40S ribosomal protein S6	9	172	31.35	2.79E-06
248	2	XP_804534.1 hypothetical protein	8	38	31.34	2.28E-05
249	2	XP_820102.1 calpain cysteine peptidase	8	32	31.06	3.37E-05
250	3	AAR89402.1 glucose-6-phosphate isomerase	8	73	31.03	0.000116
251	4	EAO00183.1 dynamin	9	33	31.01	0.000249
252	2	XP_821293.1 trans-sialidase	9	51	30.92	3.67E-05
253	5	XP_802951.1 mannose-1-phosphate guanylttransferase	8	47	30.79	7.67E-06
254	5	EAN80725.1 trans-sialidase	7	242	30.79	0.000293
255	2	EAN83138.1 calpain cysteine peptidase	8	260	30.60	4.43E-05
256	4	XP_802786.1 surface protein TolT	7	294	30.52	3.51E-06
257	4	XP_803285.1 hypothetical protein	8	30	30.44	0.000408
258	2	EAN92430.1 hypothetical protein	9	44	30.44	0.000434
259	6	EAN96014.1 trans-sialidase	9	48	30.29	3.67E-05
260	2	EAN96370.1 trans-sialidase	8	205	30.16	3.28E-05
261	2	EAN94050.1 hypothetical protein	9	40	30.08	0.000434
262	4	EAN86654.1 hypothetical protein	7	52	29.99	7.60E-06
263	2	XP_806810.1 serine/threonine-protein kinase A	7	91	29.88	7.40E-05
264	2	EAN81198.1 hypothetical protein	8	68	29.81	3.13E-05
265	4	XP_806512.1 DNAK protein	7	93	29.50	7.01E-06
266	4	XP_803906.1 GTP-binding nuclear protein rtb2	8	128	29.34	3.35E-05
267	6	EAN87804.1 60S ribosomal protein L2	7	85	29.08	2.00E-05
268	2	EAN81356.1 ATP-dependent DEAD/H RNA helicase	8	60	29.08	0.000191
269	4	EAN95380.1 hypothetical protein	7	38	29.02	4.10E-06
270	4	XP_821554.1 eukaryotic initiation factor 5a	7	214	28.80	8.20E-07

271	2	XP_803925.1 hypothetical protein	6	92	28.78	0.000164
272	2	XP_819402.1 hypothetical protein	8	24	28.75	4.27E-05
273	1	AAA19650.1 surface glycoprotein	7	322	28.71	0.000293
274	10	XP_804492.1 40S ribosomal protein SA	7	97	28.68	4.55E-06
275	2	XP_815652.1 hypothetical protein	6	86	28.03	0.000164
276	23	BAA13411.1 calcium-binding protein	7	897	27.95	9.05E-07
277	4	XP_812661.1 hypothetical protein	7	265	27.93	5.87E-06
278	4	XP_821125.1 hypothetical protein	8	45	27.88	0.000366
279	1	AAR89404.1 glucose-6-phosphate isomerase	7	73	27.81	3.14E-07
280	5	XP_806100.1 cofilin/actin depolymerizing factor	6	184	27.58	4.64E-05
281	4	XP_813577.1 ubiquitin-like protein	8	37	27.57	0.000138
282	2	EAN98484.1 hypothetical protein	8	27	27.48	0.000101
283	3	AAS44545.2 NUP-1	8	33	27.37	0.000101
284	4	XP_807771.1 calpain-like cysteine peptidase	7	796	27.27	3.61E-06
285	4	XP_812457.1 hypothetical protein	7	47	27.26	2.77E-05
286	2	XP_818548.1 surface protease GP63	7	39	27.25	0.000134
287	4	EAN97600.1 RNA-binding protein	6	50	27.11	0.000301
288	2	EAN83044.1 trans-sialidase	8	22	27.06	0.000333
289	3	XP_821190.1 UDP-glucose:glycoprotein glucosyltransferase	9	14	26.99	1.61E-05
290	5	AAD11551.1 pyruvate dehydrogenase E1 alpha subunit	8	33	26.76	0.000131
291	4	XP_814749.1 40S ribosomal protein S17	7	81	26.71	0.000124
292	2	EAN96137.1 hypothetical protein	7	76	26.53	0.000176
293	2	EAN97565.1 LA RNA binding protein	7	60	26.46	4.75E-05
294	4	XP_811646.1 pyruvate dehydrogenase E1 beta subunit	7	87	26.34	9.81E-05
295	12	EAN94386.1 retrotransposon hot spot (RHS) protein	8	37	26.22	0.000198
296	6	EAN90617.1 phosphoprotein phosphatase	7	35	26.16	1.40E-05
297	4	XP_802841.1 hypothetical protein	7	120	26.13	4.96E-05
298	7	EAN82577.1 ADP-ribosylation factor 1	7	128	26.06	0.000227
299	12	AAY84580.1 metacaspase 3	7	41	25.93	1.99E-06
300	5	XP_808554.1 isoleucyl-tRNA synthetase	8	25	25.88	1.28E-05
301	4	XP_803032.1 hypothetical protein	7	54	25.86	0.00028
302	8	EAN93883.1 enoyl-CoA hydratase/isomerase	7	32	25.80	3.86E-06
303	4	XP_816550.1 hypothetical protein	8	19	25.79	0.000162
304	4	XP_813358.1 electron-transfer-flavoprotein	7	46	25.76	1.98E-06
305	11	EAN94869.1 trans-sialidase	7	66	25.66	3.67E-05
306	10	AAD13039.1 TolT2	6	301	25.60	4.23E-06

307	4	XP_819463.1 protein kinase	7	38	25.50	1.54E-05
308	4	EAN97606.1 polyprenyl synthase	7	26	25.42	0.000363
309	2	XP_803362.1 ATP synthase, alpha chain	6	159	25.38	1.50E-06
310	4	EAN98349.1 40S ribosomal protein S5	6	62	25.33	0.000237
311	6	CAA43297.1 histone H2b	8	658	25.30	2.50E-05
312	3	EAO00026.1 elongation factor TU	7	50	25.30	3.98E-05
313	6	EAN97454.1 ribosomal protein L18	7	139	25.16	9.55E-05
314	4	XP_812031.1 RNA-binding protein	7	22	24.96	0.000244
315	1	AAD13040.1 TolT3	6	304	24.94	4.23E-06
316	8	XP_810702.1 60S ribosomal protein L30	6	101	24.86	1.87E-05
317	4	XP_811290.1 hypothetical protein	7	37	24.77	3.26E-05
318	4	EAN96274.1 ADP-ribosylation factor 3	5	39	24.54	1.47E-05
319	4	XP_804093.1 co-chaperone GrpE	7	84	24.54	1.11E-05
320	2	EAN92044.1 phosphatase 2C	7	16	24.53	0.00032
321	2	EAN81498.1 trans-sialidase	6	81	24.46	1.49E-06
322	4	EAN91919.1 proteasome beta 3 subunit	6	76	24.43	2.40E-06
323	2	EAN98432.1 nucleolar RNA binding protein	6	27	24.32	2.90E-05
324	6	EAN96565.1 60S ribosomal protein L7	7	145	24.25	1.39E-05
325	4	XP_819893.1 dihydrolipoamide acetyltransferase	8	15	24.08	0.000122
326	4	EAN88636.1 RNA-binding protein	7	45	24.06	1.66E-05
327	4	EAN86941.1 hypothetical protein	7	36	24.00	0.000353
328	2	XP_819091.1 heat shock 70 kDa protein	5	110	23.90	9.68E-10
329	2	XP_804908.1 flagellar calcium-binding protein	6	790	23.81	5.15E-05
330	4	XP_816689.1 coatomer beta subunit	7	13	23.79	1.94E-05
331	10	AAM08671.1 TC3_70K14.5	6	56	23.76	1.76E-06
332	7	XP_820872.1 cytochrome C oxidase subunit IV	6	35	23.75	6.24E-05
333	4	EAN97997.1 hypothetical protein	6	101	23.70	4.47E-06
334	4	EAN95558.1 hypothetical protein	6	32	23.58	4.32E-05
335	2	XP_802386.1 proteasome alpha 2 subunit	6	63	23.55	1.19E-05
336	4	EAN86093.1 hypothetical protein	7	35	23.51	0.000266
337	4	AAG15409.1 trypanothione synthetase	6	15	23.42	2.24E-05
338	4	EAN83217.1 trans-sialidase	6	131	23.42	4.79E-05
339	27	AAO63160.1 thiol transferase Tc52	6	20	23.41	0.000138
340	8	XP_816821.1 trans-sialidase	7	41	23.26	3.67E-05
341	2	EAN82967.1 retrotransposon hot spot (RHS) protein	6	34	23.25	2.77E-05
342	2	XP_815446.1 hypothetical protein	6	28	23.23	5.70E-05

343	5	XP_816709.1 valyl-tRNA synthetase	7	37	23.22	2.42E-05
344	12	XP_803179.1 casein kinase, delta isoform	7	58	23.13	0.000132
345	4	EAN84474.1 hypothetical protein	6	143	23.12	1.38E-07
346	4	EAN98765.1 hypothetical protein	6	25	23.11	5.85E-05
347	2	XP_818809.1 hypothetical protein	6	26	23.10	2.27E-08
348	4	XP_808539.1 oligosaccharyl transferase subunit	6	21	23.04	7.67E-05
349	2	XP_809678.1 RuvB-like DNA helicase	6	27	23.03	4.78E-05
350	2	XP_820017.1 trans-sialidase	6	43	23.00	1.32E-05
351	5	XP_810517.1 adenylate kinase	6	24	22.71	2.02E-05
352	2	EAN83964.1 40S ribosomal protein S3	6	163	22.69	9.80E-07
353	9	AAO84044.1 surface protein-2	7	135	22.65	8.72E-05
354	4	CAA46198.1 ribosomal P2-type protein	5	107	22.62	2.57E-10
355	2	XP_817806.1 RuvB-like DNA helicase	6	27	22.59	6.20E-06
356	12	EAN83293.1 casein kinase	7	51	22.55	0.000132
357	14	XP_812827.1 histone H4	6	432	22.54	3.67E-05
358	2	EAN89623.1 hypothetical protein	6	34	22.53	1.55E-05
359	2	XP_818550.1 surface protease GP63	6	27	22.49	0.000134
360	6	BAB12394.1 proteasome regulatory non-ATPase	8	19	22.45	5.16E-05
361	1	AAD30064.1 laminin receptor precursor-like protein	5	28	22.38	4.55E-06
362	2	XP_811664.1 hypothetical protein	5	36	22.37	1.05E-05
363	4	XP_817615.1 trans-sialidase	7	41	22.29	0.000208
364	4	XP_804999.1 hypothetical protein	4	122	22.28	5.95E-12
365	4	XP_821334.1 hypothetical protein	7	24	22.21	0.000106
366	4	EAN94026.1 proteasome beta 6 subunit	5	24	22.20	6.25E-06
367	8	EAN94904.1 phosphoglucomutase	6	13	22.08	8.26E-05
368	14	XP_821762.1 trans-sialidase	7	39	22.08	0.000382
369	2	XP_811191.1 proteasome regulatory non-ATPase subunit	6	41	22.04	3.81E-08
370	4	EAN97393.1 phosphoribosylpyrophosphate synthetase	6	9	21.98	4.66E-05
371	4	EAN94076.1 histone H4	6	378	21.89	1.29E-06
372	4	XP_812659.1 flagellar radial spoke component	6	22	21.85	8.71E-05
373	8	EAN88665.1 60S ribosomal protein L10	6	148	21.80	8.17E-06
374	4	XP_815632.1 hypothetical protein	7	30	21.80	0.000302
375	4	XP_821244.1 hypothetical protein	5	32	21.67	4.38E-05
376	4	EAN94556.1 hypothetical protein	6	24	21.62	5.01E-06
377	4	XP_809232.1 glutamyl-tRNA synthetase	6	36	21.56	0.000235
378	8	EAN95636.1 mitotubule-associated protein Gb4	6	26	21.55	1.64E-07

379	5	CAE17296.1 UDP-Glc 4'-epimerase	6	16	21.52	0.000134
380	5	EAN80741.1 hypothetical protein	5	20	21.50	7.93E-06
381	2	XP_804864.1 heat shock protein 78	6	43	21.45	2.86E-07
382	2	XP_821425.1 trans-sialidase	6	44	21.39	7.71E-06
383	4	EAN96753.1 40S ribosomal protein S12	5	69	21.39	2.09E-07
384	4	XP_808509.1 hypothetical protein	7	32	21.39	0.000146
385	2	EAN85135.1 aminopeptidase	7	24	21.31	2.99E-05
386	2	EAN97177.1 proteasome regulatory non-ATPase subunit 2	6	34	21.21	0.000296
387	2	EAN89705.1 hypothetical protein	5	29	21.19	0.000474
388	4	EAN89115.1 60S ribosomal protein L13	7	110	21.11	1.04E-05
389	4	EAN93281.1 hypothetical protein	7	18	21.10	0.000353
390	4	XP_819462.1 aminopeptidase	6	47	21.08	0.000154
391	8	XP_808034.1 60S ribosomal subunit protein L31	5	90	21.08	5.36E-08
392	2	EAN99751.1 dynein light chain	5	104	21.04	1.03E-07
393	4	EAN86475.1 2-oxoglutarate dehydrogenase	5	77	21.04	1.78E-05
394	2	XP_818967.1 hypothetical protein	5	22	21.03	1.28E-08
395	2	EAN98191.1 40S ribosomal protein S10	6	129	21.01	0.000282
396	2	XP_803323.1 HSP100	7	31	20.97	0.00039
397	4	XP_820056.1 40S ribosomal protein S18	6	110	20.93	0.000339
398	4	XP_806793.1 phenylalanyl-tRNA synthetase	5	12	20.85	0.000166
399	2	EAN85469.1 NADH-dependent fumarate reductase	7	55	20.82	0.000491
400	7	EAN97448.1 RNA-binding protein	5	79	20.71	6.24E-06
401	6	EAN81251.1 40S ribosomal protein S8	5	151	20.64	1.03E-06
402	5	XP_816536.1 methylthioadenosine phosphorylase	5	34	20.61	7.39E-06
403	4	EAN89566.1 beta prime COP protein	6	18	20.48	0.000301
405	6	XP_816177.1 trans-sialidase	6	65	20.42	7.71E-06
404	4	XP_812008.1 ribonucleoside-diphosphate reductase	6	16	20.42	0.000203
406	4	EAN89875.1 adenylosuccinate lyase	7	24	20.36	0.000252
407	4	XP_812798.1 40S ribosomal protein S16	6	155	20.34	4.25E-06
408	4	EAN90899.1 ribosomal protein L21E (60S)	6	208	20.18	0.000403
409	5	EAN88319.1 high mobility group protein	5	40	20.06	0.00016
410	2	EAN95314.1 trans-sialidase	6	192	19.97	0.000149
411	9	EAN83324.1 trans-sialidase	6	143	19.96	0.000133
412	4	XP_814064.1 60S ribosomal protein L23a	5	98	19.82	0.000274
413	5	EAN82258.1 calpain cysteine peptidase	5	190	19.66	0.000161
414	5	XP_807454.1 trans-sialidase	5	138	19.60	1.23E-05

415	12	AAF04810.1 KMP11	6	281	19.56	0.000125
416	2	XP_817934.1 hypothetical protein	6	67	19.55	9.48E-07
417	4	XP_812735.1 phosphomannomutase-like protein	6	16	19.50	7.39E-06
418	6	EAN83787.1 nucleolar protein	5	14	19.46	4.27E-05
419	4	XP_820608.1 hypothetical protein	5	60	19.44	0.00011
420	4	XP_808001.1 hypothetical protein	5	30	19.43	1.29E-07
421	2	EAN81218.1 trans-sialidase	5	53	19.40	0.000208
422	6	XP_815622.1 40S ribosomal protein S3	6	87	19.38	3.16E-06
423	4	XP_810238.1 hypothetical protein	6	34	19.37	2.23E-05
424	71	EAN91875.1 trans-sialidase	5	35	19.31	0.000148
425	6	XP_804734.1 3,2-trans-enoyl-CoA isomerase	4	57	19.30	1.04E-08
426	3	AAQ84130.1 pumilio protein 1	6	12	19.28	0.000164
427	6	XP_814823.1 chaperone DnaJ protein	6	11	19.21	1.05E-05
428	2	EAN96342.1 trans-sialidase	6	26	19.18	0.000153
429	8	XP_807701.1 10 kDa heat shock protein	4	570	19.17	8.89E-06
430	2	EAN87676.1 proteasome beta 7 subunit	5	16	19.12	0.000268
431	4	XP_811491.1 Hsc70-interacting protein (Hip)	5	23	19.09	0.000147
432	2	EAN81565.1 2-oxoglutarate dehydrogenase subunit	5	37	19.05	0.000483
433	4	BAD99565.1 TS-193 trans-sialidase homolog protein	4	37	19.04	1.17E-05
434	4	EAN87915.1 dolichyl-phosphate beta-D-mannosyltransferase	6	15	18.99	0.000356
436	2	XP_814968.1 dynein heavy chain	5	17	18.99	1.67E-08
435	2	EAN98362.1 protein transport protein sec31	5	13	18.99	6.94E-06
437	4	EAN82797.1 trans-sialidase	6	20	18.92	0.000137
438	2	EAN99701.1 hypothetical protein	5	43	18.92	7.39E-09
439	1	AAD45370.1 Tc45-calreticulin precursor	5	139	18.87	2.75E-05
440	4	XP_815061.1 vesicle-associated membrane protein	5	30	18.82	2.12E-05
441	4	XP_810048.1 ribosomal protein S19	6	61	18.81	5.44E-08
442	3	XP_817773.1 GTP-binding protein	5	18	18.79	0.000443
443	3	EAN86318.1 tryparedoxin	5	409	18.70	1.23E-07
444	4	XP_811758.1 hypothetical protein	4	20	18.70	3.15E-08
445	2	EAN96316.1 hypothetical protein	4	121	18.68	3.14E-05
446	2	XP_809978.1 methionyl-tRNA synthetase	5	25	18.65	2.87E-08
447	4	EAN96118.1 6-phosphogluconate dehydrogenase	5	81	18.61	1.42E-06
448	4	XP_812747.1 hypothetical protein	5	32	18.55	3.95E-05
449	7	XP_819826.1 60S ribosomal protein L18	4	61	18.49	2.71E-05

450	8	BAA21088.1 carbamyl phosphate synthase	5	17	18.46	0.000448
451	2	XP_820393.1 trans-sialidase	5	48	18.43	0.000148
452	5	CAA30335.1 unnamed protein product	5	411	18.42	0.000419
453	2	XP_813065.1 10 kDa heat shock protein	4	571	18.41	8.89E-06
454	2	EAN84471.1 transketolase	5	12	18.33	0.000151
455	7	XP_803249.1 60S ribosomal protein L11	5	41	18.31	0.000105
456	2	EAN85763.1 small GTP-binding protein Rab1	5	43	18.11	0.000337
457	4	EAN89520.1 deoxyribose-phosphate aldolase	4	14	18.08	1.38E-05
458	2	EAN80974.1 2-oxoglutarate dehydrogenase subunit	4	32	18.07	3.76E-05
459	4	XP_812197.1 NADH dehydrogenase	5	20	18.06	1.02E-08
460	4	XP_810946.1 60S acidic ribosomal protein	5	71	17.98	0.000256
461	4	XP_806132.1 trans-sialidase	5	63	17.94	3.95E-07
462	2	XP_819690.1 eukaryotic translation initiation factor 5	5	33	17.90	5.17E-07
463	4	XP_807041.1 hypothetical protein	5	56	17.80	4.22E-05
464	4	EAN81826.1 anion-transporting ATPase-like	5	20	17.77	3.77E-05
465	4	XP_804250.1 proliferative cell nuclear antigen (PCNA)	5	22	17.74	1.78E-07
466	4	XP_809846.1 trans-sialidase	5	35	17.69	0.000208
467	2	XP_818987.1 ATP synthase, epsilon chain	4	23	17.68	4.86E-05
468	4	EAN95745.1 hypothetical protein	6	15	17.67	0.000151
469	2	EAN83583.1 hypothetical protein	5	13	17.53	0.000333
470	12	XP_813612.1 aminoacylase	4	27	17.48	0.000305
471	4	EAN83270.1 hypothetical protein	6	26	17.45	0.000223
472	2	EAN92305.1 trans-sialidase	5	50	17.44	0.000208
473	4	XP_821817.1 protein phosphatase 2C	5	33	17.42	7.07E-06
474	4	EAN84714.1 ubiquitin-like protein	5	39	17.40	0.000456
475	9	XP_816871.1 spermidine synthase	4	46	17.38	2.25E-06
476	4	EAN92693.1 hypothetical protein	4	179	17.32	8.95E-08
477	2	XP_817332.1 retrotransposon hot spot (RHS) protein	5	18	17.29	0.000283
478	4	EAN91727.1 trans-sialidase	5	57	17.27	0.000148
479	2	XP_808012.1 hypothetical protein	4	12	17.24	0.000182
480	6	XP_808440.1 hypothetical protein	5	18	17.22	0.000367
481	4	XP_819338.1 hypothetical protein	4	47	17.19	3.14E-05
482	4	XP_820214.1 eukaryotic release factor 3	5	8	17.17	9.59E-07
483	2	EAN92610.1 hypothetical protein	5	18	17.09	0.000444
484	4	EAN85913.1 hypothetical protein	5	45	17.04	1.77E-06
485	8	EAN88427.1 40S ribosomal protein S33	5	93	17.00	1.88E-06

486	4	XP_818652.1 U-box domain protein	4	21	16.96	9.56E-05
487	2	XP_822060.1 serine/threonine protein kinase	4	36	16.95	3.02E-05
488	4	EAN96915.1 mitochondrial RNA-binding protein 2	5	52	16.88	4.69E-06
489	2	ABM91080.1 thiol transferase Tc52	5	14	16.84	0.000138
490	2	XP_811473.1 hypothetical protein	5	7	16.84	0.000156
491	1	AAR14054.1 Tcr300	5	20	16.83	5.62E-05
492	4	EAN99918.1 trans-sialidase	5	72	16.77	0.000293
493	15	EAN87701.1 iron superoxide dismutase	4	242	16.75	2.26E-05
494	2	XP_809719.1 trans-sialidase	5	24	16.73	0.000148
495	4	XP_820276.1 glutamyl-tRNA synthetase	5	24	16.69	2.74E-05
496	2	EAN96171.1 ATP-dependent RNA helicase	5	26	16.65	0.000262
497	2	EAN99210.1 proteasome regulatory non-ATPase subunit 6	5	14	16.61	0.000113
498	9	AAK71861.1 farnesyl pyrophosphate synthase	4	27	16.60	0.000123
499	4	EAN90878.1 hypothetical protein	4	72	16.59	3.10E-05
500	4	XP_809137.1 ubiquitin-conjugating enzyme-like	5	38	16.54	0.000113
501	4	EAN91861.1 pyrroline-5-carboxylate reductase	5	16	16.52	0.000323
502	2	EAN87422.1 hypothetical protein	5	39	16.48	0.000134
503	2	XP_814392.1 hypothetical protein	4	77	16.40	1.95E-07
504	4	EAN95873.1 coatomer gamma subunit	5	9	16.36	0.000273
505	2	XP_805577.1 trans-sialidase	5	120	16.25	2.33E-05
506	4	XP_808422.1 hypothetical protein	4	28	16.16	3.10E-05
507	4	EAN89851.1 trans-sialidase	5	28	16.15	0.000293
508	2	XP_813751.1 hypothetical protein	5	16	16.14	0.000251
509	4	XP_809583.1 hypothetical protein	5	22	16.07	5.82E-05
510	4	XP_813924.1 hypothetical protein	3	126	16.05	3.21E-10
511	4	XP_810451.1 hypothetical protein	4	12	16.01	2.71E-08
512	2	XP_818382.1 trans-sialidase	4	23	15.93	0.000208
513	8	EAN90093.1 40S ribosomal protein S13	6	58	15.93	0.000118
514	4	XP_816707.1 hypothetical protein	4	35	15.92	0.000176
515	4	EAN97098.1 dynein heavy chain	4	12	15.91	2.30E-06
516	5	XP_820169.1 proteasome regulatory ATPase subunit 3	4	20	15.89	3.25E-05
517	8	XP_808341.1 ubiquitin/ribosomal protein S27a	3	92	15.88	1.27E-09
518	4	EAN98760.1 serine/threonine protein phosphatase type 5	4	13	15.84	9.74E-06
519	6	XP_812043.1 trans-sialidase	5	58	15.81	0.000339
520	8	EAN92456.1 60S ribosomal protein L6	4	138	15.81	9.84E-07
521	2	XP_803124.1 trans-sialidase	5	47	15.80	1.48E-06

522	4	EAN85253.1 proteasome regulatory non-ATPase subunit 11	4	26	15.77	2.08E-05
523	4	XP_819083.1 sphingosine phosphate lyase-like protein	4	10	15.76	0.000223
524	16	EAN90460.1 trans-sialidase	4	8	15.75	0.000208
525	4	XP_804542.1 trans-sialidase	4	45	15.74	5.09E-06
526	4	EAN83146.1 cysteine synthase	4	16	15.73	0.000436
527	2	EAN82193.1 thioredoxin	4	23	15.71	2.08E-06
528	2	XP_814681.1 hypothetical protein	4	7	15.70	4.73E-12
529	4	EAN98515.1 heat shock protein 70	5	28	15.70	0.000487
530	8	EAN93071.1 retrotransposon hot spot (RHS) protein	4	32	15.66	0.000385
531	17	EAN87440.1 iron superoxide dismutase	4	160	15.66	2.26E-05
532	6	EAN95573.1 hypothetical protein	5	13	15.65	0.000441
533	2	EAN83785.1 mitogen-activated protein kinase	4	17	15.63	2.40E-06
534	4	EAN94452.1 trans-sialidase	4	22	15.61	3.89E-06
535	2	EAN98221.1 trans-sialidase	4	152	15.60	1.30E-08
536	8	XP_820211.1 60S ribosomal protein L28	4	115	15.56	4.38E-07
537	8	EAN96038.1 60S ribosomal protein L32	4	88	15.56	1.84E-06
538	2	XP_811687.1 trans-sialidase	5	70	15.53	0.000258
539	4	XP_809418.1 hypothetical protein	3	33	15.52	1.32E-12
540	4	EAN95383.1 reiske iron-sulfur protein precursor	5	15	15.50	7.90E-05
541	6	EAN99415.1 40S ribosomal protein S9	5	59	15.50	0.00027
542	5	EAN99772.1 protein kinase A catalytic subunit isoform 1	4	37	15.47	2.71E-08
543	8	EAN88401.1 60S ribosomal protein L13a	4	91	15.44	0.000244
544	4	XP_810308.1 UV excision repair RAD23-like protein	4	7	15.43	1.48E-09
545	4	XP_815269.1 hypothetical protein	5	16	15.41	1.34E-07
546	2	EAN93748.1 serine/threonine protein phosphatase	4	16	15.38	0.000204
547	2	XP_816048.1 hypothetical protein	5	21	15.34	0.00018
548	2	EAN98355.1 hypothetical protein	4	13	15.31	0.000136
549	24	EAN90205.1 hypothetical protein	5	14	15.31	0.000436
550	2	EAN83722.1 flagellar calcium-binding protein	5	361	15.29	5.15E-05
551	4	XP_811677.1 trans-sialidase	4	8	15.28	0.000382
552	4	XP_815374.1 hypothetical protein	4	7	15.27	0.000119
553	4	EAN95781.1 heat shock protein-like protein	4	10	15.19	0.000284
554	2	XP_812701.1 trans-sialidase	5	16	15.18	0.000353
555	2	XP_812930.1 t-complex protein 1 gamma subunit	4	12	15.14	3.90E-05
556	4	XP_807179.1 hypothetical protein	4	15	15.11	3.18E-06
557	2	XP_814301.1 thiolase protein-like protein	5	16	15.08	2.47E-05

558	4	EAN96126.1 hypothetical protein	4	16	15.08	1.18E-05
559	5	XP_802970.1 profilin	4	36	15.04	0.000185
560	2	XP_818708.1 trans-sialidase	5	32	15.03	0.000148
561	2	XP_818642.1 flagellar radial spoke protein-like	5	19	15.03	3.37E-05
562	142	EAN86328.1 trans-sialidase	4	21	14.99	0.000208
563	2	EAN99600.1 phosphomannomutase	4	23	14.98	3.76E-05
564	30	XP_813720.1 histone H2A 1JVV Trypanosoma Cruzi Macrophage Infectivity Potentiator	5	308	14.96	2.20E-05
565	5		5	61	14.94	0.000347
566	2	EAN93143.1 hypothetical protein	4	18	14.93	0.000474
567	2	XP_813913.1 flagellar radial spoke protein-like	5	18	14.91	3.37E-05
568	4	EAN88212.1 hypothetical protein	4	58	14.87	1.23E-05
569	2	XP_815117.1 trans-sialidase	4	85	14.85	4.79E-05
570	2	EAN92914.1 hypothetical protein	4	14	14.80	8.96E-05
571	2	XP_813584.1 retrotransposon hot spot (RHS) protein	4	8	14.80	0.000395
572	5	XP_816750.1 trans-sialidase	5	180	14.77	0.000171
573	2	XP_819437.1 mitochondrial processing peptidase	4	18	14.77	1.39E-06
574	2	EAN86534.1 retrotransposon hot spot (RHS) protein	4	30	14.76	0.000385
575	4	XP_818916.1 hypothetical protein	5	9	14.72	0.000269
576	2	EAN81110.1 calpain-like cysteine peptidase	4	52	14.71	1.13E-07
577	4	XP_819481.1 acetyl-CoA synthetase	3	15	14.71	9.50E-06
578	2	XP_817636.1 hypothetical protein	5	20	14.64	0.000134
579	2	XP_809010.1 trans-sialidase	4	45	14.62	2.51E-05
580	4	EAN81360.1 hypothetical protein	4	19	14.58	0.00024
581	2	XP_805228.1 peptide methionine sulfoxide reductase	4	32	14.52	0.000215
582	2	EAN88246.1 hypothetical protein	4	12	14.49	2.82E-05
583	4	XP_808225.1 ATP synthase F1 subunit gamma protein	4	20	14.49	1.74E-06
584	2	EAN81497.1 retrotransposon hot spot (RHS) protein	4	30	14.48	3.56E-05
585	7	XP_809580.1 trans-sialidase	4	32	14.45	0.000148
586	3	AAK58518.1 aldo/keto reductase	4	35	14.39	5.51E-05
587	2	XP_817289.1 hypothetical protein	4	8	14.37	7.42E-05
588	4	XP_811338.1 hypothetical protein	4	18	14.34	3.12E-05
589	3	AAA03205.1 85 kDa surface antigen	4	15	14.32	0.000333
590	8	XP_809065.1 hypothetical protein	4	15	14.31	0.000121
591	2	EAN95506.1 translation initiation factor	4	7	14.25	0.000151
592	4	EAN87865.1 protein kinase	4	6	14.21	0.000383

593	4	EAN98665.1 hypothetical protein	4	9	14.14	5.19E-06
594	8	XP_805847.1 retrotransposon hot spot (RHS) protein	4	76	14.08	0.00038
595	4	EAN83179.1 S-adenosylmethionine synthetase	4	42	14.00	1.31E-05
596	9	XP_809735.1 ras-related protein rab-2a	4	15	13.88	1.00E-05
597	5	XP_807641.1 small GTP-binding protein Rab11	4	35	13.87	1.11E-05
598	2	XP_819974.1 hypothetical protein	3	11	13.84	0.000122
599	8	EAN86913.1 ubiquitin hydrolase	4	17	13.80	5.00E-06
600	6	XP_813077.1 myo-inositol-1-phosphate synthase	4	15	13.76	0.000135
601	3	XP_818353.1 trans-sialidase	4	35	13.74	0.000148
602	9	XP_808840.1 trans-sialidase	4	29	13.68	0.000148
603	8	AAL87542.1 proton motive P-type ATPase 2	4	5	13.68	0.000344
604	5	EAN81939.1 fatty acid desaturase	5	11	13.66	0.000461
605	4	XP_813054.1 ubiquitin-conjugating enzyme E2	3	38	13.63	0.000205
606	4	XP_817835.1 ribosomal protein S25	4	47	13.62	6.30E-06
607	4	XP_817652.1 26S protease regulatory subunit	4	22	13.59	6.37E-05
608	2	XP_806554.1 valyl-tRNA synthetase	4	22	13.56	2.42E-05
609	2	XP_804448.1 trans-sialidase	4	86	13.55	0.000277
610	8	EAN87769.1 40S ribosomal protein S11	4	79	13.42	1.45E-05
611	4	XP_818232.1 trans-sialidase	4	31	13.40	0.000208
612	4	EAN83303.1 p21 antigen protein	3	12	13.38	4.62E-09
613	4	XP_806215.1 nucleosome assembly protein-like protein	4	4	13.33	3.21E-07
614	2	XP_820917.1 hypothetical protein	3	5	13.29	4.82E-10
615	9	AAM74138.1 U-rich RNA-binding protein UBP-2	3	67	13.27	3.01E-11
616	4	EAN98394.1 p22 protein precursor	4	49	13.26	2.04E-07
617	2	XP_817816.1 trans-sialidase	4	35	13.25	0.000148
618	4	EAN83691.1 ADP-ribosylation factor GTPase activating protein 1	4	8	13.20	0.000216
619	6	EAN95459.1 trans-sialidase	3	33	13.19	9.16E-05
620	2	XP_820602.1 protein kinase	4	18	13.19	4.48E-05
621	2	XP_806550.1 calmodulin	3	15	13.15	5.35E-07
622	1	AAD51095.1 microtubule-associated protein homolog	4	603	13.14	3.93E-05
623	4	EAN86644.1 trans-sialidase	4	16	13.09	0.000333
624	4	EAN85754.1 mitogen activated protein kinase	4	10	13.07	0.000199
625	5	XP_806801.1 kinetoplast DNA-associated protein	3	115	13.07	1.48E-07
626	10	EAN96142.1 trans-sialidase	4	27	13.05	0.000148
627	4	EAN84275.1 succinate dehydrogenase	4	22	13.00	0.000249

628	6	ABD78776.1 antigenic lectin-2	4	32	13.00	0.000209
629	2	EAN98728.1 hypothetical protein	4	11	13.00	0.000193
630	2	XP_817010.1 cystathione gamma lyase	3	35	12.97	9.88E-05
631	4	XP_810730.1 hypothetical protein	4	51	12.91	2.15E-07
632	4	XP_816272.1 hypothetical protein	4	6	12.88	0.000196
633	6	XP_821910.1 trans-sialidase	4	22	12.87	0.000253
634	2	XP_808915.1 hypothetical protein	4	12	12.86	0.000132
635	26	XP_819603.1 surface protease GP63	4	14	12.82	0.000196
636	2	EAN84271.1 glutamate dehydrogenase	4	6	12.81	1.07E-05
637	4	XP_813555.1 hypothetical protein	4	9	12.80	1.55E-06
638	4	EAN82460.1 eukaryotic translation initiation factor 1A	4	25	12.80	8.48E-05
639	4	XP_809619.1 phosphatase-like protein	3	24	12.79	8.42E-07
640	2	EAN97582.1 hypothetical protein	3	13	12.75	5.21E-05
641	2	XP_807920.1 heat shock protein	3	13	12.73	2.38E-08
642	7	XP_821160.1 retrotransposon hot spot (RHS) protein	4	12	12.73	0.000283
643	4	XP_817104.1 dynein intermediate chain	3	37	12.72	5.24E-06
644	8	EAN82434.1 60S ribosomal protein L12	3	40	12.68	8.22E-14
645	11	EAN87808.1 trans-sialidase	4	108	12.66	1.48E-06
646	4	EAN80990.1 hypothetical protein	3	28	12.64	2.21E-06
647	2	XP_820432.1 hypothetical protein	4	12	12.64	0.000243
648	6	XP_809990.1 ribosomal protein S20	3	83	12.60	4.11E-11
649	2	XP_814797.1 cation-transporting ATPase	4	8	12.60	3.28E-06
650	2	EAN81171.1 hypothetical protein	3	15	12.58	2.57E-08
651	4	EAN95264.1 hypothetical protein	4	15	12.57	0.000102
652	2	EAN95004.1 trans-sialidase	3	121	12.54	1.48E-06
653	2	EAO00053.1 trans-sialidase	3	10	12.49	0.000253
654	2	EAN96031.1 trans-sialidase	4	24	12.48	0.000168
655	2	EAN89843.1 hypothetical protein	4	5	12.45	0.000456
656	2	XP_821898.1 trans-sialidase	3	14	12.45	3.89E-06
657	2	EAN99233.1 trans-sialidase	4	17	12.43	3.24E-05
658	4	XP_810600.1 trypanothione synthetase	4	9	12.37	2.24E-05
659	4	EAN99175.1 uracil phosphoribosyltransferase	3	17	12.35	0.000126
660	4	XP_810436.1 hypothetical protein	3	3	12.29	2.45E-10
661	8	EAN84133.1 ADP-ribosylation factor	4	34	12.28	0.000276
662	2	XP_814528.1 hypothetical protein	4	54	12.25	0.00015
663	4	EAN92275.1 hypothetical protein	4	8	12.24	0.000242

664	4	EAN90839.1 CCR4 associated factor	3	11	12.23	1.57E-06
665	2	XP_811688.1 trans-sialidase	4	75	12.15	9.47E-05
666	4	XP_816593.1 hypothetical protein	4	34	12.15	0.000241
667	4	XP_804146.1 hypothetical protein	4	52	12.11	2.12E-08
668	2	EAN93118.1 trans-sialidase	3	16	12.10	5.73E-07
669	2	EAN85777.1 hypothetical protein	4	4	12.09	0.000124
670	2	EAN90961.1 retrotransposon hot spot (RHS) protein	3	18	12.03	0.00042
671	4	EAN88878.1 coatomer epsilon subunit	3	18	11.92	0.000404
672	2	XP_812314.1 trans-sialidase	3	7	11.86	0.000208
673	4	XP_808772.1 small GTP-binding protein	3	16	11.83	8.83E-06
674	2	EAN87328.1 trans-sialidase	3	8	11.83	3.89E-06
675	6	XP_817770.1 40S ribosomal protein S23	3	101	11.82	2.67E-06
676	4	EAN91820.1 hypothetical protein	3	15	11.81	0.00014
677	4	XP_812637.1 RNA-binding protein	3	30	11.78	0.000344
678	4	EAN94806.1 hypothetical protein	3	61	11.73	0.000101
679	2	EAN97618.1 C-terminal motor kinesin	3	11	11.70	0.000473
680	5	EAN89960.1 ribose 5-phosphate isomerase	3	24	11.69	6.56E-07
681	12	EAN88955.1 universal minicircle sequence binding protein	3	98	11.68	2.18E-07
682	12	XP_807532.1 histone H3	3	229	11.66	3.09E-05
683	4	EAN88832.1 hypothetical protein	3	13	11.65	7.42E-10
684	4	EAN93101.1 proteasome alpha 7 subunit	4	32	11.64	3.62E-05
685	4	XP_818161.1 hypothetical protein	4	6	11.63	0.000251
686	2	XP_802450.1 leucyl-tRNA synthetase	4	20	11.62	3.91E-05
687	4	EAN87853.1 hypothetical protein	3	21	11.62	5.03E-05
688	4	EAN93695.1 histone H2B variant XP_821169.1 eukaryotic translation initiation factor 3 subunit	3	17	11.61	1.53E-07
689	4	8	4	8	11.59	5.59E-05
690	3	EAN91099.1 trans-sialidase	3	21	11.57	3.07E-05
691	4	EAN95156.1 trans-sialidase	3	7	11.55	0.000208
692	2	EAN97642.1 kinesin	3	6	11.55	0.000277
693	4	EAN96705.1 hypothetical protein	3	9	11.53	7.64E-05
694	4	EAN82739.1 hypothetical protein	3	46	11.50	7.99E-11
695	4	EAN88100.1 mucin-associated surface protein (MASP)	3	12	11.48	5.58E-05
696	2	XP_822022.1 trans-sialidase	3	20	11.48	0.000148
697	4	XP_804941.1 hypothetical protein	3	9	11.47	0.000326
698	2	EAN98483.1 beta-adaptin	3	7	11.45	2.99E-06

699	4	EAN85389.1 basic transcription factor 3a	3	60	11.40	5.01E-06
700	2	EAN83374.1 proteasome regulatory non-ATPase subunit 6	3	12	11.38	3.43E-07
701	6	XP_812573.1 hypothetical protein	4	12	11.34	6.01E-06
702	2	EAN97107.1 gamma-adaptin 1	3	10	11.33	1.53E-07
703	9	EAN88266.1 histone H2A	4	304	11.29	2.20E-05
704	5	EAN87356.1 adenylate kinase	3	21	11.18	0.000259
705	2	EAN81636.1 hypothetical protein	3	6	11.10	3.18E-06
706	6	XP_819195.1 60S ribosomal protein L22	3	33	11.08	0.000136
707	5	EAN95591.1 calcium motive p-type ATPase	3	6	11.05	8.24E-06
708	6	EAN83666.1 cytosolic malate dehydrogenase	3	29	10.93	0.000431
709	4	EAN92877.1 N-myristoyl transferase	3	10	10.92	0.000413
710	11	XP_813685.1 cysteine peptidase inhibitor	3	43	10.92	3.97E-05
711	4	XP_804061.1 aspartyl aminopeptidase	3	13	10.91	0.000422
712	1	XP_820307.1 trans-sialidase	3	16	10.89	0.000148
713	7	EAN81731.1 retrotransposon hot spot (RHS) protein	3	19	10.88	3.56E-05
714	2	EAN83062.1 retrotransposon hot spot (RHS) protein	3	22	10.85	3.56E-05
715	2	EAN97996.1 hypothetical protein	3	4	10.83	6.15E-10
716	4	XP_812387.1 hypothetical protein	3	7	10.74	0.000264
717	2	EAN96967.1 hypothetical protein	3	40	10.74	1.60E-05
718	4	EAN86245.1 proteasome alpha 3 subunit	3	27	10.74	0.000363
719	4	EAN95731.1 ADP-ribosylation factor-like protein	2	8	10.71	1.37E-09
720	2	XP_818412.1 adenine phosphoribosyltransferase	3	3	10.70	1.44E-05
721	2	EAN96183.1 cystathione gamma lyase	3	13	10.69	9.88E-05
722	6	XP_805718.1 trans-sialidase	3	13	10.68	0.000137
723	2	XP_804553.1 trans-sialidase	3	13	10.66	9.42E-05
724	2	EAN90780.1 P-type H+-ATPase	3	8	10.63	0.000192
725	2	EAN81531.1 hypothetical protein	3	9	10.60	3.16E-06
726	4	XP_816271.1 hypothetical protein	3	12	10.58	6.41E-11
727	2	XP_820907.1 aminopeptidase	3	14	10.56	3.94E-07
728	4	EAN82998.1 alkyl-dihydroxyacetone phosphate synthase	3	19	10.55	2.85E-06
729	5	EAN98181.1 eukaryotic translation initiation factor 6 (eIF-6)	3	19	10.54	7.31E-05
730	4	EAN91009.1 hypothetical protein	3	16	10.50	0.000114
731	9	XP_804791.1 40S ribosomal protein S15	2	117	10.50	3.48E-12
732	2	EAN85473.1 tyrosine aminotransferase	3	23	10.48	0.000162
733	4	XP_821303.1 trans-sialidase	3	31	10.45	9.16E-05
734	2	EAN90047.1 orotidine-5-phosphate decarboxylase	3	7	10.44	8.67E-07

735	4	EAN83972.1 major vault protein	3	6	10.43	0.000229
736	9	EAN87371.1 60S acidic ribosomal protein P2	3	18	10.41	0.000172
737	2	EAN99081.1 protein transport protein sec13	3	29	10.38	2.69E-06
738	4	EAN86488.1 hypothetical protein	3	12	10.36	0.000249
739	2	XP_821108.1 proteasome regulatory non-ATPase subunit 3	3	17	10.31	4.15E-07
740	1	AAM47176.1 surface glycoprotein GP90	3	11	10.28	0.000333
741	2	XP_809098.1 hypothetical protein	3	7	10.27	8.96E-05
742	2	EAN85439.1 hypothetical protein	3	11	10.19	0.000381
743	4	EAN95592.1 kinesin	3	6	10.19	0.00037
744	2	XP_815119.1 trans-sialidase	3	31	10.18	0.000339
745	8	XP_810882.1 retrotransposon hot spot (RHS) protein	3	15	10.17	0.000125
746	2	XP_806769.1 ATPase alpha subunit	3	14	10.13	0.000358
747	4	EAN98033.1 fibrillarin	3	11	10.11	1.93E-05
748	2	EAN97710.1 retrotransposon hot spot (RHS) protein	3	13	10.11	1.12E-06
749	4	XP_805178.1 hypothetical protein	3	10	10.11	0.000171
750	2	EAN96369.1 trans-sialidase	3	13	10.10	3.47E-05
751	2	XP_821405.1 trans-sialidase	3	33	10.08	0.000382
752	4	EAN89251.1 protein phosphatase 2A, regulatory subunit B	3	5	10.08	5.72E-05
753	4	XP_805107.1 hypothetical protein	3	7	10.05	0.000119
754	2	EAN90301.1 trans-sialidase	3	39	10.05	0.000148
755	3	AAQ74641.1 c71 surface protein	3	34	10.01	0.000148
756	4	XP_804535.1 aminopeptidase P1	3	6	10.01	7.55E-07
757	5	EAN81294.1 ubiquitin-conjugating enzyme E2	3	11	9.99	1.38E-05
758	5	EAN97712.1 trans-sialidase	3	168	9.99	4.79E-05
759	4	XP_816162.1 hypothetical protein	3	11	9.98	0.000489
760	3	XP_821148.1 trans-sialidase	3	3	9.98	0.000148
761	4	XP_810458.1 nuclear transport factor 2	3	23	9.97	1.77E-05
762	4	XP_814323.1 hypothetical protein	3	3	9.93	0.000231
763	4	XP_806666.1 ABC transporter	3	5	9.87	4.48E-09
764	8	EAN86950.1 elongation factor 1-gamma (EF-1-gamma)	2	64	9.83	5.76E-09
765	2	XP_817172.1 hypothetical protein	3	7	9.74	2.82E-05
766	6	EAN82837.1 trans-sialidase	3	107	9.73	4.38E-05
767	4	EAN87921.1 peptidyl-prolyl cis-trans isomerase	3	20	9.71	3.83E-06
768	2	EAN93852.1 reticulon domain protein	2	21	9.71	3.42E-10
769	4	EAN96438.1 heat shock protein DnaJ	2	4	9.70	2.12E-07
770	4	XP_812251.1 hypothetical protein	3	17	9.69	0.000148

771	5	EAN87626.1 immunodominant antigen	2	6	9.58	2.93E-07
772	2	XP_807802.1 trans-sialidase	3	19	9.58	0.000208
773	4	EAN86445.1 hypothetical protein	3	28	9.55	1.53E-07
774	2	EAN95467.1 hypothetical protein	2	6	9.55	5.47E-05
775	4	EAN88829.1 hydrolase, alpha/beta fold family	3	13	9.54	1.96E-06
776	4	EAN90796.1 eukaryotic translation initiation factor	3	15	9.51	0.000464
777	4	XP_807058.1 hypothetical protein	3	9	9.49	4.08E-07
778	2	EAN90948.1 hypothetical protein	3	4	9.48	1.36E-05
779	5	EAN97353.1 hslvu complex proteolytic subunit-like	3	32	9.46	9.70E-07
780	4	XP_808258.1 hypothetical protein	3	10	9.43	3.39E-06
781	2	XP_806266.1 retrotransposon hot spot (RHS) protein	3	7	9.43	0.000283
782	2	EAN89453.1 trans-sialidase	3	23	9.41	1.42E-05
783	4	EAN85770.1 hypothetical protein	2	14	9.39	1.76E-09
784	2	EAN99861.1 trans-sialidase	2	21	9.30	9.16E-05
785	2	XP_815947.1 hypothetical protein	3	13	9.29	1.56E-06
786	4	EAN92482.1 radial spoke protein 3	2	3	9.28	0.000472
787	2	XP_805085.1 hypothetical protein	3	4	9.28	0.000242
788	2	EAN93392.1 ribonuclease L inhibitor	3	11	9.27	0.000342
789	2	XP_803096.1 trans-sialidase	3	109	9.25	0.000333
790	4	XP_808797.1 hypothetical protein	2	8	9.23	7.12E-08
791	4	EAN81872.1 hypothetical protein	2	10	9.23	1.94E-05
792	10	EAN90461.1 mucin-associated surface protein (MASP)	2	23	9.21	4.24E-10
793	12	XP_820496.1 mucin-associated surface protein (MASP)	2	5	9.21	2.18E-07
794	4	EAN89989.1 NAD(P)-dependent oxidoreductase	2	5	9.20	1.90E-05
795	2	EAN97577.1 hypothetical protein	3	11	9.16	3.08E-05
796	4	EAN84353.1 hypothetical protein	3	4	9.16	9.11E-05
797	2	EAN98372.1 ubiquitin hydrolase	3	3	9.15	3.76E-07
798	2	XP_818877.1 hypothetical protein	3	14	9.13	1.93E-06
799	2	EAN99871.1 elongation factor 1-gamma (EF-1-gamma)	2	64	9.13	6.74E-09
800	4	XP_808947.1 tyrosyl or methionyl-tRNA synthetase	3	7	9.10	0.000172
		XP_820928.1 succinyl-coA:3-ketoacid-coenzyme A				
801	4	transferase	3	11	9.08	3.82E-05
802	4	XP_819852.1 hypothetical protein	3	6	9.04	6.40E-05
803	2	EAN97441.1 hypothetical protein	3	23	9.03	4.74E-05
804	2	AAK97605.1 calreticulin protein	2	132	9.03	3.71E-06
805	2	EAN98772.1 diphosphomevalonate decarboxylase	3	5	9.02	0.000136

806	4	EAN89729.1 nudix hydrolase	2	7	9.00	5.38E-07
807	2	EAN97038.1 trans-sialidase	3	9	8.98	0.000414
808	4	XP_811429.1 hypothetical protein	2	37	8.93	2.18E-07
809	2	EAN88532.1 trans-sialidase	3	16	8.93	0.000148
810	4	EAN99369.1 pretranslocation protein, alpha subunit	3	10	8.92	2.62E-05
811	2	EAN82626.1 mucin-associated surface protein (MASP)	3	8	8.88	0.000335
812	4	XP_819548.1 mucin-associated surface protein (MASP)	2	7	8.87	3.16E-06
813	3	XP_813500.1 trans-sialidase	3	18	8.87	0.000148
814	4	XP_809789.1 nucleosome assembly protein	3	7	8.87	2.27E-06
815	4	EAN91985.1 phosphoacetylglucosamine mutase	2	3	8.85	0.000293
816	2	XP_820228.1 importin alpha	2	2	8.83	3.84E-09
817	2	XP_806774.1 trans-sialidase	3	19	8.81	0.000414
818	4	XP_805662.1 coatomer delta subunit	2	5	8.79	3.79E-08
819	4	EAN81315.1 hypothetical protein	2	3	8.78	7.49E-05
820	2	XP_813112.1 kinesin-like protein	3	5	8.78	0.000224
821	6	EAN86741.1 surface protease GP63	2	4	8.77	7.10E-10
822	2	XP_822026.1 trans-sialidase	2	4	8.73	0.000208
823	9	CAM33602.1 putative surface antigen YASP-A1	2	22	8.71	4.03E-07
824	4	XP_816004.1 ATP synthase	2	10	8.69	5.02E-06
825	2	EAN83682.1 trans-sialidase	3	7	8.68	1.73E-05
826	2	EAO00040.1 vacuolar ATP synthase subunit c	2	7	8.66	4.52E-05
827	4	EAN86186.1 hypothetical protein	2	21	8.62	8.30E-06
828	2	XP_818128.1 hypothetical protein	2	9	8.54	2.27E-07
829	4	EAN83415.1 hypothetical protein	3	3	8.51	0.000212
830	2	XP_821812.1 ribosomal protein L35A	3	37	8.47	8.88E-05
831	2	EAN99792.1 60S ribosomal protein L27A/L29	2	72	8.46	0.000103
833	4	XP_810188.1 hypothetical protein	2	9	8.45	2.10E-09
832	4	EAN84807.1 hypothetical protein	2	3	8.45	1.17E-08
834	78	EAN80921.1 UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase	3	20	8.44	7.43E-05
835	2	EAN96743.1 hypothetical protein	2	15	8.43	5.13E-05
836	2	EAN93956.1 trans-sialidase	3	7	8.41	0.000382
837	4	XP_818872.1 hypothetical protein	3	7	8.41	6.96E-06
838	8	XP_811573.1 60S ribosomal protein L34	2	58	8.38	1.19E-08
839	4	EAN98447.1 hypothetical protein	2	21	8.37	2.74E-09
840	2	EAN86319.1 hypothetical protein	3	22	8.37	0.000302

841	5	AAM66350.1 U2 splicing auxiliary factor; U2AF26	2	28	8.36	1.29E-05
842	4	EAN83110.1 hypothetical protein	3	6	8.35	0.000185
843	4	EAN86541.1 hypothetical protein	2	9	8.32	5.88E-06
844	2	EAN85671.1 ribonucleoprotein p18, mitochondrial precursor	2	13	8.31	1.03E-11
845	4	EAN90739.1 hypothetical protein	2	7	8.26	3.26E-06
846	4	EAN81810.1 (H+)-ATPase G subunit	2	14	8.26	7.88E-07
847	3	EAN99952.1 hypothetical protein	2	5	8.23	0.000113
848	4	EAN95271.1 hypothetical protein	2	7	8.22	2.71E-06
849	2	XP_806981.1 ubiquitin hydrolase	2	6	8.21	1.75E-05
850	6	XP_810911.1 mucin-associated surface protein (MASP)	2	18	8.21	6.05E-05
851	4	XP_810047.1 hypothetical protein	3	3	8.20	0.000483
852	4	XP_808998.1 hypothetical protein	2	2	8.19	5.01E-06
853	4	XP_806067.1 glyceraldehyde-3-phosphate dehydrogenase	2	9	8.17	1.80E-05
854	2	XP_809105.1 dynein heavy chain	3	7	8.16	6.63E-05
855	2	EAN89827.1 trans-sialidase	2	4	8.14	0.000108
856	2	EAN97971.1 hypothetical protein	2	10	8.13	7.64E-07
857	5	XP_812541.1 splicing factor TSR1	2	18	8.11	0.000348
858	4	XP_804030.1 chaperonin HSP60/CNP60	2	121	8.05	0.000161
859	2	EAN91734.1 trans-sialidase	3	11	8.03	0.000333
860	2	EAN96403.1 dynein arm light chain	3	5	8.00	0.00014
861	4	EAN85964.1 mitogen-activated protein kinase 3	2	9	7.99	1.10E-06
862	4	XP_806974.1 40S ribosomal protein S15a	2	38	7.98	1.90E-06
863	4	XP_819071.1 cytochrome b5	2	5	7.96	1.41E-05
864	4	XP_819757.1 hypothetical protein	2	7	7.95	6.77E-07
865	2	XP_821444.1 surface protease GP63	2	7	7.92	7.10E-10
866	4	EAN99237.1 trans-sialidase	2	4	7.90	0.000434
867	2	EAN89559.1 hypothetical protein	2	7	7.89	4.71E-06
868	5	AAR05659.1 translation initiation factor 34	2	4	7.85	0.000186
869	5	AAQ55215.1 21 kDa cyclophilin	3	13	7.84	2.49E-05
870	10	CAA52943.1 TcP2beta	2	60	7.83	4.78E-06
871	2	EAN99746.1 ABC transporter	2	5	7.83	5.45E-06
872	41	AAX40341.1 dihydrofolate reductase-thymidylate synthase	2	6	7.82	2.74E-08
873	2	EAN87078.1 hypothetical protein	2	15	7.76	1.53E-08
874	4	EAN97075.1 acetyltransferase	2	7	7.74	5.06E-06
875	2	EAN88604.1 lysyl-tRNA synthetase	2	3	7.74	0.000217
876	2	XP_817331.1 endomembrane protein	2	9	7.73	2.24E-09

877	2	EAO00234.1 hypothetical protein	2	7	7.72	1.77E-07
878	2	XP_814909.1 hypothetical protein	2	9	7.69	4.40E-09
879	5	EAN97907.1 lanosterol synthase	2	7	7.69	1.03E-07
880	2	XP_821629.1 hypothetical protein	2	2	7.68	1.42E-05
881	4	EAN90294.1 eukaryotic translation initiation factor 3	2	7	7.67	2.64E-08
882	5	AAC97957.2 proteasome beta 5 subunit	2	16	7.66	1.35E-06
883	30	2JJ2 F1-Atpase	2	57	7.66	1.26E-06
884	4	EAN84527.1 hypothetical protein	2	8	7.65	6.36E-05
885	2	XP_807717.1 hypothetical protein	2	2	7.65	0.00014
886	4	EAN95569.1 small ubiquitin protein	2	37	7.64	3.61E-07
887	2	XP_805281.1 hypothetical protein	2	2	7.64	3.35E-11
888	4	XP_806373.1 hypothetical protein	2	8	7.64	0.000196
889	4	EAN90347.1 phosphoribosylpyrophosphate synthetase	2	2	7.63	3.70E-08
890	3	XP_808357.1 trans-sialidase	2	36	7.60	1.30E-08
891	2	XP_817914.1 cytochrome c1, heme protein	2	14	7.59	4.21E-07
892	64	XP_805990.1 rab1 small GTP-binding protein	2	5	7.59	0.000337
893	2	EAN85114.1 mucin-associated surface protein (MASP)	2	26	7.56	1.72E-05
894	18	XP_818579.1 cruzipain precursor	2	15	7.53	1.05E-06
895	4	XP_810328.1 hypothetical protein	2	9	7.52	1.39E-05
896	4	EAN94957.1 hypothetical protein	2	4	7.51	2.70E-05
897	4	EAN90955.1 glutathione peroxidase-like protein	2	27	7.49	8.25E-07
898	4	XP_812938.1 ATP-dependent zinc metalloproteinase	2	8	7.48	1.39E-09
899	2	XP_810858.1 hypothetical protein	2	6	7.48	0.000183
900	4	EAN92104.1 acyl-CoA dehydrogenase	2	4	7.48	9.67E-06
901	2	EAN82251.1 dynein heavy chain	2	5	7.47	0.000159
902	4	EAN85198.1 hypothetical protein	2	11	7.46	9.66E-10
903	2	EAN81889.1 mucin-associated surface protein (MASP)	2	10	7.44	2.58E-08
904	2	XP_809096.1 retrotransposon hot spot (RHS) protein	2	2	7.44	2.77E-05
905	4	EAN93546.1 protein tyrosine phosphatase-like protein	2	7	7.43	5.67E-06
906	4	XP_820787.1 hypothetical protein	2	5	7.43	5.51E-09
907	2	XP_821826.1 antigenic protein	3	3	7.43	2.46E-06
908	4	XP_816576.1 mitochondrial RNA binding protein 1	2	17	7.42	1.54E-08
910	4	XP_818969.1 CAS/CSE/importin domain protein	2	4	7.41	2.68E-05
909	2	EAN95287.1 hypothetical protein	2	21	7.41	7.67E-06
911	4	XP_816085.1 hypothetical protein	2	3	7.40	0.000131
912	2	EAN95469.1 ran-binding protein 1	2	28	7.39	2.15E-05

913	6	EAN87502.1 leucine-rich repeat protein	2	6	7.39	3.42E-08
914	4	EAN88810.1 peptidyl-prolyl cis-trans isomerase/rotamase	2	3	7.34	0.000348
915	2	EAN96434.1 hypothetical protein	2	19	7.33	0.000136
916	4	XP_804162.1 eukaryotic translation initiation factor 2 subunit	2	4	7.33	4.20E-06
917	4	EAN80832.1 hypothetical protein	2	2	7.29	5.95E-06
918	2	EAN82215.1 trans-sialidase	2	6	7.28	0.00011
919	4	EAN90785.1 fructose-1,6-bisphosphatase, cytosolic	2	5	7.26	2.49E-07
920	4	EAN85226.1 histone H2A	2	23	7.23	0.000209
921	4	EAN90557.1 hypothetical protein	2	9	7.23	0.00039
922	2	XP_802801.1 hypothetical protein	2	5	7.21	2.95E-07
923	5	EAN89381.1 adenylate kinase	2	7	7.20	5.02E-06
924	2	EAN81369.1 3-oxo-5-alpha-steroid 4-dehydrogenase	2	17	7.15	0.000122
925	2	EAN88220.1 retrotransposon hot spot (RHS) protein	2	12	7.13	3.56E-05
926	4	XP_822077.1 hypothetical protein	2	2	7.09	7.39E-06
		EAN99492.1 heat shock protein HslVU, ATPase subunit				
927	4	HslU	2	6	7.08	3.13E-07
928	4	XP_816749.1 hypothetical protein	2	87	7.06	1.05E-07
929	4	EAN87281.1 hypothetical protein	2	13	7.06	2.17E-05
930	2	XP_802363.1 dynein heavy chain	2	7	7.06	5.68E-06
931	2	EAN97979.1 hypothetical protein	2	7	7.04	0.000117
932	4	XP_810897.1 calcium-binding protein	2	4	7.02	3.18E-05
933	3	XP_805177.1 flagellum-adhesion glycoprotein	2	2	7.02	3.37E-05
934	4	XP_811750.1 variant-surface-glycoprotein phospholipase C	2	4	7.00	9.41E-09
935	4	EAN86794.1 mucin-associated surface protein (MASP)	2	4	6.99	0.000303
936	4	EAN92004.1 40S ribosomal protein S12	2	13	6.98	8.37E-06
937	4	XP_815409.1 mucin TcMUCII	2	20	6.98	3.30E-07
938	4	EAN96511.1 prefoldin	2	4	6.97	2.97E-05
939	4	XP_812859.1 coatomer zeta subunit	2	7	6.96	9.88E-06
940	2	EAN92700.1 hypothetical protein	2	4	6.94	3.27E-05
941	2	EAN89980.1 hypothetical protein	2	5	6.93	3.74E-05
942	2	EAN95648.1 mucin-associated surface protein (MASP)	2	4	6.92	0.000185
943	2	EAN89395.1 histidyl-tRNA synthetase	2	4	6.92	7.54E-05
944	4	XP_813288.1 aminopeptidase P	2	6	6.89	2.80E-05
945	2	EAN94184.1 hypothetical protein	2	23	6.87	4.43E-08
946	2	XP_804574.1 glutamate dehydrogenase	2	6	6.86	7.01E-05
947	2	XP_821487.1 hypothetical protein	2	7	6.86	2.52E-05

948	4	EAN82245.1 trans-sialidase	2	3	6.86	6.71E-06
949	4	EAN82772.1 hypothetical protein	2	4	6.84	0.000121
950	2	XP_819415.1 hypothetical protein	2	3	6.80	6.71E-05
951	2	XP_806393.1 hypothetical protein	2	3	6.80	0.000279
952	2	EAN85254.1 hypothetical protein	2	4	6.78	4.90E-05
		EAN95207.1 heat shock protein HslVU, ATPase subunit				
953	4	HslU	2	4	6.77	9.39E-05
954	4	XP_813206.1 hypothetical protein	2	3	6.76	3.28E-08
955	2	XP_802960.1 glucose-6-phosphate dehydrogenase	2	7	6.76	7.72E-06
		XP_804977.1 protein kinase C substrate protein, heavy				
956	4	chain	2	13	6.73	5.87E-06
957	4	XP_819056.1 hypothetical protein	2	4	6.73	0.000309
958	4	XP_818903.1 hypothetical protein	2	5	6.72	7.30E-05
959	4	EAN95389.1 hypothetical protein	1	8	6.71	8.88E-15
960	4	EAN89233.1 hypothetical protein	2	6	6.70	0.000389
962	6	EAO00019.1 ribosomal protein S27	2	22	6.68	2.56E-06
961	2	EAN87335.1 cytochrome c	2	61	6.68	0.000132
963	2	EAN82650.1 trans-sialidase	2	2	6.67	1.35E-08
964	2	XP_811529.1 myo-inositol-1 phosphatase	2	7	6.67	0.000229
965	2	XP_815658.1 surface protease GP63	2	6	6.67	1.63E-05
966	2	XP_817687.1 50S ribosomal protein L7Ae	2	8	6.66	1.82E-05
967	2	EAN95853.1 hypothetical protein	2	3	6.66	0.000263
968	4	EAN96186.1 ubiquitin-activating enzyme E1	2	4	6.64	9.60E-05
969	4	EAN98853.1 proteasome regulatory non-ATPase subunit 7	2	5	6.63	8.00E-05
970	2	XP_807467.1 hypothetical protein	2	4	6.63	0.000183
971	4	XP_814511.1 cyclophilin	2	3	6.63	0.00018
972	4	XP_820364.1 hypothetical protein	2	4	6.62	5.26E-06
973	4	EAN81954.1 ubiquitin	2	5	6.60	1.80E-05
974	8	EAN87212.1 aspartate carbamoyltransferase	2	3	6.60	6.45E-06
975	4	XP_809257.1 hypothetical protein	2	2	6.59	0.000286
976	4	XP_812132.1 small nuclear ribonucleoprotein	2	7	6.59	1.74E-07
		EAN91546.1 hypoxanthine-guanine				
977	4	phosphoribosyltransferase	2	17	6.58	6.12E-05
978	4	XP_806068.1 hypothetical protein	2	4	6.57	3.82E-07
979	6	XP_822059.1 hypothetical protein	2	4	6.54	9.58E-06
		XP_816194.1 eukaryotic peptide chain release factor				
980	2	subunit 1	2	6	6.54	4.55E-05

981	4	XP_803006.1 trans-sialidase	2	17	6.53	7.71E-06
982	4	EAN83669.1 dynein heavy chain	2	2	6.53	9.08E-07
983	2	EAN95901.1 hypothetical protein	2	6	6.48	6.80E-05
984	4	XP_813436.1 hypothetical protein	2	3	6.48	2.89E-05
985	2	EAN82710.1 glutamamyl carboxypeptidase	2	3	6.47	3.79E-06
986	2	EAN97438.1 phenylalanyl-tRNA synthetase alpha chain	2	13	6.47	9.77E-06
987	4	XP_819866.1 hypothetical protein	2	2	6.46	0.000451
988	4	XP_815782.1 hypothetical protein	2	3	6.44	1.50E-07
989	4	XP_812363.1 hypothetical protein	2	7	6.44	5.37E-06
990	4	EAN82147.1 surface protease GP63	2	9	6.43	2.03E-05
991	2	EAN99763.1 hypothetical protein	2	3	6.42	0.000158
992	4	EAN91323.1 peptidyl-prolyl cis-trans isomerase	2	3	6.40	0.000116
993	4	XP_814235.1 ADG1-like protein	2	7	6.39	8.29E-05
994	4	EAN86754.1 hypothetical protein	2	8	6.36	0.000407
995	4	XP_804237.1 cysteinyl-tRNA synthetase	2	2	6.36	0.000371
996	2	EAN91495.1 dynein-associated protein	2	14	6.34	5.19E-06
997	4	EAN93755.1 dynein heavy chain	2	5	6.33	0.000388
998	2	XP_821107.1 hypothetical protein	2	2	6.33	0.0002
999	4	EAN99425.1 electron transfer flavoprotein-ubiquinone oxidoreductase	2	4	6.32	1.71E-06
1000	15	2E6D Dihydroorotate Dehydrogenase	2	6	6.32	9.38E-05
1001	2	BAF74646.1 glycoprotein 82 kDa	2	106	6.29	0.000264
1002	6	XP_819680.1 hypothetical protein	2	4	6.29	3.60E-05
1003	2	EAN85368.1 hypothetical protein	2	5	6.28	3.78E-08
1004	2	XP_819351.1 NADH-cytochrome B5 reductase	2	5	6.27	6.79E-06
1005	4	XP_810971.1 hypothetical protein	2	8	6.26	1.06E-06
1006	4	EAN81082.1 mevalonate kinase	2	20	6.24	0.000149
1007	6	XP_809993.1 calpain-like cysteine peptidase	2	28	6.23	0.000211
1008	2	EAN99092.1 hypothetical protein	2	3	6.21	4.21E-07
1009	4	XP_815689.1 replication factor C, subunit 2	2	4	6.21	2.48E-05
1010	4	EAN97314.1 hypothetical protein	2	2	6.20	5.81E-05
1011	2	XP_815548.1 haloacid dehalogenase hydrolase	2	3	6.20	0.000337
1012	2	XP_807657.1 retrotransposon hot spot (RHS) protein	2	12	6.19	0.000125
1013	2	XP_807727.1 cation-transporting ATPase	2	3	6.18	3.77E-06
1014	2	XP_804558.1 hypothetical protein	2	7	6.18	8.18E-05
1015	8	EAN92496.1 60S ribosomal protein L17	2	3	6.17	1.59E-05

1016	2	XP_816205.1 guanylate kinase	2	2	6.16	0.000443
1017	4	EAN94085.1 hypothetical protein	2	4	6.16	6.33E-06
1018	4	EAN93958.1 surface protease GP63	2	6	6.15	0.000134
1019	4	EAN83304.1 hypothetical protein	2	3	6.15	0.000496
1020	5	EAN96724.1 hydroxyacylglutathione hydrolase	2	10	6.13	0.00029
1021	4	XP_815388.1 hypothetical protein	2	4	6.11	2.20E-05
1022	2	XP_820534.1 hypothetical protein	2	2	6.11	0.000121
1023	4	XP_807620.1 hypothetical protein	2	24	6.10	1.17E-05
1024	8	XP_816170.1 serine/threonine protein kinase	2	4	6.09	5.16E-05
1025	2	EAN99582.1 mucin-associated surface protein (MASP)	2	6	6.08	0.000234
1026	2	EAN96121.1 RNA-binding protein	2	5	6.05	0.000463
1027	4	XP_821166.1 cytochrome c oxidase subunit IX	2	6	6.04	4.38E-05
1028	2	EAN96932.1 hypothetical protein	2	2	6.02	0.000295
1029	8	XP_816521.1 ribosomal protein L38	2	9	6.01	7.11E-05
1030	2	XP_812224.1 hypothetical protein	2	14	6.01	2.21E-05
1031	4	XP_809934.1 hypothetical protein	1	17	6.01	2.51E-11
1032	4	XP_812881.1 hypothetical protein	2	8	6.00	5.27E-05
1033	4	EAN89972.1 stress-inducible protein STI1-like	2	6	6.00	0.000131
1034	10	XP_815869.1 mucin TcMUCII	1	4	5.96	1.94E-11
1035	4	EAN84535.1 hypothetical protein	2	3	5.94	4.36E-06
1036	4	EAN99888.1 AAA ATPase	2	2	5.94	5.55E-05
1037	2	EAN87076.1 RNA guanylyltransferase	2	5	5.93	0.000139
1038	4	XP_821272.1 hypothetical protein	2	7	5.92	0.00015
1039	4	EAN88260.1 hypothetical protein	2	7	5.91	1.13E-05
1040	2	EAN95786.1 trans-sialidase	2	9	5.91	0.000297
1041	4	EAN95980.1 cytochrome c oxidase VII	2	7	5.90	7.11E-06
1042	2	EAN97166.1 hypothetical protein	2	2	5.89	0.0003
1043	4	XP_809458.1 eukaryotic translation initiation factor eIF-4E	2	7	5.89	4.99E-06
1044	2	XP_816002.1 mucin TcMUCII	1	7	5.89	1.15E-11
1045	4	EAN85242.1 hypothetical protein	2	2	5.87	1.82E-10
1046	4	EAN89743.1 hypothetical protein	2	7	5.86	3.24E-06
1047	4	XP_804157.1 hypothetical protein	2	4	5.86	0.000395
1048	4	EAN80828.1 ATP-dependent DEAD/H RNA helicase	2	12	5.84	8.47E-06
1049	4	XP_820239.1 hypothetical protein	2	4	5.84	5.79E-05
1050	2	XP_815593.1 hypothetical protein	2	5	5.84	2.11E-07
1051	4	EAN96852.1 surface protease GP63	2	5	5.83	2.03E-05

1052	2	EAN95448.1 hypothetical protein	1	2	5.83	1.34E-09
1053	2	XP_804263.1 retrotransposon hot spot (RHS) protein	2	8	5.81	0.000125
1054	4	EAN85116.1 translation initiation factor IF-2	2	3	5.81	0.00018
1055	2	XP_818674.1 hypothetical protein	2	6	5.80	4.96E-05
1056	2	EAN81944.1 lysosomal/endosomal membrane protein p67	2	6	5.79	1.12E-05
1057	4	EAN89198.1 proteasome regulatory ATPase subunit 1	2	5	5.78	1.73E-05
1058	2	XP_807600.1 hypothetical protein	2	28	5.77	0.000258
1060	2	XP_808358.1 retrotransposon hot spot (RHS) protein	2	14	5.76	4.14E-06
1059	2	EAN93571.1 endo-beta-N-acetylglucosaminidase	2	2	5.76	0.0004
1061	2	XP_807331.1 hypothetical protein	2	8	5.75	0.000422
1062	2	EAN92082.1 hypothetical protein	2	2	5.75	4.73E-06
1063	2	EAN99638.1 COP-coated vesicle membrane protein erv25	2	7	5.74	0.00034
1064	2	XP_802837.1 cystathione gamma lyase	2	7	5.74	1.63E-05
1065	2	EAN98401.1 dTDP-glucose 4_6-dehydratase	2	2	5.73	0.000303
		AAC61849.1 adenylyl cyclase; ATP pyrophosphate lyase;				
1066	1	AC1	2	4	5.71	7.87E-05
1067	14	XP_818853.1 retrotransposon hot spot (RHS) protein	2	2	5.70	0.000395
1068	4	EAN84486.1 oligopeptidase B	2	2	5.67	0.000139
1069	2	XP_817757.1 hypothetical protein	2	6	5.65	0.000205
1070	4	EAN87781.1 haloacid dehalogenase-like hydrolase	2	3	5.64	5.05E-05
1071	2	XP_814806.1 peptide methionine sulfoxide reductase	2	16	5.64	8.93E-06
1072	2	XP_817389.1 trans-sialidase	2	5	5.59	0.000339
1073	2	EAN82013.1 tyrosine aminotransferase	2	13	5.57	0.000162
1074	6	EAN97360.1 ribosomal protein L29	2	9	5.56	2.85E-05
1075	2	EAN94854.1 hypothetical protein	2	6	5.56	0.000263
1076	5	AAL29186.1 poly-zinc finger protein 2	2	5	5.53	6.34E-05
1077	2	EAN89400.1 hypothetical protein	2	3	5.53	0.000131
1078	4	EAN90233.1 hypothetical protein	2	4	5.52	7.18E-05
1079	4	EAN84647.1 hypothetical protein	2	2	5.52	0.000272
1080	2	XP_811641.1 hypothetical protein	2	2	5.51	0.000131
1081	2	XP_813370.1 protein transport protein Sec23	2	2	5.48	0.000178
1082	4	XP_819121.1 hypothetical protein	2	5	5.48	3.21E-05
1083	2	XP_820486.1 mucin-associated surface protein (MASP)	1	6	5.45	4.03E-10
1084	4	EAN83132.1 hypothetical protein	1	8	5.42	4.88E-14
1085	2	XP_818542.1 small GTP-binding protein	1	21	5.37	1.19E-09
1086	2	XP_809660.1 hypothetical protein	1	4	5.34	1.77E-09

1087	2	XP_802199.1 dynein heavy chain	2	4	5.31	0.000113
1088	4	EAN88059.1 mucin-associated surface protein (MASP)	1	2	5.31	2.23E-13
1089	6	EAN91497.1 60S ribosomal protein L26	2	37	5.31	6.11E-05
1090	4	EAN98756.1 RNA-binding protein	2	5	5.29	4.79E-05
1091	6	AAC72969.1 lysosomal alpha mannosidase	1	6	5.28	4.21E-06
1092	4	EAN92024.1 mucin-associated surface protein (MASP)	1	4	5.27	1.63E-07
1093	6	EAN99457.1 mucin TcMUCII	1	3	5.26	2.97E-12
1094	4	XP_805423.1 centrin	1	9	5.26	2.77E-12
1095	2	XP_817411.1 tyrosyl-tRNA synthetase	1	1	5.25	8.92E-06
1096	2	EAN87580.1 cytosolic leucyl aminopeptidase	1	3	5.24	2.30E-12
1097	2	EAN92727.1 hypothetical protein	1	4	5.24	1.49E-10
1098	2	EAN87059.1 hypothetical protein	2	78	5.20	0.000294
1099	10	XP_819932.1 hypothetical protein	1	22	5.20	1.79E-12
1100	4	EAN83814.1 COP-coated vesicle membrane protein gp25L	1	20	5.19	1.84E-09
1101	2	XP_820382.1 trans-sialidase	1	113	5.15	1.03E-06
1102	10	EAN90340.1 mannosyl-oligosaccharide 1,2-alpha-mannosidase IB	1	2	5.11	4.40E-08
1103	2	EAN87195.1 ribonuclease mar1	1	5	5.09	1.65E-10
1104	2	EAN98589.1 hypothetical protein	2	2	5.06	0.000247
1105	2	EAN97092.1 hypothetical protein	1	8	5.06	1.70E-10
1106	4	EAN83552.1 hypothetical protein	1	2	5.00	9.99E-08
1107	2	XP_812014.1 RNA polymerase B subunit RPB8	1	2	4.97	9.02E-05
1108	2	XP_820626.1 hypothetical protein	1	1	4.97	5.60E-11
1109	8	EAN87527.1 protein phosphatase 2A catalytic subunit	1	1	4.94	0.000155
1110	4	EAN90324.1 hypothetical protein	1	5	4.88	4.28E-07
1111	2	EAN86401.1 hypothetical protein	1	1	4.87	8.46E-06
1112	2	XP_819424.1 vacuolar protein sorting-associated protein 35	1	1	4.87	6.40E-11
1113	2	EAN88285.1 hypothetical protein	1	3	4.85	1.15E-08
1114	9	EAN99145.1 60S ribosomal protein L19	1	15	4.85	2.72E-12
1115	2	EAN93040.1 ATP-binding cassette protein	2	3	4.85	0.000384
1116	4	XP_806477.1 hypothetical protein	2	9	4.84	5.13E-06
1117	4	EAN98243.1 mucin-associated surface protein (MASP)	1	1	4.80	2.33E-13
1118	2	XP_819496.1 hypothetical protein	1	7	4.78	2.08E-07
1119	2	EAN88971.1 ATP-dependent DEAD/H RNA helicase	1	1	4.78	1.80E-07
1120	2	XP_820835.1 hypothetical protein	1	12	4.77	3.04E-09
1121	2	XP_805476.1 protein phosphatase 2C-like	1	1	4.76	3.93E-06

1122	2	XP_813909.1 ABC transporter	1	11	4.74	0.000195
1123	4	EAN82832.1 RNA-binding protein	1	6	4.74	2.57E-09
1124	2	EAN93110.1 acyl carrier protein, mitochondrial precursor	1	2	4.74	7.33E-07
1125	4	EAN94755.1 hypothetical protein	1	2	4.73	5.61E-08
1126	4	EAN97508.1 hypothetical protein	1	1	4.71	7.13E-08
1127	4	EAN85256.1 proteasome regulatory non-ATPase subunit 8	1	6	4.71	3.22E-13
1128	2	EAN94987.1 ubiquitin carboxyl-terminal hydrolase	1	2	4.70	3.01E-09
1129	8	EAN98199.1 ribosomal protein L24	1	43	4.70	6.67E-09
1130	4	XP_811511.1 vacuolar ATP synthase	1	32	4.70	2.06E-10
1131	2	EAN84035.1 mucin TcMUCII	1	1	4.69	3.21E-11
1132	6	AAF05607.1 glutathione-S-transferase/glutaredoxin	1	2	4.69	2.19E-08
1133	4	XP_807722.1 hypothetical protein	1	3	4.67	7.30E-06
1134	4	XP_816577.1 hypothetical protein	1	2	4.66	5.35E-10
1135	2	XP_818127.1 signal peptide peptidase	1	8	4.65	1.80E-10
1136	2	XP_818955.1 hypothetical protein	2	5	4.65	6.67E-05
1137	2	EAN98585.1 hypothetical protein	1	6	4.61	2.94E-11
1138	4	EAN81193.1 heat shock protein 100	1	13	4.59	2.02E-07
1139	4	XP_807887.1 hypothetical protein	1	2	4.54	4.07E-08
1140	4	EAN89910.1 autophagocytosis associated protein	1	4	4.54	9.84E-05
1141	4	XP_815076.1 hypothetical protein	1	3	4.53	2.91E-11
1142	2	XP_817690.1 hypothetical protein	1	6	4.53	4.09E-11
1143	2	XP_814675.1 hypothetical protein	2	3	4.53	0.000403
1144	4	EAN91956.1 aldehyde dehydrogenase	1	2	4.52	3.06E-06
1145	2	XP_811189.1 coatomer alpha subunit	1	4	4.52	2.73E-10
1146	4	XP_817556.1 hypothetical protein	1	2	4.50	3.84E-11
1147	4	EAN88417.1 lactoylglutathione lyase-like protein	1	2	4.49	5.00E-07
1148	3	AAQ04681.1 80 kDa prolyl oligopeptidase	1	10	4.47	1.53E-09
1149	3	EAN84062.1 hypothetical protein	1	3	4.46	1.87E-11
1150	1	AAA30389.1 H ⁺ -ATPase B subunit	1	15	4.46	8.44E-09
1151	2	XP_820739.1 mucin-associated surface protein (MASP)	1	1	4.44	2.02E-08
1152	2	EAN82084.1 cytidine triphosphate synthase	1	6	4.44	1.79E-09
1153	4	XP_820934.1 syntaxin 5	1	2	4.44	3.68E-07
1154	2	XP_808107.1 small nuclear ribonucleoprotein protein	1	5	4.42	8.64E-10
1155	2	XP_813680.1 mucin-like glycoprotein	1	2	4.42	2.68E-08
1156	4	EAN92422.1 ribose-phosphate pyrophosphokinase	1	12	4.42	4.30E-08
1157	2	XP_807901.1 5'-3' exonuclease XNRA	1	1	4.41	1.53E-06

1158	4	XP_819110.1 hypothetical protein	1	2	4.39	5.16E-05
1159	4	XP_815994.1 cytochrome-B5 reductase	1	4	4.39	1.76E-05
1160	8	XP_805715.1 hypothetical protein	1	1	4.39	8.92E-06
1161	4	EAN81704.1 hypothetical protein	1	4	4.38	4.14E-10
1162	6	EAN92707.1 mucin-associated surface protein (MASP)	1	7	4.37	5.48E-07
1163	4	XP_813439.1 phosphoprotein phosphatase	1	1	4.37	9.66E-05
1164	5	EAN88501.1 hypothetical protein	1	10	4.35	4.25E-08
1165	4	XP_809176.1 hypothetical protein	1	2	4.33	2.29E-07
1166	2	EAN87798.1 hypothetical protein	1	4	4.33	0.000106
1167	2	EAN96412.1 mercaptopyruvate sulfurtransferase	1	4	4.31	1.36E-09
1168	2	EAN99756.1 hypothetical protein	1	3	4.31	7.57E-05
1169	4	EAN85094.1 hypothetical protein	1	9	4.31	2.19E-09
1170	4	XP_814903.1 hypothetical protein	1	3	4.30	3.42E-07
1171	2	EAN82927.1 retrotransposon hot spot (RHS) protein	1	4	4.29	1.04E-08
1172	8	EAN87948.1 hypothetical protein	1	13	4.29	3.44E-08
1173	2	XP_810207.1 hypothetical protein	1	1	4.29	0.000132
1174	4	XP_802301.1 mucin TcMUCII	1	1	4.28	1.03E-05
1175	2	XP_810134.1 hypothetical protein	1	2	4.26	2.29E-11
1176	2	EAN88793.1 hypothetical protein	1	4	4.25	3.05E-09
1177	2	XP_820860.1 peroxin-2	1	3	4.23	8.36E-07
1178	14	XP_813237.1 trans-sialidase	1	2	4.22	1.61E-06
1179	2	EAN97386.1 hypothetical protein	1	2	4.21	1.53E-06
1180	2	EAN90232.1 hypothetical protein	1	1	4.19	4.20E-07
1181	2	XP_821366.1 hypothetical protein	1	6	4.18	2.68E-07
1182	4	XP_821729.1 hypothetical protein	1	11	4.18	3.51E-05
1183	3	AAA67560.1 140/116 kDa antigen	1	4	4.17	2.44E-08
1184	4	XP_816006.1 hypothetical protein	1	4	4.16	5.85E-13
1185	4	EAN99250.1 diacylglycerol acyltransferase	1	5	4.16	3.18E-07
1186	6	EAN92849.1 ribosomal protein L36	1	8	4.15	9.79E-05
1187	2	EAN90998.1 NADH-cytochrome b5 reductase	1	1	4.14	7.64E-06
1188	4	XP_809102.1 hypothetical protein	1	3	4.13	8.65E-08
1189	10	EAN84856.1 retrotransposon hot spot (RHS) protein	1	5	4.13	7.18E-08
1190	2	EAN84728.1 hypothetical protein	1	1	4.13	5.55E-07
1191	2	XP_819607.1 cell cycle associated protein MOB1	1	1	4.12	1.11E-05
1192	2	EAN80780.1 dynein heavy chain	1	1	4.12	4.72E-06
1193	6	XP_819249.1 mucin-associated surface protein (MASP)	1	5	4.11	8.17E-08

1194	4	XP_809122.1 glycosomal membrane protein	1	26	4.08	1.15E-06
1195	4	XP_809147.1 hypothetical protein	1	9	4.07	9.98E-07
1196	2	XP_821480.1 developmentally regulated GTP-binding protein	1	4	4.07	7.26E-05
1197	4	EAN98302.1 proteasome beta-1 subunit	1	2	4.07	1.79E-07
1198	2	EAN97462.1 lysophospholipase	1	9	4.07	1.52E-06
1199	4	EAN96785.1 hypothetical protein	1	3	4.06	1.96E-06
1200	2	EAN95892.1 heat shock protein 70	1	5	4.06	2.71E-10
1201	4	XP_814357.1 hypothetical protein	1	3	4.04	1.89E-05
1202	4	XP_809482.1 hypothetical protein	1	4	4.04	2.10E-06
1203	3	EAN82516.1 endoribonuclease L-PSP (pb5)	1	1	4.04	1.91E-11
1205	4	EAN91779.1 ribulose-5-phosphate 3-epimerase	1	3	4.03	3.93E-08
1204	2	EAN88522.1 hypothetical protein	1	4	4.03	6.10E-09
1206	4	EAN89667.1 hypothetical protein	1	4	4.01	6.75E-05
1207	2	XP_815905.1 trans-sialidase	1	7	4.01	0.000166
1208	4	XP_813005.1 oxidoreductase	1	5	4.01	1.12E-09
1209	4	XP_808620.1 hypothetical protein	1	14	4.00	6.44E-05
1210	4	XP_814297.1 cytochrome c oxidase VIII (COX VIII)	1	3	3.99	7.03E-07
1211	9	XP_805519.1 40S ribosomal protein L14	1	32	3.99	6.59E-08
1212	4	EAN93257.1 hypothetical protein	1	5	3.98	0.000401
1213	2	XP_819388.1 mucin TcMUCII	1	10	3.98	5.95E-07
1214	2	EAN84308.1 pyrroline-5-carboxylate synthetase-like protein	1	6	3.97	2.17E-07
1215	4	EAN89862.1 ATP-dependent DEAD/H RNA helicase	1	3	3.96	3.80E-06
1216	2	EAN95113.1 hypothetical protein	1	1	3.94	2.63E-07
1217	2	XP_820949.1 hypothetical protein	1	5	3.94	3.19E-07
1218	4	EAN93241.1 hypothetical protein	1	4	3.93	3.11E-05
1219	2	EAN88355.1 hypothetical protein	1	3	3.93	8.02E-08
1220	4	XP_809393.1 mucin-associated surface protein (MASP)	1	2	3.91	4.13E-05
1221	4	EAN87022.1 protein kinase	1	9	3.90	3.81E-05
1222	4	XP_807137.1 hypothetical protein	1	6	3.89	3.07E-09
1223	4	EAN88213.1 hypothetical protein	1	12	3.88	5.28E-09
1225	4	XP_808549.1 hypothetical protein	1	3	3.87	0.000496
1224	2	EAN97613.1 protein kinase	1	1	3.87	1.37E-07
1226	27	XP_821289.1 surface protease GP63	1	2	3.86	1.16E-08
1227	4	XP_806627.1 outer arm dynein-like	1	2	3.86	1.56E-05
1228	4	XP_815294.1 dynein arm light chain	1	2	3.85	1.15E-07

1229	2	EAN83580.1 hypothetical protein	1	2	3.85	3.60E-05
1230	2	XP_819494.1 hypothetical protein	1	1	3.85	6.05E-06
1231	2	XP_817294.1 hypothetical protein	1	4	3.83	4.65E-05
1232	2	EAN81951.1 hypothetical protein	1	17	3.81	3.68E-06
1233	4	EAN95369.1 prefoldin subunit 2	1	5	3.78	6.93E-10
1234	4	XP_809750.1 proteasome beta 2 subunit	1	5	3.78	4.79E-11
1235	2	EAN97993.1 hypothetical protein	1	2	3.77	1.18E-07
1236	4	EAN98717.1 ATP synthase	1	1	3.77	0.000332
1237	2	XP_803020.1 hypothetical protein	1	4	3.77	0.00026
1238	4	EAN97939.1 protein translation factor SUI1 homolog	1	2	3.76	2.86E-05
1239	4	EAN81582.1 hypothetical protein	1	1	3.76	2.34E-05
1240	2	EAN94661.1 hypothetical protein	1	1	3.75	3.35E-07
1241	2	EAN89288.1 hypothetical protein	1	2	3.75	0.00015
1242	4	XP_809802.1 hypothetical protein	1	3	3.75	8.86E-06
1243	5	2A0M Arginase Superfamily Protein	1	1	3.74	8.56E-06
1244	4	EAN87082.1 hypothetical protein	1	2	3.74	1.92E-07
1245	2	XP_819694.1 hypothetical protein	1	2	3.74	0.000167
1246	2	XP_817406.1 protein kinase	1	3	3.74	0.000376
1247	2	XP_809654.1 myosin heavy chain	1	10	3.73	8.66E-05
1248	2	EAN94440.1 hypothetical protein	1	1	3.72	1.52E-05
1249	4	EAN82274.1 mucin-associated surface protein (MASP)	1	6	3.72	7.86E-06
1250	2	EAN98634.1 mucin-associated surface protein (MASP)	1	1	3.71	2.13E-05
1252	2	XP_818287.1 hypothetical protein	1	2	3.71	1.17E-05
1251	2	XP_802245.1 lathosterol oxidase	1	1	3.71	8.12E-06
1253	4	EAN89929.1 hypothetical protein	1	1	3.70	0.000204
1254	3	XP_813599.1 hypothetical protein	1	2	3.70	1.54E-08
1255	4	EAN94007.1 hypothetical protein	1	1	3.69	2.26E-06
1256	2	EAN91740.1 hypothetical protein	1	4	3.69	4.38E-07
1257	2	EAN98488.1 hypothetical protein	1	4	3.69	0.000206
1258	2	XP_806505.1 hypothetical protein	1	2	3.69	2.17E-06
1259	4	XP_804770.1 hypothetical protein	1	4	3.68	6.59E-06
1260	2	XP_810479.1 OSM3-like kinesin	1	1	3.68	2.23E-08
1261	4	XP_813700.1 small nuclear ribonucleoprotein Sm-F	1	4	3.67	2.26E-07
1262	4	XP_805811.1 GPI transamidase component Tta1	1	4	3.67	5.10E-06
1263	4	EAN85675.1 hypothetical protein	1	1	3.66	3.00E-05
1265	11	XP_809804.1 ribosomal protein L27	1	12	3.66	8.26E-06

1264	4	EAN92943.1 hypothetical protein	1	32	3.66	0.000102
1266	4	XP_817869.1 mucin-associated surface protein (MASP)	1	7	3.65	5.15E-07
1267	4	XP_814412.1 hypothetical protein	1	3	3.64	1.56E-05
1268	2	XP_820500.1 OSM3-like kinesin	1	1	3.64	1.88E-06
1270	2	XP_803893.1 phosphoribosylpyrophosphate synthetase	1	2	3.63	9.94E-06
1271	2	XP_809744.1 hypothetical protein	1	2	3.63	1.90E-06
1269	2	EAN96165.1 endoplasmic reticulum oxidoreductin	1	1	3.63	2.09E-07
1272	4	XP_821127.1 hypothetical protein	1	1	3.63	3.89E-05
1273	4	XP_805353.1 mucin-associated surface protein (MASP)	1	2	3.62	4.78E-05
1274	2	EAN88633.1 hypothetical protein	1	1	3.62	1.22E-06
1275	4	EAN84311.1 adenosine kinase	1	1	3.62	2.72E-06
1276	4	XP_811513.1 centromere/microtubule binding protein cbf5	1	1	3.62	5.43E-05
1277	4	EAN87603.1 protein kinase A catalytic subunit	1	2	3.61	0.000273
1278	2	EAN80706.1 dynein heavy chain	1	6	3.61	5.11E-08
1279	4	XP_803138.1 dihydroxyacetone kinase 1-like	1	2	3.61	0.000421
1280	4	EAN90723.1 adrenodoxin precursor	1	2	3.61	9.77E-09
1281	2	EAN87051.1 SPFH domain / Band 7 family protein	1	5	3.60	0.000429
1282	2	XP_818844.1 protein transport protein Sec24A	1	2	3.60	3.53E-06
1284	4	XP_805313.1 divalent cation tolerance protein	1	1	3.60	0.000116
1283	2	EAN85244.1 trypanothione peroxidase	1	30	3.60	2.71E-06
1285	4	XP_820192.1 hypothetical protein	1	3	3.60	1.26E-05
1286	2	EAN94035.1 hypothetical protein	1	1	3.59	1.27E-07
1287	2	EAN95881.1 hypothetical protein	1	5	3.58	8.89E-07
1288	8	EAN83919.1 protein kinase	1	64	3.56	0.000226
1289	4	EAN96033.1 hypothetical protein	1	1	3.55	6.25E-10
1290	2	XP_815543.1 hypothetical protein	1	1	3.55	0.000146
1291	4	EAN88001.1 hypothetical protein	1	6	3.54	0.000121
1292	2	EAN88986.1 mucin-like glycoprotein	1	2	3.54	8.23E-08
1293	4	EAN88079.1 hypothetical protein	1	7	3.53	4.29E-07
1294	6	XP_815416.1 mucin TcMUCII	1	1	3.53	0.000142
1295	4	XP_811878.1 hypothetical protein	1	1	3.53	3.98E-05
1296	2	XP_811555.1 hypothetical protein	1	1	3.53	8.38E-06
1297	4	EAO00283.1 hypothetical protein	1	39	3.52	1.72E-06
1298	2	EAN90038.1 hypothetical protein	1	1	3.52	0.000259
1300	2	XP_820913.1 hypothetical protein	1	2	3.52	0.000108
1299	2	XP_807687.1 hypothetical protein	1	1	3.52	0.00036

1301	2	EAN98788.1 mucin-associated surface protein (MASP)	1	1	3.51	0.000208
1302	2	XP_808900.1 vacuolar ATP synthase	1	9	3.51	0.000413
1305	5	XP_820609.1 hypothetical protein	1	5	3.50	2.74E-05
1304	4	EAN87062.1 kinetoplast DNA-associated protein	1	25	3.50	6.52E-07
1303	2	EAN82818.1 hypothetical protein	1	2	3.50	2.98E-09
1306	2	XP_810607.1 mucin TcMUCII	1	1	3.49	5.84E-05
1307	4	EAN98010.1 ubiquitin activating E1 enzyme	1	1	3.49	0.000179
1308	4	EAN92428.1 ubiquinone biosynthesis methyltransferase	1	1	3.49	7.01E-05
1310	2	XP_819185.1 biotin/lipoate protein ligase	1	3	3.49	0.000268
1309	2	XP_816708.1 hypothetical protein	1	2	3.49	1.61E-05
1311	4	EAN85465.1 hypothetical protein	1	5	3.49	7.44E-10
1312	2	EAN96125.1 DNA topoisomerase type IB small subunit	1	2	3.48	0.000238
1313	2	EAN96283.1 hypothetical protein	1	1	3.48	2.13E-05
1314	4	EAN82657.1 trichohyalin	1	3	3.48	1.51E-05
1315	2	XP_819746.1 hypothetical protein	1	1	3.48	0.00038
1316	4	XP_804842.1 peroxisome assembly protein	1	1	3.47	5.84E-05
1317	2	EAN96262.1 hypothetical protein	1	1	3.47	1.77E-07
1318	2	EAN95019.1 hypothetical protein	1	10	3.47	9.88E-07
1320	2	XP_810236.1 mucin-associated surface protein (MASP)	1	3	3.47	6.88E-05
1319	2	EAN84631.1 centrin	1	1	3.47	0.00023
1321	6	XP_802713.1 zinc finger protein 2	1	15	3.47	4.67E-08
1322	4	XP_817750.1 hypothetical protein	1	1	3.46	9.57E-05
1323	4	EAN91129.1 hypothetical protein	1	2	3.46	1.05E-07
1324	2	XP_810449.1 hypothetical protein	1	1	3.46	0.000105
1325	2	XP_806681.1 mucin-associated surface protein (MASP)	1	5	3.45	0.000198
1326	2	XP_818165.1 hypothetical protein	1	3	3.45	4.02E-06
1327	2	XP_809537.1 hypothetical protein	1	1	3.44	0.00011
1328	2	XP_810298.1 hypothetical protein	1	1	3.44	0.000308
1329	4	EAN81256.1 surface protease GP63	1	3	3.43	5.43E-08
1330	4	EAN96156.1 prohibitin	1	1	3.42	7.09E-05
1331	4	XP_811763.1 hypothetical protein	1	2	3.42	1.46E-05
1332	2	XP_804023.1 hypothetical protein	1	1	3.42	8.51E-09
1333	3	EAN94223.1 paraflagellar rod protein	1	2	3.41	3.87E-06
1334	2	EAN96741.1 serine carboxypeptidase S28	1	2	3.41	0.000133
1335	2	XP_817899.1 mucin TcMUCII	1	14	3.40	3.71E-07
1336	2	XP_812803.1 hypothetical protein	1	2	3.40	1.37E-06

1337	4	XP_818899.1 mu-adaptin 1	1	1	3.38	0.000198
1338	2	XP_818951.1 hypothetical protein	1	4	3.38	4.78E-07
1339	4	EAN94583.1 hypothetical protein	1	3	3.37	1.52E-07
1340	4	EAN94927.1 hypothetical protein	1	3	3.37	7.73E-08
1341	4	XP_803273.1 hypothetical protein	1	3	3.36	0.000194
1342	4	XP_805503.1 hypothetical protein	1	2	3.35	4.06E-05
1343	4	XP_806537.1 C-1-tetrahydrofolate synthase, cytoplasmic	1	1	3.35	0.000152
1344	4	EAN95355.1 leucine-rich repeat protein	1	3	3.35	1.68E-05
1345	2	XP_813540.1 hypothetical protein	1	2	3.35	0.000361
1346	4	XP_814231.1 ubiquitin-activating enzyme	1	1	3.34	0.000143
1347	2	XP_811506.1 cytochrome C oxidase subunit VI	1	4	3.34	2.91E-08
1348	2	XP_821567.1 hypothetical protein	1	2	3.34	0.000153
1349	2	EAN98051.1 mucin-associated surface protein (MASP)	1	37	3.34	5.98E-06
1350	2	XP_809167.1 mucin-associated surface protein (MASP)	1	4	3.34	1.87E-06
1351	4	EAN86163.1 protein kinase	1	2	3.33	1.25E-05
1352	2	EAN93185.1 hypothetical protein	1	3	3.32	4.09E-06
1353	2	EAN82374.1 dynein heavy chain	1	1	3.32	0.000236
1354	4	XP_816141.1 vesicle-associated membrane protein	1	2	3.32	2.97E-07
1355	4	XP_811957.1 cysteine proteinase	1	1	3.32	4.54E-05
1357	4	XP_819876.1 hypothetical protein	1	8	3.32	7.76E-07
1356	4	EAN85172.1 eukaryotic translation initiation factor 3 interacting protein	1	1	3.32	1.69E-07
1358	4	XP_809691.1 hypothetical protein	1	9	3.31	2.44E-07
1359	2	EAN92115.1 nucleolar RNA-binding protein	1	3	3.30	0.000487
1360	4	XP_809882.1 hydrogenase	1	3	3.30	0.00027
1361	2	XP_818676.1 hypothetical protein	1	2	3.30	0.000496
1362	4	EAN94942.1 mucin-associated surface protein (MASP)	1	2	3.30	0.00015
1363	2	XP_802257.1 mucin-associated surface protein (MASP)	1	2	3.29	1.32E-08
1364	26	EAN89197.1 kinesin	1	1	3.28	4.57E-07
1365	4	EAN97787.1 vacuolar ATP synthase subunit D	1	1	3.28	3.10E-06
1366	2	EAN82478.1 mucin-associated surface protein (MASP)	1	3	3.28	2.64E-05
1367	4	EAN87609.1 hypothetical protein	1	1	3.27	0.000256
1368	2	XP_810215.1 aquaporin-like protein	1	1	3.27	6.06E-05
1369	2	XP_810290.1 hypothetical protein	1	3	3.26	7.58E-09
1370	2	XP_809129.1 ADP/ATP translocase	1	1	3.26	6.06E-05
1371	4	XP_815383.1 hypothetical protein	1	12	3.26	3.22E-05

1372	2	XP_812272.1 hypothetical protein	1	1	3.24	1.15E-05
1373	4	XP_809140.1 hypothetical protein	1	5	3.23	0.000319
1374	2	EAN85710.1 hypothetical protein	1	3	3.23	2.24E-06
1375	4	EAN84944.1 hypothetical protein	1	1	3.23	0.000212
1376	2	XP_806703.1 hypothetical protein	1	3	3.21	1.29E-06
1377	2	EAN93745.1 hypothetical protein	1	2	3.20	0.000159
1378	2	XP_804136.1 hypothetical protein	1	3	3.20	0.000217
1379	5	AAQ10955.1 putative splicing factor XB2	1	1	3.19	1.69E-05
1380	2	XP_819752.1 hypothetical protein	1	1	3.19	4.42E-07
1381	2	EAN83160.1 phosphomannose isomerase	1	3	3.19	2.95E-06
1382	4	XP_809059.1 hypothetical protein	1	1	3.19	0.000335
1383	2	XP_817916.1 Golgi/lysosome glycoprotein	1	3	3.19	1.25E-05
1384	2	XP_821399.1 mucin-associated surface protein (MASP)	1	2	3.19	2.76E-06
1385	2	XP_812625.1 hypothetical protein	1	3	3.19	2.07E-05
1386	2	EAN96066.1 pitrilysin-like metalloprotease	1	2	3.18	0.000267
1387	4	EAN95232.1 acetyltransferase	1	2	3.18	5.31E-09
1388	3	AAK48428.1 putative CAAX prenyl protease 1	1	1	3.18	3.94E-07
1389	8	XP_821251.1 40S ribosomal protein S21	1	5	3.18	7.82E-05
1390	4	EAN90498.1 alpha/beta-hydrolase	1	1	3.18	2.54E-05
1391	4	XP_815796.1 hypothetical protein	1	2	3.17	2.15E-05
1392	2	XP_820022.1 mucin-associated surface protein (MASP)	1	1	3.17	0.000437
1393	2	EAN91340.1 hypothetical protein	1	2	3.17	0.000151
1394	4	XP_812323.1 hypothetical protein	1	3	3.17	0.000494
1395	2	EAN97875.1 mucin-associated surface protein (MASP)	1	8	3.16	0.00011
1396	10	EAN96703.1 mucin-like glycoprotein	1	1	3.16	0.000244
1397	4	XP_820241.1 ubiquitin fusion degradation protein 2	1	1	3.16	7.22E-07
1398	2	EAN86607.1 methionine aminopeptidase	1	1	3.15	4.32E-08
1399	4	XP_807452.1 mucin-associated surface protein (MASP)	1	5	3.15	4.91E-05
1400	4	EAN95902.1 katanin	1	6	3.15	9.53E-07
1401	4	XP_804930.1 glycine cleavage system H protein	1	43	3.14	0.000103
1402	4	XP_807664.1 dynein light chain	1	6	3.14	7.68E-09
1403	4	EAN94710.1 hypothetical protein	1	2	3.14	0.00047
1404	2	EAN91326.1 hypothetical protein	1	1	3.13	1.34E-05
1405	4	EAN88481.1 hypothetical protein	1	2	3.12	4.37E-05
1406	2	XP_811108.1 serine/threonine protein kinase	1	11	3.12	0.000474
1407	3	EAN99335.1 C-8 sterol isomerase	1	9	3.12	7.22E-07

1408	2	EAN93790.1 hypothetical protein	1	4	3.12	0.000316
1409	4	XP_807341.1 hypothetical protein	1	6	3.11	2.68E-08
1410	2	EAN84633.1 hypothetical protein	1	1	3.11	1.08E-05
1411	4	EAN89432.1 phosphatidyl serine synthase	1	1	3.11	2.42E-05
		EAN98512.1 DNA-damage inducible protein DDI1-like				
1412	4	protein	1	2	3.11	8.62E-08
1413	4	XP_804643.1 serine/threonine protein phosphatase 2A	1	1	3.11	6.54E-07
1414	2	EAN97672.1 hypothetical protein	1	5	3.10	9.74E-06
1415	5	1XQ7 Cyclophilin	1	4	3.10	0.000213
1416	8	XP_808829.1 dispersed gene family protein 1 (DGF-1)	1	1	3.10	0.000104
1417	2	EAN81584.1 hypothetical protein	1	3	3.10	4.57E-06
1418	2	EAN84402.1 mucin TcMUCII	1	3	3.09	3.42E-05
1419	2	EAN84259.1 hypothetical protein	1	1	3.09	0.000229
1420	2	XP_814683.1 hypothetical protein	1	1	3.09	1.39E-05
1422	4	XP_819616.1 acyl-CoA dehydrogenase	1	2	3.09	3.37E-06
1421	2	EAN97096.1 small nuclear ribonucleoprotein SmD3	1	1	3.09	9.67E-05
1423	2	EAO00041.1 hypothetical protein	1	2	3.09	6.88E-05
1424	4	XP_809465.1 hypothetical protein	1	2	3.08	3.02E-05
1425	2	XP_816394.1 hypothetical protein	1	3	3.08	8.86E-06
1426	4	EAN89339.1 hypothetical protein	1	3	3.07	0.000393
1427	2	XP_817213.1 hypothetical protein	1	17	3.07	0.000269
1428	4	EAN93467.1 trans-sialidase	1	1	3.06	0.00013
1429	2	EAN90735.1 protein kinase	1	4	3.06	0.000182
1430	2	XP_816032.1 hypothetical protein	1	1	3.06	3.59E-06
1431	10	XP_817863.1 mucin-associated surface protein (MASP)	1	4	3.05	7.45E-06
1433	4	XP_813045.1 ADP-ribosylation factor-like protein	1	2	3.05	0.000389
1432	2	EAN84083.1 hypothetical protein	1	1	3.05	0.000284
1434	8	XP_802657.1 cytochrome c oxidase subunit 10	1	5	3.05	0.000167
1435	4	XP_817654.1 quinone oxidoreductase	1	1	3.04	3.03E-06
1436	4	XP_805889.1 hypothetical protein	1	4	3.03	0.000236
1437	2	XP_814243.1 hypothetical protein	1	9	3.03	3.56E-07
1438	2	EAN84505.1 hypothetical protein	1	5	3.02	0.000326
1439	2	XP_805102.1 mucin-associated surface protein (MASP)	1	1	3.02	9.22E-05
1440	6	EAN85696.1 60S ribosomal protein L37	1	17	3.02	3.37E-05
1441	4	EAN86789.1 mucin TcMUCII	1	1	3.02	1.02E-07
1442	2	EAN99079.1 hypothetical protein	1	1	3.02	0.000114

1443	2	EAN94262.1 Zn-finger protein	1	31	3.01	0.000275
1444	2	EAN99986.1 tubulin-specific chaperone	1	10	3.01	3.52E-05
1445	2	XP_803984.1 trans-sialidase	1	2	3.01	2.20E-05
1446	4	EAN94113.1 hypothetical protein	1	1	3.01	0.000215
1447	4	EAN82788.1 hypothetical protein	1	1	3.01	6.67E-05
1448	4	XP_806827.1 hypothetical protein	1	1	3.01	1.73E-05
1449	5	EAN92503.1 hypothetical protein	1	1	3.00	0.000326
1450	4	EAN91032.1 hypothetical protein	1	1	3.00	6.52E-06

Table 4.4 Gene ontology analysis of identified *T. cruzi*-specific protein groups

N	Level	GO	Name	Count	% from total
1	1	GO:0003674	Molecular function	1032	71.3
2	2	GO:0003774	motor activity	34	2.3
3	3	GO:0003777	microtubule motor activity	34	2.3
4	2	GO:0003824	catalytic activity	635	43.9
5	3	GO:0004386	helicase activity	10	0.7
6	3	GO:0008641	small protein activating enzyme activity	2	0.1
7	3	GO:0008907	integrase activity	2	0.1
8	3	GO:0009975	cyclase activity	3	0.2
9	3	GO:0016491	oxidoreductase activity	91	6.3
10	3	GO:0016740	transferase activity	149	10.3
11	3	GO:0016787	hydrolase activity	324	22.4
12	3	GO:0016829	lyase activity	27	1.9
13	3	GO:0016853	isomerase activity	37	2.6
14	3	GO:0016874	ligase activity	39	2.7
15	2	GO:0004871	signal transducer activity	25	1.7
16	3	GO:0004872	receptor activity	4	0.3
17	3	GO:0005057	receptor signaling protein activity	13	0.9
18	3	GO:0005102	receptor binding	5	0.3
19	2	GO:0005198	structural molecule activity	78	5.4
20	3	GO:0003735	structural constituent of ribosome	68	4.7
21	3	GO:0005199	structural constituent of cell wall	5	0.3
22	3	GO:0005200	structural constituent of cytoskeleton	2	0.1
23	3	GO:0008307	structural constituent of muscle	3	0.2
24	2	GO:0005215	transporter activity	109	7.5
25	3	GO:0005344	oxygen transporter activity	3	0.2
26	3	GO:0005386	carrier activity	85	5.9
27	3	GO:0005478	intracellular transporter activity	1	0.1
28	3	GO:0005489	electron transporter activity	34	2.3
29	3	GO:0008565	protein transporter activity	11	0.8
30	3	GO:0015075	ion transporter activity	54	3.7
31	3	GO:0015144	carbohydrate transporter activity	3	0.2
32	3	GO:0015267	channel or pore class transporter activity	1	0.1

33	3	GO:0043492	ATPase activity, coupled to movement of substances	47	3.2
34	2	GO:0005488	binding	546	37.7
35	3	GO:0000166	nucleotide binding	230	15.9
36	3	GO:0003676	nucleic acid binding	132	9.1
37	3	GO:0003682	chromatin binding	1	0.1
38	3	GO:0005496	steroid binding	1	0.1
39	3	GO:0005515	protein binding	117	8.1
40	3	GO:0008289	lipid binding	3	0.2
41	3	GO:0019842	vitamin binding	1	0.1
42	3	GO:0030246	carbohydrate binding	11	0.8
43	3	GO:0031406	carboxylic acid binding	2	0.1
44	3	GO:0042277	peptide binding	1	0.1
45	3	GO:0043167	ion binding	138	9.5
46	3	GO:0043176	amine binding	1	0.1
47	3	GO:0046906	tetrapyrrole binding	7	0.5
48	3	GO:0048037	cofactor binding	24	1.7
49	3	GO:0050824	water binding	5	0.3
50	3	GO:0051540	metal cluster binding	4	0.3
51	2	GO:0016209	antioxidant activity	7	0.5
52	3	GO:0004601	peroxidase activity	6	0.4
53	3	GO:0004791	thioredoxin-disulfide reductase activity	1	0.1
54	2	GO:0030234	enzyme regulator activity	5	0.3
55	3	GO:0008047	enzyme activator activity	2	0.1
56	3	GO:0019207	kinase regulator activity	2	0.1
57	3	GO:0019208	phosphatase regulator activity	1	0.1
58	3	GO:0030695	GTPase regulator activity	1	0.1
59	2	GO:0030528	transcription regulator activity	21	1.5
60	3	GO:0003700	transcription factor activity	38	2.6
61	3	GO:0016564	transcriptional repressor activity	4	0.3
62	2	GO:0045182	translation regulator activity	25	1.7
63	3	GO:0008135	translation factor activity, nucleic acid binding	33	2.3
64	2	GO:0045735	nutrient reservoir activity	1	0.1
65	1	GO:0005575	Cellular component	412	28.5
66	2	GO:0005576	extracellular region	3	0.2
67	3	GO:0005615	extracellular space	1	0.1
68	2	GO:0005623	cell	402	27.8

69	3	GO:0000267	cell fraction	1	0.1
70	3	GO:0005622	intracellular	349	24.1
71	3	GO:0016020	membrane	118	8.1
72	3	GO:0030312	external encapsulating structure	1	0.1
73	3	GO:0042597	periplasmic space	1	0.1
74	3	GO:0042995	cell projection	11	0.8
75	2	GO:0019012	virion	4	0.3
76	3	GO:0019028	viral capsid	4	0.3
77	2	GO:0031974	membrane-enclosed lumen	5	0.3
78	3	GO:0043233	organelle lumen	5	0.3
79	2	GO:0031975	envelope	16	1.1
80	3	GO:0030313	cell envelope	1	0.1
81	3	GO:0031967	organelle envelope	15	1.0
82	2	GO:0043226	organelle	242	16.7
83	3	GO:0031982	vesicle	7	0.5
84	3	GO:0043227	membrane-bound organelle	136	9.4
85	3	GO:0043228	non-membrane-bound organelle	120	8.3
86	3	GO:0043229	intracellular organelle	241	16.6
87	3	GO:0043233	organelle lumen	5	0.3
88	2	GO:0043234	protein complex	190	13.1
89	3	GO:0000015	phosphopyruvate hydratase complex	16	1.1
90	3	GO:0000151	ubiquitin ligase complex	3	0.2
91	3	GO:0000178	exosome (RNase complex)	1	0.1
92	3	GO:0000502	proteasome complex (sensu Eukaryota)	21	1.5
93	3	GO:0000786	nucleosome	13	0.9
94	3	GO:0005643	nuclear pore	29	2.0
95	3	GO:0005667	transcription factor complex	7	0.5
96	3	GO:0005787	signal peptidase complex	29	2.0
97	3	GO:0005838	proteasome regulatory particle (sensu Eukaryota)	5	0.3
98	3	GO:0005852	eukaryotic translation initiation factor 3 complex	2	0.1
99	3	GO:0005853	eukaryotic translation elongation factor 1 complex	3	0.2
100	3	GO:0005875	microtubule associated complex	32	2.2
101	3	GO:0005941	unlocalized protein complex	10	0.7
102	3	GO:0005945	6-phosphofructokinase complex	4	0.3
103	3	GO:0005960	glycine cleavage complex	3	0.2
104	3	GO:0005971	ribonucleoside-diphosphate reductase complex	1	0.1

105	3	GO:0008180	signalosome complex	27	1.9
106	3	GO:0016272	prefoldin complex	2	0.1
107	3	GO:0016282	eukaryotic 43S preinitiation complex	2	0.1
108	3	GO:0016459	myosin	11	0.8
109	3	GO:0016469	proton-transporting two-sector ATPase complex	53	3.7
110	3	GO:0016471	hydrogen-translocating V-type ATPase complex	15	1.0
111	3	GO:0030529	ribonucleoprotein complex	76	5.2
112	3	GO:0043235	receptor complex	1	0.1
113	3	GO:0045252	oxoglutarate dehydrogenase complex	6	0.4
114	3	GO:0045254	pyruvate dehydrogenase complex	3	0.2
115	3	GO:0045259	proton-transporting ATP synthase complex	3	0.2
116	3	GO:0045261	proton-transporting ATP synthase complex, catalytic core F(1)	3	0.2
117	3	GO:0045275	respiratory chain complex III	2	0.1
118	3	GO:0045285	ubiquinol-cytochrome-c reductase complex	3	0.2
119	1	GO:0008150	Biological Process	810	55.9
120	2	GO:0000003	reproduction	2	0.1
121	3	GO:0019953	sexual reproduction	2	0.1
122	2	GO:0007275	development	6	0.4
123	3	GO:0007389	pattern specification	1	0.1
124	3	GO:0009653	morphogenesis	2	0.1
125	3	GO:0009790	embryonic development	2	0.1
126	3	GO:0009791	post-embryonic development	1	0.1
127	3	GO:0009888	tissue development	1	0.1
128	3	GO:0030154	cell differentiation	4	0.3
129	3	GO:0040029	regulation of gene expression, epigenetic	1	0.1
130	3	GO:0048513	organ development	2	0.1
131	3	GO:0048731	system development	3	0.2
132	2	GO:0007582	physiological process	788	54.4
133	3	GO:0008152	metabolism	701	48.4
134	3	GO:0016265	death	4	0.3
135	3	GO:0040011	locomotion	7	0.5
136	3	GO:0042592	homeostasis	4	0.3
137	3	GO:0050791	regulation of physiological process	53	3.7
138	3	GO:0050874	organismal physiological process	4	0.3
139	3	GO:0050875	cellular physiological process	747	51.6
140	3	GO:0051179	localization	162	11.2

141	2	GO:0009987	cellular process	758	52.3
142	3	GO:0007154	cell communication	38	2.6
143	3	GO:0007155	cell adhesion	13	0.9
144	3	GO:0030154	cell differentiation	4	0.3
145	3	GO:0050794	regulation of cellular process	52	3.6
146	3	GO:0050875	cellular physiological process	747	51.6
147	2	GO:0040007	growth	1	0.1
148	3	GO:0040008	regulation of growth	1	0.1
149	2	GO:0050789	regulation of biological process	56	3.9
150	3	GO:0040008	regulation of growth	1	0.1
151	3	GO:0040029	regulation of gene expression, epigenetic	1	0.1
152	3	GO:0048518	positive regulation of biological process	1	0.1
153	3	GO:0048519	negative regulation of biological process	2	0.1
154	3	GO:0050790	regulation of enzyme activity	3	0.2
155	3	GO:0050791	regulation of physiological process	53	3.7
156	3	GO:0050794	regulation of cellular process	52	3.6
157	2	GO:0050896	response to stimulus	37	2.6
158	3	GO:0006950	response to stress	23	1.6
159	3	GO:0007600	sensory perception	1	0.1
160	3	GO:0007610	behavior	11	0.8
161	3	GO:0009605	response to external stimulus	11	0.8
162	3	GO:0009607	response to biotic stimulus	4	0.3
163	3	GO:0009628	response to abiotic stimulus	29	2.0
164	3	GO:0009719	response to endogenous stimulus	4	0.3
165	3	GO:0051606	detection of stimulus	1	0.1
166	2	GO:0051704	interaction between organisms	147	10.2
167	3	GO:0044419	interspecies interaction between organisms	147	10.2

Table 4.5 Analysis of surface protein and GPI-anchoring and MHC-presentation sequence prediction.

Protein/Family	Protein groups				Unique proteins		
	Surface proteins	GPI-anchored proteins	MHC Class I epitopes	MHC Class II epitopes	Surface proteins	GPI-anchored p	MHC Class I epitopes
<i>TS/gp85</i>	140	114	23	5621	789	620	133
trans-sialidase	130	111	22	5279	744	604	131
Tc85	3	0	0	97	7	1	0
85 kDa surface protein	2	1	0	60	10	1	0
c71	1	1	1	51	1	1	1
gp90	1	0	0	43	1	0	0
gp82	1	1	0	33	11	11	0
surface protein	0	0	0	0	2	0	0
surface protein-2	1	0	0	45	8	1	1
surface glycoprotein	1	0	0	13	3	0	0
amastigote surface protein-2	0	0	0	0	1	0	0
Shed-acute-phase-antigen	0	0	0	0	1	1	0
<i>MASP</i>	37	36	3	615	136	134	18
<i>TcMUCII</i>	12	10	0	136	46	42	8
<i>mucin-like</i>	3	1	0	94	12	8	0
<i>putative surface antigen YASP</i>	1	1	0	15	5	5	0
<i>ToIT</i>	4	3	0	70	16	14	0
<i>gp63</i>	11	9	3	377	83	51	8
<i>hypothetical proteins</i>		4	0	71		12	0
<i>unknown</i>		0	0	0		1	1
<i>amastigote cytoplasmic antigen</i>		0	0	0		1	0
<i>cytochrome b5-like</i>		0	0	0		2	0
Total	208	178	29	6999	1087	890	168

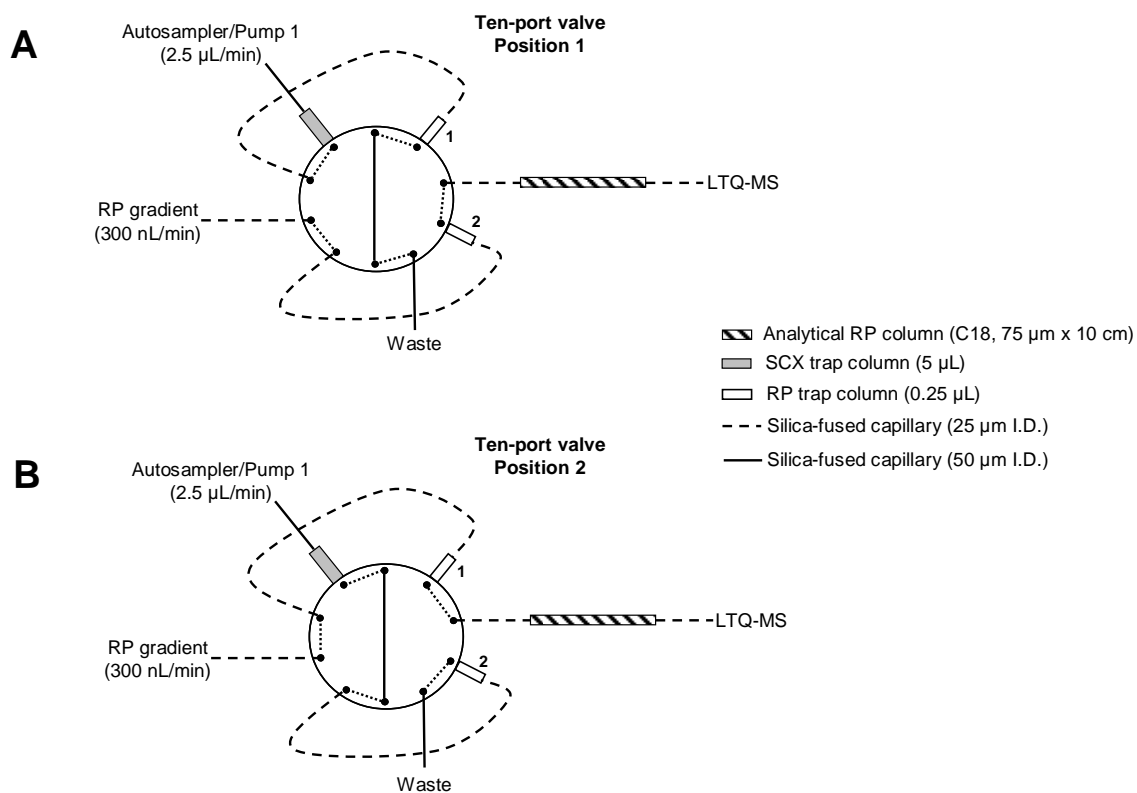


Figure 4.1 Two dimensional liquid chromatography tandem mass spectrometry system. **(A)** Ten-port valve in position 1. While the SCX gradient is running in the RP trap column 1, the RP gradient is running through the RP trap column 2. **(B)** Ten-port valve in position 2. While the SCX gradient is running in the RP trap column 2, the RP gradient is running through the RP trap column 3. See Table 4.1 for more details about the first dimension gradient.



Figure 4.2 Gene Ontology analysis. GO analysis was performed using GOblet algorithm. The most representative categories (≤ 10 sequences) for “molecular function” **(A)**, “cellular component” **(B)**, and “biological process” **(C)** are shown in the graph. For the complete list, see Table 4.4.

Chapter 5: Phosphoproteomic analysis of the human pathogen *Trypanosoma cruzi* at the epimastigote stage

Ernesto S. Nakayasu¹⁺, Matthew R. Gaynor^{1,2+}, Jeremy A. Ross¹ and Igor C. Almeida¹

¹Department of Biological Sciences, University of Texas at El Paso, El Paso, TX, USA,

²Division of Science and Mathematics, Alice Lloyd College, Pippa Passes, KY, USA.

Correspondence: Dr. Igor C. Almeida. Department of Biological Sciences, University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968 USA. Telephone: +1 (915) 747-6086. Fax: +1 (915) 747-5808. E-mail: icalmeida@utep.edu

⁺Both these authors contributed equally to this work.

Abbreviations: **FA**, formic acid; **GO**, gene ontology; **MSA**, multistage activation; **SCX**, strong-cation exchange;

Key words: Cell Signaling / Chagas disease / Phosphoproteome / Posttranslational Modification / *Trypanosoma cruzi*

5.1 Abstract

Trypanosoma cruzi, the etiologic agent of Chagas disease, affects millions of people in Latin America and is becoming a public health concern in the United States and Europe. The possibility that kinase inhibitors represent novel anti-parasitic agents is currently being explored. However, fundamental understanding of the cell signaling networks requires the detailed analysis of the involved phosphorylated proteins. Here we have performed a comprehensive mass spectrometry (MS)-based phosphorylation mapping study of phosphoproteins from *T. cruzi* epimastigote forms. Our liquid chromatography (LC)-tandem MS (MS/MS, MS/MS/MS, and multistage activation) analysis has identified 247 phosphopeptides from 144 distinct proteins. Furthermore, 220 phosphorylation sites were unambiguously mapped: 163 on serine, 49 on threonine, and 8 on tyrosine. In addition, immunoprecipitation and Western blotting analysis confirmed the presence of at least seven tyrosine-phosphorylated proteins in *T. cruzi*. The identified phosphoproteins were subjected to Gene Ontology, InterPro, and BLAST analysis, and categorized based on their role in cell structure, motility, transportation, metabolism, pathogenesis, DNA/RNA/protein turnover, and signaling. Taken together our phosphoproteomic data provide new insights into the molecular mechanisms governed by protein kinases and phosphatases in *T. cruzi* development and function. We discuss the potential roles of the identified phosphoproteins in parasite physiology and drug development.

5.2 Introduction

Chagas disease or American trypanosomiasis is one of the most prevalent tropical illnesses with an estimated 11 million people infected and approximately 120 million people living in high risk areas. Chagas disease is a major public health problem in Latin America, where up to 50,000 people may die every year due to complications in the acute or chronic phases of the disease (Barrett *et al*, 2003; Dias *et al*, 2002). In addition, due to the migration of chronically infected people from endemic areas and lack of screening in blood banks, Chagas disease is becoming a public health concern in United States and Europe (Bern *et al*, 2007; Piron *et al*, 2008).

Chagas disease is caused by the protozoan parasite *Trypanosoma cruzi*, which is naturally transmitted by hematophagous Reduviidae insects; however, the parasite may also be transmitted congenitally or via blood transfusion or organ transplantation (Barrett *et al*, 2003). The *T. cruzi* life cycle comprises two stages in the insect vector and two stages in the human host. In the insect vector, epimastigote forms replicate in the midgut, whereas in the more distal portion of the gut, they are transformed into infective metacyclic trypomastigotes under nutritional stress. These metacyclic forms are expelled together with the insect's feces during a bloodmeal, and invade the host through the bite wound or exposed mucosas. Inside the human host, the parasite can infect different types of nucleated cells, where they immediately differentiate into amastigote forms, which reproduce by binary fission. After several divisions, amastigotes transform into trypomastigote forms, which are eventually released into the extracellular milieu and then reach the bloodstream, where they can be taken up by the insect vector or infect new host cells, thus completing the natural cycle (Tyler and Engman, 2001).

The treatment of Chagas disease is currently restricted to two drugs, nifurtimox and benznidazole, which have limited efficacy and cause severe side effects (Urbina and Docampo, 2003; Wilkinson *et al*, 2008). In addition, there is no effective vaccine against *T. cruzi* (Dumonteil, 2007; Garg and Bhatia, 2005). Therefore, there is a critical need to develop new therapeutic strategies for treating Chagas disease.

In line with the current interest in protein kinases as molecular targets for the treatment of a variety of diseases (Arslan *et al*, 2006; Cohen and Goedert, 2004; Force *et al*, 2004; Kumar *et al*, 2003), the possibility that kinase inhibitors represent novel anti-parasitic agents is currently being explored (Doerig, 2004). Reversible protein phosphorylation is a key mechanism for the regulation of major biological processes including proliferation and differentiation. Approximately 2% of the *T. cruzi* genome encodes protein kinases, suggesting a major regulatory role in controlling parasite development and function (Parsons *et al*, 2005). A comparative study of the kinomes of trypanosomatids showed that *T. brucei*, *T. cruzi*, and *Leishmania major* have 176, 190, and 199 protein kinase genes, respectively (Naula *et al*, 2005; Parsons *et al*, 2005). Of these kinases, approximately 12% are unique to trypanosomatids (Naula *et al*, 2005; Parsons *et al*, 2005). Among the protein phosphatases, *T. brucei*, *T. cruzi* and *L. major* have 78, 86 and 88 genes, respectively. About 40% of these phosphatase genes were atypical with no clear orthologs in other eukaryote genomes (Brenchley *et al*, 2007). Taken together, the significant differences between *T. cruzi* and host cell protein kinases suggest that parasite specific inhibition can be achieved and, therefore, may represent a viable therapeutic approach to control Chagas disease.

Despite the importance of protein phosphorylation in many cellular processes, few studies have identified phosphorylation sites in trypanosomatid proteins. Recently, proteomic analysis of promastigote and amastigote forms of *Leishmania donovani* identified 73

phosphoproteins with diverse biological functions, however the specific phosphorylation sites (or phosphosites) were not identified (Morales *et al*, 2008). Another proteomic study in *L. donovani* was able to identify 18 phosphosites from 16 distinct phosphopeptides (Rosenzweig *et al*, 2008); however, to our knowledge the only phosphorylation site determined in a *T. cruzi* protein is serine phosphorylation of the linker histone H1 (da Cunha *et al*, 2005). In order to gain insight into the signaling networks that govern parasite function, we analyzed the phosphoproteome of *T. cruzi* at the epimastigote stage. Total epimastigote lysate was digested with trypsin, followed by the enrichment of phosphopeptides by strong cation exchange (SCX) and immobilized metal affinity chromatography (IMAC), and analysis by LC-MS. Here, we show a comprehensive phosphorylation mapping study and discuss the potential roles of the identified phosphoproteins in parasite physiology and drug development.

5.3 Materials and Methods

5.3.1 Cell culture and protein extraction

T. cruzi epimastigotes (Y stain) were grown in liver infusion tryptose medium containing 10% fetal bovine serum at 28°C for 3-4 days (Camargo, 1964). The parasites were harvested and washed 3 times with phosphate-buffered saline (PBS), pH 7.4. Epimastigotes (2.4×10^{10}) were lysed with 3 ml 8 M urea, 400 mM NH_4HCO_3 , containing phosphatase inhibitor cocktail (Sigma-Aldrich), by vigorous vortexing for approximately 5 min. Disulfide bonds were reduced with 5 mM dithiothreitol for 15 min at 50°C, and the free thiol groups were alkylated with 10 mM iodoacetamide for 30 min at room temperature and protected from the light. Protein content was measured by the Micro BCA assay (Pierce), according to the manufacturer's protocol.

5.3.2 Trypsin Digestion

After BCA quantification, an equivalent of 10 mg total protein extract was diluted to 10 ml with 100 mM NH_4HCO_3 and digested with 100 µg sequencing-grade trypsin (Promega) for 24 h at 37°C. After the digestion, peptides were acidified by adding 100 µl formic acid (FA) and desalted in a C18 cartridge (DSC-18, Supelco, Sigma-Aldrich). The cartridge was activated with 4 ml methanol and equilibrated with 4 ml 0.05% TFA. After loading and washing with 4 ml 0.05% TFA, the sample was eluted with 2 ml 80% ACN / 0.05% TFA and dried in a vacuum centrifuge (Vacufuge, Eppendorf).

5.3.3 Strong Cation-Exchange (SCX) Chromatography

SCX fractionation was performed with 100 µl POROS HS 50 resin (Applied Biosystems) placed in a SPE support cartridge. After equilibrating the column with 25% ACN / 0.5% FA

(SCX buffer), the samples were loaded and eluted with 1 ml 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 250, and 500 mM NaCl dissolved in SCX buffer. The fractions were dried in a vacuum centrifuge, dissolved in 200 μ l 0.05% TFA, desalted in POROS R2 50 ziptips as described elsewhere (Jurado *et al*, 2007), and dried again.

5.3.4 Immobilized Metal-Affinity Chromatography (IMAC)

Fifty microliters of IMAC resin (PHOS Select Iron Affinity Gel, Sigma-Aldrich) was washed 2X with 450 μ l 0.25 M acetic acid / 30% ACN (IMAC buffer). Each fraction from SCX chromatography was redissolved in 100 μ l IMAC buffer and incubated with the resin for 60 min with constant shaking. After the incubation, the resin was loaded onto ziptips, washed 5X with 100 μ l IMAC buffer and eluted 2X with 100 μ l 5% TFA / 45% ACN. Samples were dried in a vacuum centrifuge prior to LC-MS analysis.

5.3.5 Liquid Chromatography-Mass Spectrometry (LC-MS)

Following IMAC purification each fraction was redissolved in 30 μ l 0.05% TFA and 8 μ l were used for each injection. The peptides were loaded in a C18 trap column (0.25 μ l, Opti-Pak, Optimized Technologies) coupled to a nanoHPLC system (1D Plus, Eksigent). The separation was carried in a capillary RP column (Acclaim, 3 μ m C18, 75 μ m x 25 cm, LC Packings, Dionex) in a gradient of 2-33.2% ACN / 0.1% FA for 120 min. The eluted peptides were analyzed online in a linear ion trap mass spectrometer (LTQ XL/ETD, Thermo Fisher Scientific). The five most abundant ions were submitted to data-dependent CID MS/MS, multistage activation (MSA) or MS/MS/MS fragmentation before being dynamically excluded for 2 min. The CID was set to 40% normalized collision energy. The MSA was set at -98.0, -49.0, and -32.6 Thompson (Th) relative to the precursor ion, and MS/MS/MS scans were

triggered by the appearance of phosphate neutral loss (-98.0, -49.0, and -32.6 Th) in the MS² scan event.

5.3.6 Bioinformatics analysis

MS/MS spectra were converted to DTA files using Bioworks (v3.3.1, Thermo Fisher Scientific) with the following criteria: peptide masses from 800 to 3500 Da, at least 15 ions, and a minimum of 10 counts. TurboSequest searches were done against the forward and reverse *T. cruzi*, human keratin, and porcine trypsin sequences (total of 84948 sequences, downloaded on March 17th, 2008 from GenBank). The parameters for database search were: fully trypsin digestion; up to 1 missed cleavage site; 2 Da for peptide mass tolerance; 1 Da for fragment mass tolerance; cysteine carbamidomethylation (+57 Da) as fixed modification; and methionine oxidation (+16 Da) and serine, tyrosine and threonine phosphorylation (+80 Da for MS/MS and -18 Da for MS/MS/MS) as variable modifications. The datasets were filtered with DCn \geq 0.05, peptide probability \leq 0.1 (\leq 0.05 for MS/MS/MS data), and Xcorr of 1.5, 2.2, and 2.7 for singly-, doubly- and triply-charged peptides, respectively. All phosphopeptide spectra were carefully examined for diagnostic b and y fragments to determine the exact modification site. For phosphorylation sites that could not be determined, the possible modified amino acid residue was indicated by lower cap and the number of phosphate groups in parenthesis. The phosphorylation motifs were searched using Phosida phosphorylation site database (<http://www.phosida.com/>) (Gnad *et al*, 2007). All valid phosphoproteins were submitted to Gene Ontology (GO) analysis using GOblet tool (<http://goblet.molgen.mpg.de/cgi-bin/goblet/webapp-goblet.cgi>) (Hennig *et al*, 2003). GO annotation was performed on October 2nd, 2008, against the invertebrate taxonomy and an e-value \leq 1e-10. The protein sequences

were also submitted to an automated Blast ($e\text{-value} \leq 1e\text{-}5$) and InterPro annotation using Blast2go algorithm (October 3rd, 2008, <http://www.blast2go.de/>) (Gotz *et al*, 2008).

5.3.7 Immunoprecipitation and Western blot analysis

An equivalent of 1.2×10^9 epimastigote cells were solubilized in Triton lysis buffer (10 mM Tris-HCl (pH 7.6), 5 mM EDTA (pH 8.0), 50 mM NaCl, 30 mM $\text{Na}_4\text{P}_2\text{O}_7$, 50 mM NaF, 1 mM Na_3VO_4 , 1% Triton X-100) containing 1 mM phenylmethylsulfonyl fluoride, 5 $\mu\text{g/ml}$ aprotinin, 2 $\mu\text{g/ml}$ leupeptin, 1 $\mu\text{g/ml}$ pepstatin A, and clarified by centrifugation ($16,000 \times g$, 10 min, 4°C). The supernatants were rotated with 10 μg of anti-phosphotyrosine monoclonal antibody (4G10, Upstate) for 2 h at 4 °C. The immune complex was captured by incubation for 1 h at 4°C with protein A-Sepharose beads (Rockland Immunochemicals). The beads were then washed 3 times with cold lysis buffer and eluted by boiling in 2x SDS sample buffer (50 mM Tris-HCl (pH 6.8), 100 mM dithiothreitol, 2% SDS, 0.02% bromophenol blue, 10% glycerol, pH 6.8). Samples were resolved by 10% SDS-PAGE and transferred to PVDF membrane (Amersham Biosciences, GE Healthcare Life Sciences). After blocking with 1% BSA, the membrane was incubated overnight with anti-phosphotyrosine monoclonal antibody (4G10, Upstate), followed by 2 h incubation with horseradish peroxidase-conjugated goat anti-mouse IgG (H+L) (KPL), and visualized by using enhanced chemiluminescence reagent (ECL) and X-ray film (Phenix Research Products).

5.4 Results and Discussion

5.4.1 Phosphopeptide enrichment and phosphorylation site mapping

The identification of phosphoproteins can be limited due to their low abundance and the low stoichiometry of phosphorylation. In order to map the phosphorylation sites in *T. cruzi* proteins, whole epimastigote extracts were digested with trypsin, fractionated by SCX chromatography, and the resulting phosphopeptides enriched using IMAC. The enriched fractions were analyzed by LC-MS/MS, phosphate neutral loss-triggered LC-MS/MS/MS, or LC-MSA. It is well documented that MS/MS analysis of phosphopeptides leads to a strong neutral loss of the phosphate group (DeGnore and Qin, 1998). In many cases, the resulting spectra do not provide sufficient fragments to determine the phosphopeptide sequences. To circumvent this problem, neutral loss-triggered MS/MS/MS (Beausoleil *et al*, 2004) and MSA (also known as pseudo-MS³) (Schroeder *et al*, 2004) were developed. To increase the phosphoproteome coverage, we applied all three dissociation strategies that resulted in the identification of similar numbers of non-phosphorylated peptides (Table 5.1). However, MSA identified more phosphopeptides (160 peptides) compared to MS/MS (41 peptides) or MS/MS/MS (112 peptides). In our analyses, 247 phosphopeptides were sequenced from 226 phosphorylation sites in 144 proteins (Table 5.1). Of those 226 newly identified phosphorylation sites, 163 (72.1%) were on serine, 49 (21.7%) on threonine, and 8 (3.5%) on tyrosine (Table 5.1). Six phosphorylation sites could not be assigned to specific amino acid residues, 5 on serine or threonine and 1 on serine or tyrosine (Table 5.1). From the identified phosphoproteins, approximately 33% had multiple phosphorylation sites (Table 5.1). The identification of these phosphoproteins further supports the notion that kinase-mediated signal transduction pathways are important in the regulation of parasite biological processes. Unlike

metazoa and yeast, which utilize regulated transcription factors to direct the expression of certain genes, trypanosomatids indiscriminately transcribe most genes in large polycistronic units, thus emphasizing the critical role of posttranslational modifications (PTM) in the regulation of *T. cruzi* proteins.

5.4.2 Annotation of phosphorylation motifs

The specificity of protein kinases is determined by the amino acid residues adjacent to the phosphosite and is termed the consensus phosphorylation motif. To determine the kinase(s) responsible for modification of the identified phosphosites, the phosphorylation motifs were annotated using the Phosida search algorithm [15]. The most abundant motifs are those from CAMK2 (14.3%), CK1 (12.0%), PKA (9.0%), CK2 (7.9%), GSK3 (7.9%), ERK (6.7%), and CDK1 (6.1%) (Table 5.2). Fifty-six (16.3%) phosphorylation sites did not match any known motif (Table 2), which may be due to kinases with undetermined specificity or atypical kinases that might only be present in trypanosomatids (Naula *et al*, 2005; Parsons *et al*, 2005). Interestingly, there was a significant amount of redundancy among the identified motifs, therefore more motifs (343) are present compared to phosphorylation sites (226). In these cases, it is probable that specificity is achieved by different spatial and temporal expression of the kinase(s). Intriguingly, a number of tyrosine-kinase phosphorylation motifs (ALK, SRC, and EGFR) were identified (Table 5.2). A key difference between host and parasite kinomes is the absence of receptor-linked and cytoplasmic tyrosine kinases in trypanosomatids (Parsons *et al*, 2005); however, the presence of protein-tyrosine phosphatases indicates that tyrosine phosphorylation is a key regulatory mechanism in *T. cruzi* (Brenchley *et al*, 2007). It has been proposed that atypical tyrosine kinases such as Wee1 and dual-specificity kinases such as the

DYRKs and CLKs, which are all present in the *T. cruzi* genome, are responsible for this activity (Parsons *et al*, 2005).

5.4.3 Functional categorization of identified phosphoproteins

Of 144 sequenced phosphoproteins, 85 (59%) were annotated in the GenBank database as hypothetical proteins, making it difficult to infer their biological function (Table 5.3). Therefore, we submitted all identified phosphoproteins to automated Blast, InterPro and Gene Ontology (GO) analysis (see Material and Methods for details). Since the majority of *T. cruzi* sequences were compiled in 2005 with the completion of the genome project (El-Sayed *et al*, 2005), new entries in the GenBank database could help in the annotation of the identified sequences. Indeed, Blast analysis resulted in the annotation of 33 sequences previously described as hypothetical proteins (Table 5.3). In addition, InterPro analysis resulted in 84 entries, 30 of which were previously annotated as hypothetical proteins (Table 5.3). Finally, GO analysis generated 90 annotations, 38 of which were previously annotated as hypothetical proteins (Table 5.3, Table 5.4). Taken together, Blast, InterPro and GO analyses, assigned 49 (58%) of the 85 hypothetical sequences to a predicted biological function (Table 5.3). The overall function of the identified phosphoproteins is discussed in more detail in the following subsections.

5.4.3.1 Gene ontology analysis

The overall functional distribution of the identified phosphoproteins, as determined by GO analysis (Table 5.4), is shown in Figure 5.1. The majority of proposed functions involved molecular interactions with proteins, nucleic acid, nucleotide, ions, and lipids. Enzymatic or catalytic activity and transport functions were also prevalent (Fig. 5.1, Table 5.4). The most

abundant cellular components were intracellular, membrane bound, and localized in organelles (Fig. 5.1, Table 5.4). For the biological process category, a large proportion of the phosphoproteins were related to metabolism and other physiological processes. There were also proteins related to cell reproduction, development, differentiation and death; cell communication, response to stimulus, and locomotion (Fig. 5.1, Table 5.4). Overall, GO analysis further illustrates that protein phosphorylation is a significant regulatory mechanism governing a variety of *T. cruzi* biological functions.

5.4.3.2 Signaling transduction-related proteins

Signal transduction pathways are highly dynamic protein networks that integrate information from various stimuli. Twenty-one of the identified proteins are proposed to be cell signaling modulators. Of these, eight (38%) are classified as either protein kinases or phosphatases by Blast, InterPro, or GO analysis (Table 5.3), suggesting a number of signaling pathways are active. Indeed, calcium-binding proteins such as calmodulin and EF-hand domain-containing phosphoproteins were identified. Consistent with these observations, 14% of the phosphorylation motifs were characterized as CAMK2 consensus sites, indicating Ca^{2+} plays an important role in the regulation of many processes in *T. cruzi* physiology. Interestingly, a number of WD40 domain-containing phosphoproteins were identified. WD40 domain is an amino acid repeat with conserved tryptophan and aspartic acid residues (Cole, 2003). This repeat has been shown to bind to phosphoserine/threonine and to act as a adaptor motif to anchor ubiquitin ligase, thus playing an important role in phosphorylation-dependent protein degradation pathways (Yaffe and Elia, 2001). WD40 repeats are also present in several proteins from *Chlamydomonas reinhardtii* intraflagellar transport system (Cole, 2003), indicating that these proteins might also be involved in phosphorylation/ubiquitination-regulated

cellular trafficking. Taken together, we have established a functional relationship between the proposed signaling molecules that provides unique insights into parasite biology with relevance to future drug development.

5.4.3.3 Cytoskeleton, flagellum and trafficking proteins

Fifteen of the identified phosphoproteins were annotated as cytoskeleton, flagellum, and trafficking proteins (Table 5.3). In mammalian sperm, hyperactivated motility is mediated by Ca^{2+} (Hill, 2003); however, it remains unclear whether the same phenomenon occurs in trypanosomatid flagellum. The presence of Ca^{2+} sensors in *Trypanosoma sp.* flagellum, such as calmodulin and flagellar calcium-binding protein (FCaBP), suggest that this phenomenon does occur in these protozoa (Broadhead *et al*, 2006; Hill, 2003; Maldonado *et al*, 1999). We found that one paraflagellar rod protein (PAR1b) and three proteins with orthologs in mammalian sperm are phosphorylated in a CAMK2 consensus motif, supporting the role of Ca^{2+} -mediated protein phosphorylation in the regulation of motility in trypanosomes.

5.4.3.4 Transporters

We found 8 transporters in our analysis, 3 ABC, 4 ion and 1 acetate transporter. The ABC transporters are widely correlated with multidrug resistance from bacteria to humans (Jones and George, 2005). The *T. cruzi* genome has 33 ABC transporter genes (Aguero *et al*, 2006). Two ABC-related transporters (tcpgp1 and tcpgp2) were characterized in *T. cruzi*, but their expression was not correlated with the efflux of nifurtimox or benznidazole (Jones *et al*, 2005). Lara *et al.* proposed that the heme uptake by epimastigotes was dependent on ABC transporters (Lara *et al*, 2007). More research is necessary to elucidate the ABC transporter

functions in drug resistance, nutrient uptake, and metabolic secretion, and the role of phosphorylation in these processes.

5.4.3.5 DNA, RNA, and protein turnover

We have found three proteins with function corresponding to nucleosome and chromatin assembly (i.e., histone H2B, subunit of chromatin-modifying protein, nucleosome assembly protein). Marques Porto et al. showed by ^{32}P -labeling the phosphorylation of *T. cruzi* histone H1 and H2B (Marques Porto et al, 2002). Although we have previously mapped the phosphorylation of histone H1 and now report herein the phosphorylation of histone H2B, the phospho-histone H1 was not detected in the current analysis. This could be because the phosphopeptide derived from histone H1 has an acetylated N-terminus and several miscleaved sites by trypsin digestion (da Cunha et al, 2005), thus they were not considered in current our analysis. Identification of heterogeneous nuclear ribonucleoprotein H/F, ATP-dependent RNA helicase, and pab1 binding protein (poly-A binding) suggests phosphorylation-mediated signal transduction pathways play a key role in *T. cruzi* RNA synthesis, processing, and degradation. Additionally, identification of protein synthesis modulators such as translation elongation factor and tRNA synthase as well as protein folding and processing modulators such as heat shock protein 70 and calpain indicates that phosphorylation plays an important role in these processes in *T. cruzi*.

5.4.3.6 Metabolism

Among the metabolic proteins, we identified numerous enzymes involved in the late stages of the glycolytic pathway (pyruvate phosphate dikinase, pyruvate dehydrogenase E1 alpha subunit, and NADH-dependent fumarate reductase) (Table 5.3). Two phosphoproteins

involved in nucleotide synthesis (ribose-phosphate pyrophosphokinase and phosphoribosylpyrophosphate synthetase) and lipid metabolism (phosphatidyl synthase, AT-hook motif family protein) were also identified. These results suggest that these metabolic pathways are controlled at least in part by phosphorylation and are highly active, thus feeding the machinery responsible for parasite cell growth and proliferation.

5.4.3.7 Pathogenesis

Five phosphoproteins belonging to 3 protein groups/families (*trans*-sialidase (TS), type-III secretion chaperone, and dispersed gene family) were identified and related to pathogenesis (Table 5.3). TS is a well known virulence factor (Risso *et al*, 2004) and comprises a protein superfamily (TS/gp85) encoded by about 1400 genes in *T. cruzi* genome (El-Sayed *et al*, 2005). Members of the TS/gp85 superfamily have the enzymatic activity to transfer sialic acid residues from the host to parasite glycoconjugates, which is related to protection against the host immune response (Pereira-Chioccola *et al*, 2000). Other members have been related to host cell recognition and invasion to protect the parasite from the host complement system (Acosta-Serrano *et al*, 2007). Type-III secretion chaperone is part of a complex that exports virulence factors from pathogenic intracellular prokaryotes to host cells (Galan and Wolf-Watz, 2006). Prior to the current analysis, the presence and phosphorylation of this complex in protozoan parasites had not yet been reported. However, the function of this complex in *T. cruzi* remains to be determined. Disperse gene family (DGF) protein 1 is a multigene family that was previously shown to be *N*-glycosylated (Atwood *et al*, 2006). Here, we show that members of this family are also phosphorylated. The InterPro annotation suggests a pectin-lyase fold/virulence factor function, but its role in *T. cruzi* virulence yet remains to be elucidated (Table 5.3).

5.4.3.8 Proteins with other functions

A number of the phosphoproteins identified in our analysis seem to bind ions and other proteins; however, their exact functions are unknown. In addition, several of the identified phosphoproteins are from the retrotransposon hot spot (RHS) family. RHS is encoded by a large multigene family localized mainly in telomeric regions of trypanosomal chromosomes, and it is believed to be involved in gene duplication of multigene families, including virulence factors such as mucins and TS/gp85 glycoproteins (Bringaud *et al*, 2002; Kim *et al*, 2005). The finding that several members of RHS family are phosphorylated suggests that these gene products are highly active, which accelerates gene duplication and evolution (Bringaud *et al*, 2002; Kim *et al*, 2005). This might explain the high divergence and differences in virulence between *T. cruzi* strains and phylogenetic lineages (Buscaglia and Di Noia, 2003; Kim *et al*, 2005).

5.4.4 Analysis of phosphotyrosine proteome

It is estimated that 30% of all eukaryotic proteins contain covalently bound phosphate at any given time with a ratio of 1800:200:1 for pSer, pThr, and pTyr, respectively (Venter *et al*, 2001). The rarity of phosphorylation on tyrosine residues suggests there is a higher gain in signaling pathways because they are more tightly regulated. A genome-wise prediction of protein kinases of trypanosomatid parasites (*T. cruzi*, *T. brucei*, and *Leishmania major*) has shown that these protozoa lack typical tyrosine kinases; however, the presence of tyrosine phosphorylated proteins has been reported by a number of research groups (Das *et al*, 1996; Favoreto *et al*, 1998). Das *et al.* showed that the major *T. brucei* tyrosine-phosphorylated protein is a nuclear RNA-binding protein (Nopp44/46) (Das *et al*, 1996). In *T. cruzi*, a 175-kDa

protein is tyrosine-phosphorylated upon contact with the host cells, but the identity and the function of this protein during host cell invasion remains unknown (Favoreto *et al*, 1998).

As described above, our MS/MS analysis identified 8 tyrosine phosphorylation sites from 7 distinct proteins: protein kinase (glycogen synthase kinase 3, GSK3) (Fig. 5.2); heterogeneous nuclear ribonucleoprotein H/F; dehydrogenase; phosphatidyl synthase; AT-hook motif family protein; and two hypothetical proteins EAN84292.1 and EAN94368.1 (Table 5.3). To further investigate tyrosine phosphorylation in *T. cruzi*, we performed an immunoprecipitation experiment using a monoclonal anti-phosphotyrosine antibody (see Material and Methods for details). Indeed, Western blotting analysis revealed 7 distinct tyrosine phosphorylated proteins; p250, p150, p90, p55, p50, p45, and p35, which were assigned names according to their relative molecular mass (Fig. 5.3). Interestingly, except for the p250, the predicted molecular masses of the MS/MS identified tyrosine phosphorylated proteins corresponded to those of the phosphoproteins detected by Western blot: protein kinase (GSK3) (40.4 kDa), heterogeneous nuclear ribonucleoprotein H/F (50.3 kDa), dehydrogenase (42.3 kDa), phosphatidyl synthase (60.0), AT-hook motif family protein (112 kDa), EAN84292.1 (103 kDa), and EAN94368.1 (61.3 kDa). However, due to other PTMs that could alter the relative mobility of these proteins during SDS-PAGE, the identity of these proteins cannot be inferred.

The identification of GSK3 phosphorylation at Y187 suggests that this signaling pathway is important in *T. cruzi* biology. This enzyme is conserved throughout evolution; however, the parasite sequences are slightly truncated (353-355 residues) compared to the human, mouse, and *Arabidopsis thaliana* GSK orthologs (410-420 residues) (Fig. 5.4). The amino acid sequence alignment also shows that *T. cruzi* GSK3 Y187 (arrow) is conserved in each of the included species (Fig 5.4). The corresponding tyrosine residue in human GSK3 (Y216) is in the “activation loop” of this kinase and its phosphorylation is critical for GSK3 catalytic activity

(Hughes *et al*, 1993; Wang *et al*, 1994). The significant proportion of GSK3 consensus phosphorylation sites (Table 5.2) further supports the presence of activated GSK3 in *T. cruzi*. Taken together, our results support the role of GSK3 in *T. cruzi* physiology and a selective inhibitor to this enzyme could potentially be used as an anti-parasitic agent. Indeed, GSK3 has been targeted for the development of selective drugs against *Plasmodium* *ssp.* and trypanosomatids (Droucheau *et al*, 2004; Knockaert *et al*, 2002).

5.5 Conclusions

In this study, we successfully mapped 247 phosphopeptides from 144 distinct proteins of epimastigote forms of *T. cruzi*. The results indicate that propagation of cell-signaling cascades by protein kinases and phosphatases play an important role in *T. cruzi* physiological processes, including cell motility, metabolism, ion transport, differentiation, survival, and pathogenesis. We are currently analyzing the phosphoproteomes of other developmental stages of *T. cruzi* to expand the fundamental knowledge of the mechanisms regulating this medically relevant protozoan parasite. In addition to contributing to the understanding of the molecular aspects of *T. cruzi* biology, the information presented here will aid in the development of potential kinase-directed therapeutic strategies to treat Chagas disease.

5.6 Acknowledgements

This work was funded by grants from the National Institutes of Health (5S06GM08012-37, 1R01AI070655, and 5G12RR008124). E.S.N. was supported by the George A. Krutilek memorial graduate scholarship from Graduate School, UTEP. M.R.G. was supported by the Research Experience for Undergraduates (REU) Program, UTEP (National Science Foundation grant # DBI-0353887). 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.

5.7 References

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**: D428-431.

Arslan MA, Kutuk O, Basaga H (2006) Protein kinases as drug targets in cancer. *Curr Cancer Drug Targets* **6**: 623-634.

Atwood JA, 3rd, Minning T, Ludolf F, Nuccio A, Weatherly DB, Alvarez-Manilla G, Tarleton R, Orlando R (2006) Glycoproteomics of *Trypanosoma cruzi* trypomastigotes using subcellular fractionation, lectin affinity, and stable isotope labeling. *J Proteome Res* **5**: 3376-3384.

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

Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* **101**: 12130-12135.

Bern C, Montgomery SP, Herwaldt BL, Rassi A, Jr., 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.

Brenchley R, Tariq H, McElhinney H, Szoor B, Huxley-Jones J, Stevens R, Matthews K, Taberner L (2007) The TriTryp phosphatome: analysis of the protein phosphatase catalytic domains. *BMC Genomics* **8**: 434.

Bringaud F, Biteau N, Melville SE, Hez S, El-Sayed NM, Leech V, Berriman M, Hall N, Donelson JE, Baltz T (2002) A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryot Cell* **1**: 137-151.

Broadhead R, Dawe HR, Farr H, Griffiths S, Hart SR, Portman N, Shaw MK, Ginger ML, Gaskell SJ, McKean PG, Gull K (2006) Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* **440**: 224-227.

Buscaglia CA, Di Noia JM (2003) *Trypanosoma cruzi* clonal diversity and the epidemiology of Chagas' disease. *Microbes Infect* **5**: 419-427.

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.

Cohen P, Goedert M (2004) GSK3 inhibitors: development and therapeutic potential. *Nat Rev Drug Discov* **3**: 479-487.

Cole DG (2003) The intraflagellar transport machinery of *Chlamydomonas reinhardtii*. *Traffic* **4**: 435-442.

da Cunha JP, Nakayasu ES, Elias MC, Pimenta DC, Tellez-Inon MT, Rojas F, Munoz MJ, Almeida IC, Schenkman S (2005) Trypanosoma cruzi histone H1 is phosphorylated in a typical cyclin dependent kinase site accordingly to the cell cycle. *Mol Biochem Parasitol* **140**: 75-86.

Das A, Peterson GC, Kanner SB, Frevert U, Parsons M (1996) A major tyrosine-phosphorylated protein of Trypanosoma brucei is a nucleolar RNA-binding protein. *J Biol Chem* **271**: 15675-15681.

DeGnore JP, Qin J (1998) Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J Am Soc Mass Spectrom* **9**: 1175-1188.

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.

Doerig C (2004) Protein kinases as targets for anti-parasitic chemotherapy. *Biochim Biophys Acta* **1697**: 155-168.

Droucheau E, Primot A, Thomas V, Mattei D, Knockaert M, Richardson C, Sallicandro P, Alano P, Jafarshad A, Baratte B, Kunick C, Parzy D, Pearl L, Doerig C, Meijer L (2004) Plasmodium falciparum glycogen synthase kinase-3: molecular model, expression, intracellular localisation and selective inhibitors. *Biochim Biophys Acta* **1697**: 181-196.

Dumontel 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, Wortley 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, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B (2005) The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. *Science* **309**: 409-415.

Favoreto S, Jr., Dorta ML, Yoshida N (1998) Trypanosoma cruzi 175-kDa protein tyrosine phosphorylation is associated with host cell invasion. *Exp Parasitol* **89**: 188-194.

Force T, Kuida K, Namchuk M, Parang K, Kyriakis JM (2004) Inhibitors of protein kinase signaling pathways: emerging therapies for cardiovascular disease. *Circulation* **109**: 1196-1205.

Galan JE, Wolf-Watz H (2006) Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **444**: 567-573.

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

Gnad F, Ren S, Cox J, Olsen JV, Macek B, Oroshi M, Mann M (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol* **8**: R250.

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* **36**: 3420-3435.

Hennig S, Groth D, Lehrach H (2003) Automated Gene Ontology annotation for anonymous sequence data. *Nucleic Acids Res* **31**: 3712-3715.

Hill KL (2003) Biology and mechanism of trypanosome cell motility. *Eukaryot Cell* **2**: 200-208.

Hughes K, Nikolakaki E, Plyte SE, Totty NF, Woodgett JR (1993) Modulation of the glycogen synthase kinase-3 family by tyrosine phosphorylation. *Embo J* **12**: 803-808.

Jones PM, George AM (2005) Multidrug resistance in parasites: ABC transporters, P-glycoproteins and molecular modelling. *Int J Parasitol* **35**: 555-566.

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.

Kim D, Chiurillo MA, El-Sayed N, Jones K, Santos MR, Porcile PE, Andersson B, Myler P, da Silveira JF, Ramirez JL (2005) Telomere and subtelomere of *Trypanosoma cruzi* chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of *T. cruzi* telomeres. *Gene* **346**: 153-161.

Knockaert M, Wieking K, Schmitt S, Leost M, Grant KM, Mottram JC, Kunick C, Meijer L (2002) Intracellular Targets of Paullones. Identification following affinity purification on immobilized inhibitor. *J Biol Chem* **277**: 25493-25501.

Kumar S, Boehm J, Lee JC (2003) p38 MAP kinases: key signalling molecules as therapeutic targets for inflammatory diseases. *Nat Rev Drug Discov* **2**: 717-726.

Lara FA, Sant'anna C, Lemos D, Laranja GA, Coelho MG, Reis Salles I, Michel A, Oliveira PL, Cunha ESN, Salmon D, Paes MC (2007) Heme requirement and intracellular trafficking in *Trypanosoma cruzi* epimastigotes. *Biochem Biophys Res Commun* **355**: 16-22.

Maldonado RA, Mirzoeva S, Godsel LM, Lukas TJ, Goldenberg S, Watterson DM, Engman DM (1999) Identification of calcium binding sites in the trypanosome flagellar calcium-acyl switch protein. *Mol Biochem Parasitol* **101**: 61-70.

Marques Porto R, Amino R, Elias MC, Faria M, Schenkman S (2002) Histone H1 is phosphorylated in non-replicating and infective forms of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **119**: 265-271.

Morales MA, Watanabe R, Laurent C, Lenormand P, Rousselle JC, Namane A, Spath GF (2008) Phosphoproteomic analysis of *Leishmania donovani* pro- and amastigote stages. *Proteomics* **8**: 350-363.

Naula C, Parsons M, Mottram JC (2005) Protein kinases as drug targets in trypanosomes and *Leishmania*. *Biochim Biophys Acta* **1754**: 151-159.

Parsons M, Worthey EA, Ward PN, Mottram JC (2005) Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics* **6**: 127.

Pereira-Chioccola VL, Acosta-Serrano A, Correia de Almeida I, Ferguson MA, Souto-Padron T, Rodrigues MM, Travassos LR, Schenkman S (2000) Mucin-like molecules form a negatively charged coat that protects *Trypanosoma cruzi* trypomastigotes from killing by human anti-alpha-galactosyl antibodies. *J Cell Sci* **113 (Pt 7)**: 1299-1307.

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.

Risso MG, Garbarino GB, Mocetti E, Campetella O, Gonzalez Cappa SM, Buscaglia CA, Leguizamon MS (2004) Differential expression of a virulence factor, the trans-sialidase, by the main *Trypanosoma cruzi* phylogenetic lineages. *J Infect Dis* **189**: 2250-2259.

Rosenzweig D, Smith D, Myler PJ, Olafson RW, Zilberstein D (2008) Post-translational modification of cellular proteins during *Leishmania donovani* differentiation. *Proteomics* **8**: 1843-1850.

Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ (2004) A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal Chem* **76**: 3590-3598.

Tyler KM, Engman DM (2001) The life cycle of *Trypanosoma cruzi* revisited. *Int J Parasitol* **31**: 472-481.

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

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratt E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C,

Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. *Science* **291**: 1304-1351.

Wang QM, Fiol CJ, DePaoli-Roach AA, Roach PJ (1994) Glycogen synthase kinase-3 beta is a dual specificity kinase differentially regulated by tyrosine and serine/threonine phosphorylation. *J Biol Chem* **269**: 14566-14574.

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 U S A* **105**: 5022-5027.

Yaffe MB, Elia AE (2001) Phosphoserine/threonine-binding domains. *Curr Opin Cell Biol* **13**: 131-138.

Table 5.1 Statistics of phosphoproteome analysis

	Non-phosphorylated proteins	Monophosphoproteins	Diphosphoproteins	Triphosphoproteins	Tetraphosphoproteins	Pentaphosphoproteins	Total phosphoproteins	
MS/MS	100	23	4	3	0	0	30	
MS/MS/MS	89	51	13	5	2	1	72	
MSA	73	81	20	10	3	0	114	
Total	122	97	21	22	3	1	144	
	Non-phosphorylated peptides	Phosphorylation sites	Monophosphopeptide	Diphosphopeptide	Triphosphopeptide	Total phosphopeptides	Reverse	FDR
MS/MS	142	39	40	1	0	41	3	1.64%
MS/MS/MS	135	101	106	3	3	112	7	2.75%
MSA	113	164	151	4	5	160	7	2.50%
Total	209	226	231	8	8	247	16	3.34%
	Phosphoserine	Phosphothreonine	Phosphotyrosine	Ser or Thr	Ser or Tyr			
Sites	163	49	8	5	1			
% from Total	72.1	21.7	3.5	2.2	0.4			

Table 5.2 Distribution of kinase specific motifs

Kinase (motif)	Number	% from total
CAMK2 (R-X-X- S/T)	49	14.3
CK1 (S-X-X- S/T)	41	12.0
PKA (R-X- S/T , R-R/K-X- S/T)	31	9.0
CK2 (S/T -X-X-E)	27	7.9
GSK3 (S -X-X-X-S)	27	7.9
ERK (V-X- S/T -P)	23	6.7
CDK1 (S/T -P-K/R)	21	6.1
NEK6 (L-X-X- S/T)	13	3.8
Aurora (R/K-X- S/T -I/L/V)	9	2.6
PKD (L/V/I-X-R/K-X-X- S/T)	8	2.3
CHK1 (M/I/L/V-X-R/K-X-X- S/T)	8	2.3
CDK2 (S/T -P-X-K/R)	8	2.3
Aurora-A (R/K/N-R-X- S/T -M/L/V/I)	4	1.2
AKT (R-R/S/T-X- S/T -X-S/T)	6	1.7
PLK1 (E/D-X- S/T -F/L/I/Y/W/V/M)	4	1.2
EGFR (D/P/S/A/E/N-X- Y -V/L/D/E/I/N/P)	3	0.9
ALK (Y -X-X-I/L/V/M)	2	0.6
SRC (E/D-X-X- Y -X-X-D/E/A/G/S/T)	1	0.3
S7CK1 (S/T-X-X-X- S)	1	0.3
S4PKA (R-R/K-X- S/T)	1	0.3
Unknown motif	56	16.3
Total	343	100.0

Table 5.3 Functional categorization of identified phosphoproteins

N	Protein description	Accession number	Phosphorylation sites	Blast Analysis		InterPro	Gene Ontology
				Sequence description	min. e-value		
1	regulatory subunit of protein kinase a-like protein	EAN92816.1	4	protein kinase a regulatory	0	Cyclic nucleotide-binding Signal recognition particle, SRP54 subunit, GTPase	protein serine/threonine kinase activity
2	hypothetical protein	EAN84292.1	3	hypothetical protein	0		
3	protein kinase	EAN91026.1	2	protein kinase	0	Serine/threonine protein kinase-related	protein serine/threonine kinase activity
4	hypothetical protein	EAN97116.1	2	calmodulin-related protein	4.94E-175		calcium ion binding
5	protein kinase-A catalytic subunit	AAL17691.2	1	protein kinase a catalytic subunit	0	Serine/threonine protein kinase	protein serine/threonine kinase activity
6	hypothetical protein	EAN83100.1	1	EF-hand domain (c-terminal) containing 1	0	Region of unknown function DM10	calcium ion binding
7	hypothetical protein	EAN83235.1	1	hypothetical protein	0	C2 calcium-dependent membrane targeting	
8	hypothetical protein	EAN85866.1	1	hypothetical protein	0	Sel1-like/Tetratricopeptide-like helical	binding
9	calmodulin	EAN86242.1	1	calmodulin	7.12E-79	Calcium-binding EF-hand	calcium ion binding
10	hypothetical protein	EAN88571.1	1	tor kinase binding protein	0	WD40/YVTN repeat-like	
11	phosphoprotein phosphatase	EAN90617.1	1	serine threonine protein phosphatase	0	Serine/threonine-specific protein phosphatase	hydrolase activity
12	hypothetical protein	EAN91302.1	1	chromosome 6 open reading frame 224	0	Shikimate kinase	ATP binding
13	hypothetical protein	EAN88376.1	1	hypothetical protein	0	C2 calcium-dependent membrane targeting	
14	rac serine-threonine kinase	EAN95526.1	1	rac serine-threonine	0	Zinc finger, RING/FYVE/PHD-type	protein serine/threonine kinase activity
15	protein kinase	EAN98751.1	1	GSK-3	0	Serine/threonine protein kinase	MAP kinase activity
16	hypothetical protein	EAN99395.1	1	ef hand family protein	0	Calcium-binding EF-hand	calcium ion binding
17	hypothetical protein	EAN95573.1	2	wd repeat domain 17	0	WD40/YVTN repeat-like	signal transducer activity
18	hypothetical protein	EAN96688.1	1	hypothetical protein	0	WD40 repeat-like	ATP binding/catalytic activity
19	hypothetical protein	EAN87422.1	1	erythrocyte binding	0		signal transducer activity
20	hypothetical protein	EAN88004.1	1	hypothetical protein	0		membrane/signal transducer activity
21	hypothetical protein	EAN93781.1	2	hypothetical protein	0	Forkhead-associated	

Cytoskeleton, flagellum and traffic proteins

22	kinesin	EAN86063.1	2	kinesin	0	Kinesin, motor region	microtubule associated complex
23	cofilin/actin depolymerizing factor	EAN88777.1	2	cofilin actin depolymerizing	3.04E-74	Actin-binding, cofilin/tropomyosin type	actin binding
24	beta tubulin 2.3	AAL75957.1	1	beta-tubulin	0	Tubulin	microtubule-based movement
25	paraflagellar rod component PAR1b	AAC32103.1	1	69 kda paraflagellar rod protein	0	Paraflagellar rod	flagellum (sensu Eukaryota)
26	paraflagellar rod protein PFR5	AAV28395.1	1	69 kda paraflagellar rod protein	0	Paraflagellar rod	flagellum (sensu Eukaryota)
27	mitotubule-associated protein Gb4	EAN95636.1	1	mitotubule-associated protein	0		microtubule-based movement
28	dynein heavy chain	EAN97081.1	1	dynein heavy	0	Dynein heavy chain, N-terminal region 2	microtubule-based movement
29	clathrin coat assembly protein	EAN98227.1	1	clathrin coat assembly dc2-related axonemal dynein intermediate chain 4	0	Clathrin adaptor, phosphoinositide-binding, GAT-like	phospholipid binding
30	hypothetical protein	EAN84227.1	1		0		cytokinesis
31	hypothetical protein	EAN93004.1	1	myosin heavy chain	0		
32	hypothetical protein	EAN93132.1	1	spermatogenesis associated 4	0	Protein of unknown function DUF1042	
33	hypothetical protein	EAN95700.1	1	hypothetical protein af154847_133 kda vamp-associated protein	0	Major sperm protein	structural molecule activity
34	hypothetical protein	EAN82309.1	5		3.00E-129	Major sperm protein	structural molecule activity
35	hypothetical protein	EAN99904.1	3	loc495400 protein	1.14E-128	Major sperm protein	structural molecule activity
36	hypothetical protein	EAN81574.1	1	ankyrin repeat	0	Ankyrin	cytoskeleton/signal transduction

Transporters

37	ABC transporter	EAN92058.1	3	ABC transporter	0	ABC transporter-like	permease activity
38	ABC transporter	EAN89676.1	2	ABC transporter	0	ABC transporter-like	permease activity
39	ABC transporter	EAN96763.1	1	ABC transporter	0	ABC transporter-like	permease activity
40	Na+-ATPase	BAC98847.1	1	calcium motive p-type	0	ATPase, P-type, K/Mg/Cd/Cu/Zn/Na/Ca/Na/H -transporter	monovalent inorganic cation transport
41	metal-ion transporter	EAN85275.1	1	zinc transporter-like protein	0	Cation efflux protein	cation transporter activity
42	chloride channel protein	EAN94356.1	1	chloride channel	0	Cystathionine beta-synthase, core	voltage-gated chloride channel activity
43	GPR1/FUN34/yaaH family	EAN82352.1	1	gpr1 fun34 yaaH family protein	1.48E-132	GPR1/FUN34/yaaH (acetate	membrane

					transporter)	
44	calcium channel protein	EAN97848.1	1	calcium channel	0	Ion transport voltage-gated potassium channel
DNA, RNA and protein turnover						
45	heterogeneous nuclear ribonucleoprotein H/F	EAN91678.1	5	heterogeneous nuclear ribonucleoprotein H/F	0	RNA recognition motif, RNP-1 protein/nucleic acid binding
46	GTP-binding elongation factor Tu family	EAN90873.1	3	GTP-binding protein 1	0	Translation elongation factor EFTu/EF1A GTP binding
47	hypothetical protein	EAN95745.1	3	chromosome segregation protein	0	cytolysis
48	hypothetical protein	EAN99674.1	2	protein transport protein sec31	0	Nucleic acid/zinc binding
49	Tcc1a22.3	AAL82703.1	1	elongation factor 1-gamma (ef-1-gamma)	0	translation elongation factor activity
50	hypothetical protein	EAN81890.1	1	subunit of chromatin modifying protein	2.89E-64	Dpy-30
51	25 kDa translation elongation factor 1-beta	EAN82455.1	1	translation elongation factor 1-	1.84E-77	Translation elongation factor EF1B translation elongation factor activity
52	histone H2B	EAN83534.1	1	histone H2B	8.33E-56	Histone H2B
53	ATP-dependent RNA helicase	EAN96171.1	1	ATP-dependent RNA helicase	0	DNA/RNA helicase, C-terminal nucleosome ATP-dependent helicase activity
54	tyrosyl or methionyl-tRNA synthetase	EAN96489.1	1	tyrosyl-tRNA synthetase	1.65E-125	tRNA-binding region RNA recognition motif, RNP-1 nucleic acid/tRNA binding
55	hypothetical protein	EAN95162.1	1	hypothetical protein	0	nucleotide/RNA binding
56	nucleosome assembly protein	EAN95410.1	1	nucleosome assembly	6.73E-155	Nucleosome assembly protein (NAP) nucleosome assembly
57	hypothetical protein	EAN89957.1	3	nod3 protein	0	Leucine-rich repeat
58	hypothetical protein	EAN94862.1	3	hypothetical protein	0	nucleotide binding
59	hypothetical protein	EAN96978.1	1	hypothetical protein	0	nucleic acid/zinc ion binding
60	hypothetical protein	EAN86654.1	1	pab1 binding protein (poly-A binding)	0	Ataxin-2, N-terminal
61	hypothetical protein	EAN96152.1	2	calpain-like cysteine peptidase	0	proteolysis
62	calpain-like cysteine peptidase	EAN92790.1	2	calpain-like cysteine peptidase	3.98E-136	Region of unknown function DUF1935 proteolysis
63	calpain cysteine peptidase	EAN83138.1	1	calpain cysteine peptidase	0	Peptidase C2, calpain proteolysis
64	hypothetical protein	EAN87586.1	1	calpain-like cysteine peptidase	0	proteolysis
65	calpain-like cysteine peptidase	EAN88142.1	1	calpain-like cysteine peptidase	0	Peptidase C2, calpain proteolysis
66	proteasome regulatory non-ATPase subunit	EAN92953.1	1	proteasome regulatory non-ATPase	1.99E-153	von Willebrand factor, type A / Ubiquitin interacting motif protein complex
67	glutamamyl carboxypeptidase	EAN85499.1	1	acetylornithine deacetylase	5.39E-149	Peptidase M20 proteolysis
68	Heat shock protein 70	CAA47952.1	1	heat shock protein 70	0	Heat shock protein Hsp70 protein folding

Metabolism

69	ribose-phosphate pyrophosphokinase	EAN92422.1	3	ribose-phosphate pyrophosphokinase	0	Phosphoribosyltransferase Pyruvate phosphate dikinase, PEP/pyruvate-binding	ribose phosphate diphosphokinase activity
70	pyruvate phosphate dikinase 2	AAG12986.1	1	pyruvate phosphate dikinase	0	Pyruvate dehydrogenase (acetyl-transferring) E1 component	kinase activity
71	pyruvate dehydrogenase E1 alpha subunit	AAD11551.1	1	pyruvate dehydrogenase e1 component alpha	0		oxidoreductase activity, acting on the aldehyde 1-aminocyclopropane-1-carboxylate synthase activity
72	B Chain Of Tyrosine Aminotransferase	1BW0	1	tyrosine aminotransferase NADH:flavin oxidoreductase NADH oxidase	0	Tyrosine aminotransferase	
73	dehydrogenase NADH-dependent fumarate reductase	AAX55669.1	1		0	Aldolase-type TIM barrel	oxidoreductase activity succinate
74		EAN82648.1	1	NADH-dependent fumarate ubiquinone menaquinone biosynthesis-related protein	0	ApbE-like lipoprotein	dehydrogenase activity methyltransferase activity
75	hypothetical protein	EAN95143.1	1		0	Methyltransferase type 11	ribose phosphate diphosphokinase activity
76	phosphoribosylpyrophosphate synthetase	EAN97393.1	1	phosphoribosylpyrophosphate	0	Phosphoribosyl pyrophosphokinase	ribose phosphate diphosphokinase activity
77	ribose-phosphate pyrophosphokinase	EAN95566.1	1	ribose-phosphate pyrophosphokinase	0	Phosphoribosyltransferase HAD-superfamily hydrolase, subfamily IIA, CECR5	hydrolase activity
78	hypothetical protein	EAN95748.1	1	phosphatidyl synthase transcriptional regulatory protein not1	0	CCR4-Not complex component, Not1	carboxylic acid metabolism
79	hypothetical protein	EAN88126.1	1		0	HMG-I and HMG-Y, DNA-binding	
80	hypothetical protein	EAN91502.1	3	AT hook motif family protein	0		fatty acid metabolism

Pathogenesis

81	trans-sialidase	EAN91292.1	3	trans-sialidase	0	Concanavalin A-like lectin/glucanase / Neuraminidase	pathogenesis
82	trans-sialidase	EAN91764.1	1	trans-sialidase type III secretion system chaperone	0	Concanavalin A-like lectin/glucanase / Neuraminidase	pathogenesis
83	hypothetical protein dispersed gene family protein 1 (DGF-1)	EAN92340.1	1	dispersed gene family protein 1 (dgf-1)	3.56E-177	Ubiquitin	pathogenesis/regulation of protein secretion
84	dispersed gene family protein 1 (DGF-1)	EAN86627.1	1	dispersed gene family protein 1 (dgf-1)	0	Pectin lyase fold/virulence factor	membrane
85	dispersed gene family protein 1 (DGF-1)	EAN97078.1	1		0	Parallel beta-helix repeat	membrane

Binding

86	hypothetical protein	EAN97824.1	3	c11orf60 protein	0		protein binding
87	hypothetical protein	EAO00106.1	2	proteophosphoglycan ppg4	0		protein binding
88	hypothetical protein	EAN92411.1	2	p25-alpha family protein	1.42E-71	P25-alpha	calcium ion binding
89	I/6 autoantigen	EAN91318.1	1	I/6 autoantigen	1.18E-78		calcium ion binding
90	I/6 autoantigen	EAN94637.1	1	I/6 autoantigen	9.00E-79		calcium ion binding
91	hypothetical protein	EAN83610.1	1	mgc84420 protein	6.81E-152	Protein of unknown function DUF298	ubiquitin binding
92	hypothetical protein	EAN90373.1	1	mucoidy inhibitor a	0	Conserved hypothetical protein CHP02231	protein binding
93	hypothetical protein	EAN99066.1	3	hypothetical protein	0		calcium ion binding
94	hypothetical protein	EAN86750.1	1	hypothetical protein	0		integral to membrane
95	hypothetical protein	EAN98546.1	1	hypothetical protein	0		calcium/heat shock protein binding
96	hypothetical protein	EAN99719.1	1	hypothetical protein	0		binding
97	hypothetical protein	EAN94135.1	3	hypothetical protein	0		binding
Other functions							
98	retrotransposon hot spot (RHS) protein	EAN86249.1	2	retrotransposon hot spot	0	Trypanosome RHS	
99	retrotransposon hot spot (RHS) protein	EAN94562.1	2	retrotransposon hot spot	0	Trypanosome RHS	
100	retrotransposon hot spot (RHS) protein	EAN99306.1	2	retrotransposon hot spot	0	Trypanosome RHS	
101	retrotransposon hot spot (RHS) protein	EAN81665.1	1	retrotransposon hot spot	0	Trypanosome RHS	
102	retrotransposon hot spot (RHS) protein	EAN94997.1	1	retrotransposon hot spot	0	Trypanosome RHS	
103	hypothetical protein	EAN85046.1	1	hypothetical multipass transmembrane protein	0		integral to membrane
104	hypothetical protein	EAN93909.1	1	hypothetical multipass transmembrane protein	0		integral to membrane
105	peroxin 14	EAN89658.1	1	peroxin	3.31E-173	Peroxisome membrane anchor protein Pex14p	peroxisome
106	hypothetical protein	EAN89431.1	1	cingulin	0		
107	hypothetical protein	EAN91356.1	1	glycoprotein 96- t-complex-associated testis expressed 1	0		
108	hypothetical protein	EAN92548.1	1		0	Leucine-rich repeat	
Proteins without known function or domain							
109	hypothetical protein	EAN93581.1	4	hypothetical protein	4.37E-156		
110	hypothetical protein	EAN94661.1	4	hypothetical protein	0		
111	hypothetical protein	EAN85364.1	3	hypothetical protein	6.49E-119		

112	hypothetical protein	EAN88585.1	3	hypothetical protein	0
113	hypothetical protein	EAN90722.1	3	hypothetical protein	0
114	hypothetical protein	EAN91901.1	3	hypothetical protein	0
115	hypothetical protein	EAN94097.1	3	hypothetical protein	1.10E-105
116	hypothetical protein	EAN94312.1	3	hypothetical protein	0
117	hypothetical protein	EAN95418.1	3	hypothetical protein	0
118	hypothetical protein	EAN96967.1	3	hypothetical protein	2.91E-122
119	hypothetical protein	EAN99202.1	3	hypothetical protein	0
120	hypothetical protein	EAN83060.1	2	hypothetical protein	1.75E-111
121	hypothetical protein	EAN91358.1	2	hypothetical protein	3.99E-50
122	hypothetical protein	EAN94368.1	2	hypothetical protein	0
123	hypothetical protein	EAN95229.1	2	hypothetical protein	0
124	hypothetical protein	EAN99051.1	2	hypothetical protein	1.64E-143
125	hypothetical protein	EAN99070.1	2	hypothetical protein	8.34E-168
126	hypothetical protein	EAN82852.1	1	hypothetical protein	0
127	hypothetical protein	EAN85405.1	1	hypothetical protein	0
128	hypothetical protein	EAN86037.1	1	hypothetical protein	1.90E-68
129	hypothetical protein	EAN86093.1	1	hypothetical protein	0
130	hypothetical protein	EAN94518.1	1	hypothetical protein	0
131	hypothetical protein	EAN90456.1	1	hypothetical protein	0
132	hypothetical protein	EAN93058.1	1	hypothetical protein	8.09E-67
133	hypothetical protein	EAN95093.1	1	hypothetical protein	0
134	hypothetical protein	EAN95103.1	1	hypothetical protein	0
135	hypothetical protein	EAN95354.1	1	hypothetical protein	0
136	hypothetical protein	EAN96262.1	1	hypothetical protein	6.41E-90
137	hypothetical protein	EAN96518.1	1	hypothetical protein	3.60E-180
138	hypothetical protein	EAN97191.1	1	hypothetical protein	0
139	hypothetical protein	EAN98665.1	1	hypothetical protein	0
140	hypothetical protein	EAN99075.1	1	hypothetical protein	0
141	hypothetical protein	EAN99770.1	1	hypothetical protein	0
142	hypothetical protein	EAN99804.1	1	hypothetical protein	3.24E-173
143	hypothetical protein	EAN99811.1	1	hypothetical protein	1.78E-163
144	hypothetical protein	EAN80998.1	1	hypothetical protein	0

Table 5.4 Gene Ontology analysis using GOblet algorithm

N	Level	GO	Name	Count	% from Total
1	1	GO:0003674	Molecular Function	80	55.6
2	2	GO:0003774	motor activity	3	2.1
3	3	GO:0003777	microtubule motor activity	3	2.1
4	2	GO:0003824	catalytic activity	38	26.4
5	3	GO:0004386	helicase activity	1	0.7
6	3	GO:0016491	oxidoreductase activity	5	3.5
7	3	GO:0016740	transferase activity	14	9.7
8	3	GO:0016787	hydrolase activity	18	12.5
9	3	GO:0016829	lyase activity	2	1.4
10	2	GO:0004871	signal transducer activity	4	2.8
11	3	GO:0005057	receptor signaling protein activity	2	1.4
12	2	GO:0005198	structural molecule activity	4	2.8
13	2	GO:0005215	transporter activity	8	5.6
14	3	GO:0005386	carrier activity	5	3.5
15	3	GO:0015075	ion transporter activity	8	5.6
16	3	GO:0015267	channel or pore class transporter activity	2	1.4
17	3	GO:0015646	permease activity	2	1.4
18	3	GO:0043492	ATPase activity, coupled to movement of substances	5	3.5
19	2	GO:0005488	binding	56	38.9
20	3	GO:0000166	nucleotide binding	23	16.0
21	3	GO:0003676	nucleic acid binding	11	7.6
22	3	GO:0005515	protein binding	14	9.7
23	3	GO:0008289	lipid binding	1	0.7
24	3	GO:0043167	ion binding	15	10.4
25	2	GO:0030234	enzyme regulator activity	1	0.7
26	3	GO:0019207	kinase regulator activity	1	0.7
27	2	GO:0045182	translation regulator activity	2	1.4
28	3	GO:0008135	translation factor activity, nucleic acid binding	4	2.8
29	1	GO:0005575	Cellular Component	39	27.1
30	2	GO:0005623	cell	38	26.4
31	3	GO:0005622	intracellular	23	16.0
32	3	GO:0009986	cell surface	1	0.7
33	3	GO:0016020	membrane	16	11.1
34	3	GO:0042995	cell projection	2	1.4
35	3	GO:0045177	apical part of cell	1	0.7
36	2	GO:0043226	organelle	13	9.0
37	3	GO:0043227	membrane-bound organelle	6	4.2
38	3	GO:0043228	non-membrane-bound organelle	8	5.6
39	3	GO:0043229	intracellular organelle	13	9.0
40	2	GO:0043234	protein complex	10	6.9
41	3	GO:0000786	nucleosome	2	1.4
42	3	GO:0001518	voltage-gated sodium channel complex	1	0.7
43	3	GO:0005853	eukaryotic translation elongation factor 1 complex	4	2.8
44	3	GO:0005875	microtubule associated complex	5	3.5
45	3	GO:0005891	voltage-gated calcium channel complex	10	6.9

46	3	GO:0005941	unlocalized protein complex	1	0.7
47	3	GO:0008076	voltage-gated potassium channel complex	1	0.7
48	3	GO:0016459	myosin	1	0.7
49	1	GO:0008150	Biological Process	48	33.3
50	2	GO:0000003	reproduction	1	0.7
51	3	GO:0019953	sexual reproduction	1	0.7
52	2	GO:0007275	development	1	0.7
53	3	GO:0009653	morphogenesis	1	0.7
54	3	GO:0009790	embryonic development	1	0.7
55	3	GO:0030154	cell differentiation	1	0.7
56	3	GO:0035295	tube development	1	0.7
57	3	GO:0048513	organ development	1	0.7
58	3	GO:0048731	system development	1	0.7
59	2	GO:0007582	physiological process	45	31.3
60	3	GO:0008152	metabolism	35	24.3
61	3	GO:0016265	death	1	0.7
62	3	GO:0040011	locomotion	1	0.7
63	3	GO:0050791	regulation of physiological process	1	0.7
64	3	GO:0050874	organismal physiological process	1	0.7
65	3	GO:0050875	cellular physiological process	42	29.2
66	3	GO:0051179	localization	16	11.1
67	2	GO:0009987	cellular process	45	31.3
68	3	GO:0007154	cell communication	4	2.8
69	3	GO:0030154	cell differentiation	1	0.7
70	3	GO:0050875	cellular physiological process	42	29.2
71	2	GO:0050789	regulation of biological process	1	0.7
72	3	GO:0050791	regulation of physiological process	1	0.7
73	2	GO:0050896	response to stimulus	4	2.8
74	3	GO:0006950	response to stress	1	0.7
75	3	GO:0007610	behavior	2	1.4
76	3	GO:0009605	response to external stimulus	2	1.4
77	3	GO:0009607	response to biotic stimulus	1	0.7
78	3	GO:0009628	response to abiotic stimulus	3	2.1
79	2	GO:0051704	interaction between organisms	3	2.1
80	3	GO:0044419	interspecies interaction between organisms	3	2.1

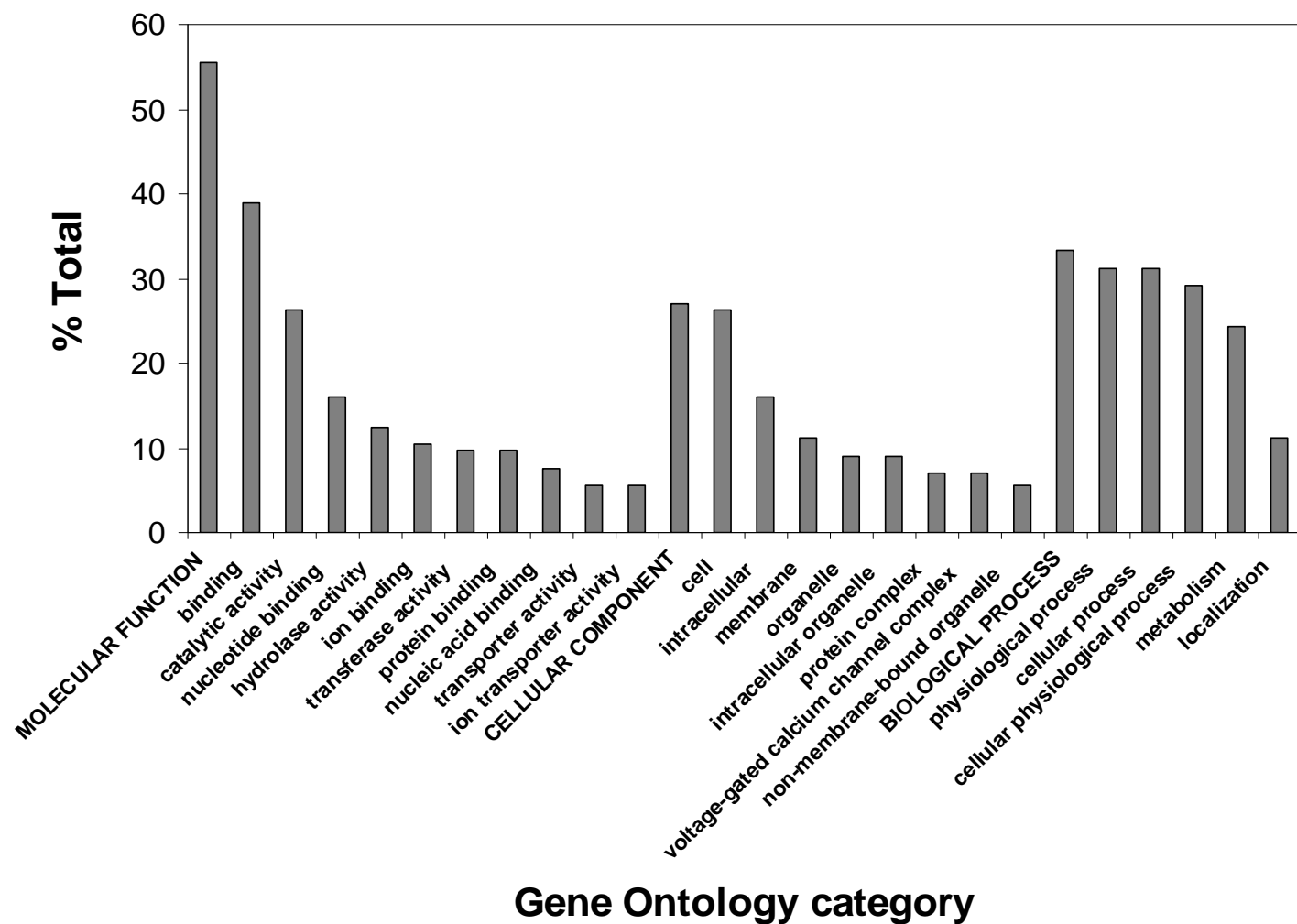


Figure 5.1 Gene ontology analysis. GO analysis was performed using GOblet algorithm. The representative categories for “molecular function”, “cellular component”, and “biological process” are shown in the graph. For the complete list, see Table 4.

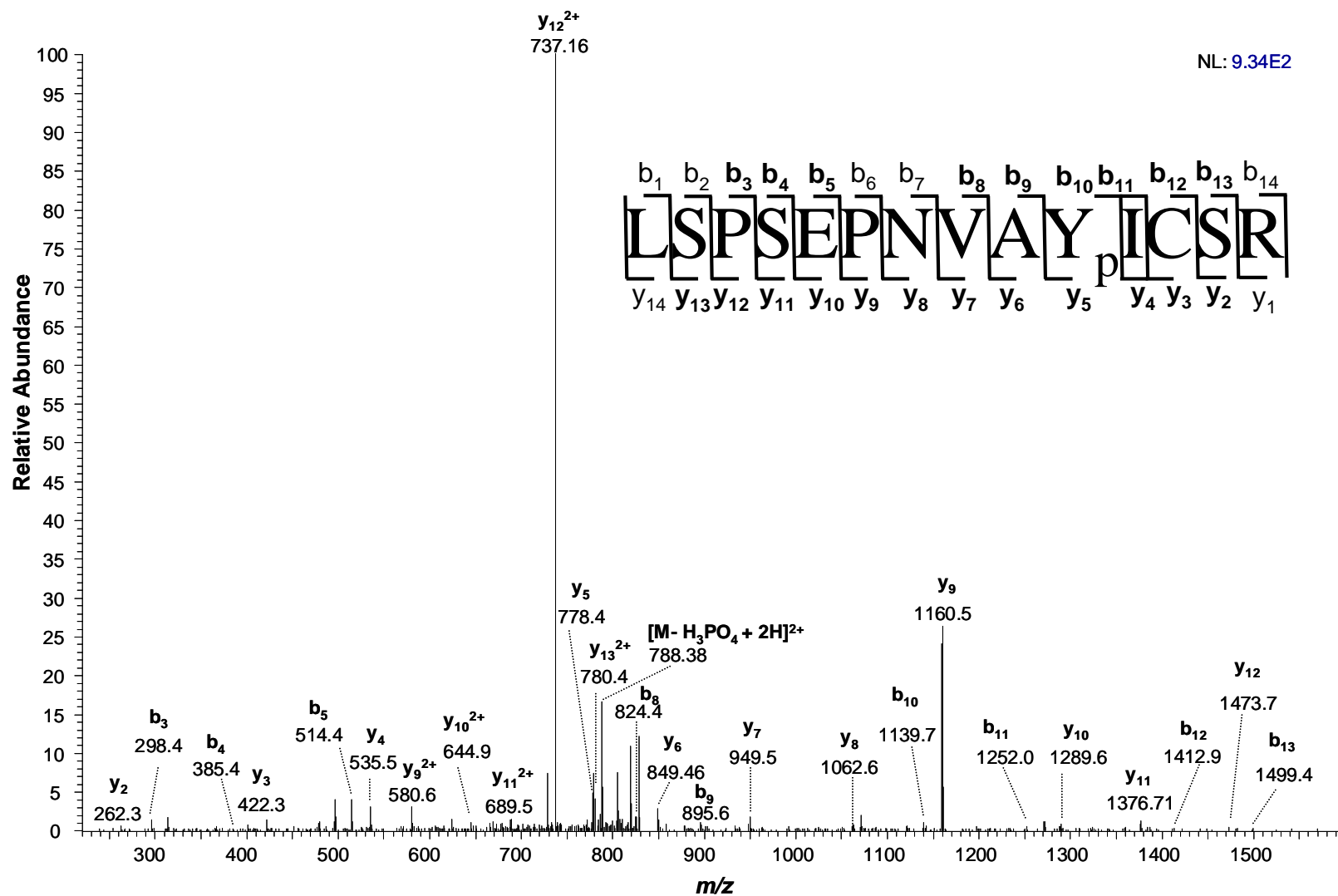


Figure 5.2 MS/MS spectrum from the tyrosine-phosphorylated peptide LSPSEPNVAYpICSR from glycogen synthase kinase 3 (GSK3). Matched b and y fragments are shown in bold.

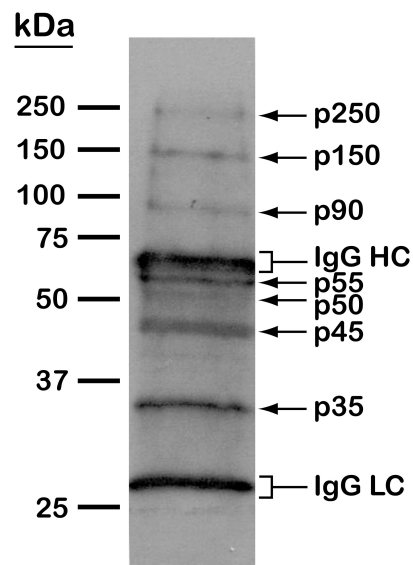


Figure 5.3 Western blotting analysis of tyrosine phosphorylated proteins in *T. cruzi*. Epimastigote extracts were subjected to anti-phosphotyrosine immunoprecipitation and separation by 10% SDS-PAGE. After blocking with BSA, the membrane was probed with the anti-phosphotyrosine antibody followed by detection with horseradish peroxidase conjugated anti-mouse IgG and chemiluminescence. Arrows denote the positions of the tyrosine phosphorylated proteins. Brackets denote the locations of the Immunoglobulin G heavy chains (IgG HC) and light chains (IgG LC)


```

T.cruzi      : -----MSLNLAEEADE-----SQREMEKYLVERVAGCGTCTVQLAREKNTGVNVAIKKVIQDPREKKNRELQIMQDLSRLRHHENIVLKNWFYTV : 86
T. brucei    : -----MSLNLTDAADDR-----SYKEMEKYTVVERVAGCGTCTVQLARDKSTGSLVAIKKVIQDPREKKNRELQIMQHLARLRHHENIVMLKNWFYTV : 86
L.mexicana   : -----MSLNAAADAADER-----SRKEMDRFLVERMAGCGTCTVQLGREKSTGMSVAIKKVIQDPREKKNRELQIMQDLAVLHHEINIVQLQSFYTVL : 86
mouse        : --MSGRPRTTSFA-----ESCKPVQPSAFGSMKVSRRD-----DGSKVTT--VVATPGQGPDRP--QEVSYTDTKVIKNGSPGVVYQARLDCSGELVAIKKVLQDKREKKNRELQIMR---KLDHCNIVRLRYEFYSS : 119
human        : -----TTSFA-----ESCKPVQPSAFGSMKVSRRD-----DGSKVTT--VVATPGQGPDRP--QEVSYTDTKVIKNGSPGVVYQARLDCSGELVAIKKVLQDKREKKNRELQIMR---KLDHCNIVRLRYEFYSS : 113
A.thaliana   : MASVGTLPASSMATKQSNASICAELPEGINEMKIKDDMEMEAAVVDGNGTETGHIIVTTIGGKNGQPKQTI SYMAERIVGCGSFCIVPQAKCLETGETVAIKKVLQDKRYKKNRELQIMR---LLDHEENVVSLKHCFEST : 137

                                         ↓

T.cruzi      : LGDHEHNDVYLVNVVMEFVPETLHRCORAMVRRLASPEHILVKVFYQQLRGTACLHLPAVNVCHRDIKPEHNVLVDESGDLKLCDFGSAKKLSPEENVAIKCSRYRRAPELIPGNOYYSTAVDVWSVGCIPAEMLLGEF : 226
T. brucei    : GGEGRNDVYLVNVVMEFVPETLHRCORAMVRRMTNPEHILVKVFYQQLRGTACLHLPAVNVCHRDIKPEHNVLVDESGDLKLCDFGSAKKLSPEENVAIKCSRYRRAPELIPGNOFYTTAVDINSVGCIPAEMLLGEF : 226
L.mexicana   : GERDRR-DIYLVNVVMEFVPETLHRCORAMVRRQVAPPEHILVKVFYQQLRGTACLHLPSNVVCHRDIKPEHNVLVNEAEGTLKLCDFGSAKKLSPEENVAIKCSRYRRAPELIPGNOHYTTSVDINSVGCIPAEMLLGEF : 225
mouse        : --GBRKEDEVYLVNLVLDYVPETVYRVARHYSRAKQTLQVYVVKLYYQQLERSLAYIHS--FGICHRDIKPEONLLDPDTAVLKLCDPFGSAKQIVRGEPNVSYICSRYRRAPELIPGATDYTSSIDVWSAGCVLAELLGQF : 255
human        : --GBRKEDEVYLVNLVLDYVPETVYRVARHYSRAKQTLQVYVVKLYYQQLERSLAYIHS--FGICHRDIKPEONLLDPDTAVLKLCDPFGSAKQIVRGEPNVSYICSRYRRAPELIPGATDYTSSIDVWSAGCVLAELLGQF : 249
A.thaliana   : --TEK-DELYLVNLVLEFVPETVYRVARHYSRAQRMPETIYVVKLYYQQLERSLAYIHS--FGVCHRDIKPEONLLVNPHTQVVKLCDFGSAKVIVKGEPNISYICSRYRRAPELIPGATEYTTTIDINSAGCVLAELLGQF : 273

T.cruzi      : IFCGENTSGQIREIVRVLGRPSREELHKLSTSNVNLNVTNNSTPWEDVFKRP---LAEVYDLCATIFKYLEBORITPLEALCHFFFDDELHATVKLPSNALPAHLFTLELPEETNEMTEAQRSOLVQK----- : 353
T. brucei    : IFCGENTSGQIREIVKILCKPTKEELHKLNGSSTEINAN-ARATPWENVEKQP---LAEVYDLCGATIFKYVVDQRIETPIDALCHFFFNELRETTKLPSNPPLPAHLVQETPDEVEAMTEAQREYLLKK----- : 352
L.mexicana   : IFRGDSAGQDHEIVRVLGCPSREVLKRLNPSHTDVDLYNSAGIPWSTVECDHSLKDAKEAYDLSALLQYLEDNRMKYEALCHBYFDELHSATKLENHKNLEEDLFRLESEIEVMSQAQKAKLVK----- : 355
mouse        : IFRGDSGVDQIVEIKVLCPTREBOIREMNPNTYEFKFPQIKAHPTWKVFRPR---TPEAIALCSRLLEYTETARLTPLEACAHSEFFDELQDENVKLPNCRDTEA-LFNETTQELS-SNPPLATILIPPHARIQAAASP : 390
human        : IFRGDSGVDQIVEIKVLCPTREBOIREMNPNTYEFKFPQIKAHPTWKVFRPR---TPEAIALCSRLLEYTETARLTPLEACAHSEFFDELQDENVKLPNCRDTEA-LFNETTQELS-SNPPLATILIPPHARIQAAAST : 384
A.thaliana   : LFRGDSGVDQIVEIKVLCPTREBOIREMNPNTYEFKFPQIKAHPTWKVFRPR---TPEAVDQVSRLLQYSNLRSTAMEALVHFFFDDELQDENTRLPNCRALEP-LFNKPEQLKGALELLSKLIPDHARKQCSFLA : 409

T.cruzi      : ----- : -
T. brucei    : ----- : -
L.mexicana   : ----- : -
mouse        : PANATAASDTNAGDRGQTNNAAASASASNST : 420
human        : PTNATAASDANTGDRGQTNNAAASASASNST : 414
A.thaliana   : L----- : 410

```

Figure 5.4 Alignment of glycogen synthase kinase 3 from several organisms: protozoa: *T. cruzi*, *T. brucei* and *L. mexicana*; higher eukaryotes: mouse and human; and plant: *Arabidopsis thaliana*. The sequences were aligned using ClustalW, available in BioEdit v7.0.9.0 and exported to GeneDoc. Arrow denotes the MS/MS identified phosphorylated tyrosine.

FINAL DISCUSSION AND CONCLUDING REMARKS

In this dissertation I focused in the analysis of glycoconjugates of the *T. cruzi* surface and secreted vesicles (TcαGalves), as well in the analysis of the parasite phosphoproteome. First, the proteomic analysis of bloodstream trypomastigote secreted vesicles was carried out (Chapter 1). These vesicles were shown to be rich in members of TS/gp85 superfamily and other putative GPI-anchored proteins, such as gp63, MASP and TcMUC II mucins. Next, the GPIomic analysis showed that epimastigotes are covered by a high abundant and diverse layer of GIPLs. In addition, these analyses have shown that the major surface glycoproteins from epimastigote are the mucin TcSMUG S (Chapter 2). Next, we compared membrane proteins from epimastigotes and metacyclic trypomastigotes by extracting them with detergent and analyzing by LC-MS/MS (Chapter 3). In the Chapter 4, a large-scale proteomic analysis was performed with the whole bloodstream trypomastigote lysate. The results from this chapter showed that bloodstream trypomastigotes express hundreds, perhaps more than one thousand, surface proteins including members of TcMUC II mucins, TS/gp85, MASP, mucin-like, gp63, and ToIT families. The last part of my dissertation (Chapter 5) comprises the analysis of the phosphoproteome of epimastigote forms of *T. cruzi*.

T. cruzi has a very different glycocalyx depending on the developmental stage (Fig. 6.1 and 6.2). For instance, epimastigotes have high amounts of GIPLs and express short and highly glycosylated GPI-anchored polypeptides, such as TcSMUG S mucins and NETNES sequences (Chapter 2, and Macrae *et al*, 2005) (Fig. 6.1 and 6.2). Interestingly, the epimastigote GPI-anchors were shown to have ether phospholipids as the lipid tails, while the GIPLs had mainly ceramide lipids (Chapter 2). The function of the GIPLs and mucins in the interaction of the parasite with the insect vector is still poorly explored. The absence of long,

unmodified (non-glycosylated) polypeptide sequences in the cell surface of epimastigotes may be defense mechanism of the parasite against the proteases present in the kissing bug midgut (Fig. 6.1 and 6.2). For instance, some preliminary data from our laboratory has shown that the TcSMUG S mucins are almost completely resistant to trypsin and other proteases (Almeida *et al.*, unpublished data). On the other hand, a recent report has shown that epimastigote GPIs help the parasite to attach to the insect midgut (Nogueira *et al.*, 2007). Moreover, GPIs from *Plasmodium falciparum* and *Toxoplasma gondii* were shown to activate mosquito innate immune leading to the overexpression of antimicrobial peptides (Arrighi *et al.*, 2008).

Metacyclic trypomastigotes have mucins with similar migration in SDS-PAGE compared to the epimastigote TcSMUG S, but they have not been identified yet [reviewed in (Buscaglia *et al.*, 2006)]. However, the major difference in the GPI-anchors from the metacyclic trypomastigote stage is the presence of ceramide as the lipid tail, which has not been so far described in epimastigotes (Serrano *et al.*, 1995). In addition to the mucins, metacyclic trypomastigotes have proteins related to host-cell invasion, such as gp82 and gp90 (Yoshida, 2006) (Fig. 6.1 and 6.2). Proteomic analysis of detergent-solubilized membrane proteins (Chapter 3), and biotinylated and affinity purified surface proteins (Marques, Nakayasu and Almeida, unpublished data) also supports these information (Fig. 6.1 and 6.2).

Little is known about the composition of the surface proteins and glycolipids from amastigotes. For instance, immunolabeling experiments have shown that the major glycoproteins seem to be from the mucin TcMUC I family (Campo *et al.*, 2006). A proteomic analysis from the surface proteins from all four developmental stages of the parasite, suggest that the amastigotes express more proteins compared to epimastigotes or metacyclic trypomastigotes, but less compared to bloodstream trypomastigotes (Marques, Nakayasu and

Almeida, unpublished data). The most abundant proteins seems to be members of TS/gp85 and MASP superfamilies (Fig. 6.1 and 6.2).

Thus far, bloodstream trypomastigotes seem to have the most complex glycocalyx. Although the GPIome of this stage has not been fully characterized, there is evidence that bloodstream trypomastigotes have many more different species of protein-linked GPI anchors compared to the epimastigote ones (Almeida *et al*, 2000; Buscaglia *et al*, 2004) (Fig. 6.1 and 6.2). Although to date only GPI anchors with alkylacylglycerol as the lipid tail have been described, a big proportion of the acyl chains are unsaturated (Almeida *et al*, 2000; Buscaglia *et al*, 2004), while in epimastigotes they are mostly saturated (Chapter 2). The proteomic analysis of the whole cell lysate (Chapter 3) and biotinylated and affinity-purified surface proteins (Marques, Nakayasu and Almeida, unpublished data), have been shown that TcMUC II mucins, TS/gp85, MASP, mucin-like, gp63, and ToIT families are present on the parasite plasma membrane. The biggest challenge now is to determine the abundance of each of those proteins on the cell surface. Although by protein coverage members of TS/gp85 superfamily are likely to be highly expressed, we can not affirm that they are most abundant ones. The problem is the uncertainty of number of genes expressed and the fact the mucins, MASPs, and mucin-like apparently are highly glycosylated, thus decreasing the proteome coverage. An interesting alternative to study these surface molecules is by GPIomic analysis (Chapter 2). Since the GPI structures are more conserved, the ionization of these species by electrospray are very similar (Chapter 2), thus having a better correlation between the sample amount and the signal in the mass spectrometer.

Taken together, our data and others from the literature indicate that the trypomastigote stage of *T. cruzi* plasma membrane is covered by a much more complex coat of surface glycoproteins compared to other protozoan parasites. *T. brucei* is coated mainly by a clonally

expressed variant surface glycoprotein (VSG) (bloodstream trypomastigote forms) or procyclin (promastigote forms), but also contain minor proteins as TS and gp63, and protein free GPIs (glycoinositolphospholipids – GIPLs) (Acosta-Serrano *et al*, 2006). *Leishmania* *ssp.* are covered by gp63 and proteophosphoglycans (PPG), and glycolipids, such as GIPLs and lipophosphoglycans (LPGs) (Acosta-Serrano *et al*, 2006). This difference in the diversity of structures of surface molecules may reflect the mechanism used by each parasite to evade the host immune response. *T. cruzi* expresses a great variety of glycoproteins leading to a “dilution” of cell surface epitopes. So the immune response is raised against some isoforms and not against all of them (Martin *et al*, 2006; Ratier *et al*, 2008). It is still to be determined whether the parasite changes its coat in the course of the infection, and what are the surface coat variations between different strains and phylogenetic lineages. However, it is also important to point out that *T. cruzi* infection produces high titers of anti- α Gal antibodies that control the parasitemia (Almeida *et al*, 1994a, b; Almeida *et al*, 1993; Almeida *et al*, 1991; Gazzinelli *et al*, 1991). In contrast to *T. cruzi*, *T. brucei* expresses only a single VSG gene, which is switched in the course of the infection (Taylor and Rudenko, 2006). Therefore, when antibodies are raised against one VSG isoform leading to a decrease of the parasitemia, the parasite switches the expression to new isoform of VSG leading to a persisting infection (Taylor *et al*, 2006). Unlike *T. cruzi* or *T. brucei* infection, antibodies produced against *Leishmania* molecules fail to provide a protective immune response. Moreover, the production of immunoglobulin G has been correlated to the production of IL-10 and the progression of the infection (Miles *et al*, 2005).

Bloodstream *T. cruzi* trypomastigotes also shed membrane vesicles (Tc α GalVes) rich in surface antigens. By proteomic analysis and immunoassays, these vesicles were shown here to be rich in members of TS/gp85 superfamily and α -galactosyl containing glycans (Chapter 1).

MAASP, gp63, and TcMUC II mucins were also detected, but since a quantitative analysis was not performed, we cannot infer about the abundance of those proteins (Chapter 1). Tc α GalVes have been shown to strongly activate the proinflammatory response by activating TLR2-mediated pathways. Surprisingly, this activation was shown to increase the parasite infectivity *in vitro* (Torrecilhas and Nakayasu *et al.*, unpublished results). It is also possible that these vesicles have other functions. A recent report in *T. brucei* showed that the hydrodynamic flow generated by the parasite swimming helps the parasite to eliminate antibodies against VSG (Engstler *et al.*, 2007). It is possible that Tc α GalVes are been formed by a similar mechanism, thus removing antibodies from the cell surface of trypomastigotes. Another possibility is that trypomastigotes shed Tc α GalVes as decoys to the strong lytic antibody response. Tc α GalVes are rich in shed-acute-phase-antigens (SAPA), which are TS that have tandem repeats in the C-terminal. These tandem repeats were shown to be highly immunogenic, thus the antibody production is directed to this region and not to the enzyme catalytic site (Buscaglia *et al.*, 1999).

We are currently analyzing the large-scale and surface proteome, the GPIome, and the glycome of other stages and strains of *T. cruzi*. This will allow us to obtain a more complete qualitative and quantitative profile of the molecules expressed on the parasite surface. This information would be essential for the discovery of conserved epitopes among different glycoconjugates that could be tested as vaccine candidates for experimental Chagas' disease. To date, most of the tested antigens were based on single polypeptide sequences, which were not even checked whether they are expressed by the parasite. Thus, these vaccines showed only a partial protection against the parasite infection (Garg and Bhatia, 2005).

Next, since kinases sequences were found in the proteomic analysis of Tc α GalVes, we speculate that protein phosphorylation would be an important PTM in these vesicles. However, lack of a straightforward and universal method to analyze phosphoproteome lead us to test the

method first in epimastigotes, which are easier to obtain in high amounts in cell-free cultures. The analysis of the epimastigote phosphoproteome showed a possible involvement of protein phosphorylation in many physiological processes, some of them seem to be attractive targets for chemotherapeutic development. In the evolutionary point of view, it is interesting that 4% of the mapped phosphorylation sites were found in tyrosine residues, even *T. cruzi* does not have typical tyrosine kinase or tyrosine kinase-like genes (Naula *et al*, 2005; Parsons *et al*, 2005). We are currently investigating the phosphoproteome of different stages of the parasite and time-points during the cell differentiation. Furthermore, we intend to analyze the phosphorylated proteins of TcαGalVes. This will provide us a profile of phosphorylation sites and kinases involved in different aspects of the parasite physiology, such as cell growth, differentiation, and virulence process.

References:

Acosta-Serrano A, Hutchinson C, Nakayasu ES, Almeida IC, Carrington M (2006) Comparison and evolution of the surface architecture of trypanosomatid parasites. In *Trypanosomes: After the genome*: Horizon Press.

Almeida IC, Camargo MM, Procopio DO, Silva LS, Mehler 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, Ferguson MA, Schenkman S, Travassos LR (1994a) GPI-anchored glycoconjugates from *Trypanosoma cruzi* trypomastigotes are recognized by lytic anti-alpha-galactosyl antibodies isolated from patients with chronic Chagas' disease. *Braz J Med Biol Res* **27**: 443-447.

Almeida IC, Ferguson MA, Schenkman S, Travassos LR (1994b) Lytic anti-alpha-galactosyl antibodies from patients with chronic Chagas' disease recognize novel O-linked oligosaccharides on mucin-like glycosyl-phosphatidylinositol-anchored glycoproteins of *Trypanosoma cruzi*. *Biochem J* **304 (Pt 3)**: 793-802.

Almeida IC, Krautz GM, Krettli AU, Travassos LR (1993) Glycoconjugates of *Trypanosoma cruzi*: a 74 kD antigen of trypomastigotes specifically reacts with lytic anti-alpha-galactosyl antibodies from patients with chronic Chagas disease. *J Clin Lab Anal* **7**: 307-316.

Almeida IC, Milani SR, Gorin PA, Travassos LR (1991) Complement-mediated lysis of *Trypanosoma cruzi* trypomastigotes by human anti-alpha-galactosyl antibodies. *J Immunol* **146**: 2394-2400.

Arrighi RB, Debierre-Grockiego F, Schwarz RT, Faye I (2008) The immunogenic properties of protozoan glycosylphosphatidylinositols in the mosquito *Anopheles gambiae*. *Dev Comp Immunol*.

Buscaglia CA, Alfonso J, Campetella O, Frasch AC (1999) Tandem amino acid repeats from *Trypanosoma cruzi* shed antigens increase the half-life of proteins in blood. *Blood* **93**: 2025-2032.

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.

Campo VA, Buscaglia CA, Di Noia JM, Frasch AC (2006) Immunocharacterization of the mucin-type proteins from the intracellular stage of *Trypanosoma cruzi*. *Microbes Infect* **8**: 401-409.

Engstler M, Pfohl T, Herminghaus S, Boshart M, Wiegertjes G, Heddergott N, Overath P (2007) Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell* **131**: 505-515.

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

Gazzinelli RT, Pereira ME, Romanha A, Gazzinelli G, Brener Z (1991) Direct lysis of *Trypanosoma cruzi*: a novel effector mechanism of protection mediated by human anti-gal antibodies. *Parasite Immunol* **13**: 345-356.

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.

Martin DL, Weatherly DB, Laucella SA, Cabinian MA, Crim MT, Sullivan S, Heiges M, Craven SH, Rosenberg CS, Collins MH, Sette A, Postan M, Tarleton RL (2006) CD8⁺ T-Cell responses to *Trypanosoma cruzi* are highly focused on strain-variant trans-sialidase epitopes. *PLoS Pathog* **2**: e77.

Miles SA, Conrad SM, Alves RG, Jeronimo SM, Mosser DM (2005) A role for IgG immune complexes during infection with the intracellular pathogen *Leishmania*. *J Exp Med* **201**: 747-754.

Naula C, Parsons M, Mottram JC (2005) Protein kinases as drug targets in trypanosomes and Leishmania. *Biochim Biophys Acta* **1754**: 151-159.

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.

Parsons M, Worthey EA, Ward PN, Mottram JC (2005) Comparative analysis of the kinomes of three pathogenic trypanosomatids: Leishmania major, Trypanosoma brucei and Trypanosoma cruzi. *BMC Genomics* **6**: 127.

Ratier L, Urrutia M, Paris G, Zarebski L, Frasch AC, Goldbaum FA (2008) Relevance of the diversity among members of the Trypanosoma cruzi trans-sialidase family analyzed with camelids single-domain antibodies. *PLoS ONE* **3**: e3524.

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.

Taylor JE, Rudenko G (2006) Switching trypanosome coats: what's in the wardrobe? *Trends Genet* **22**: 614-620.

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

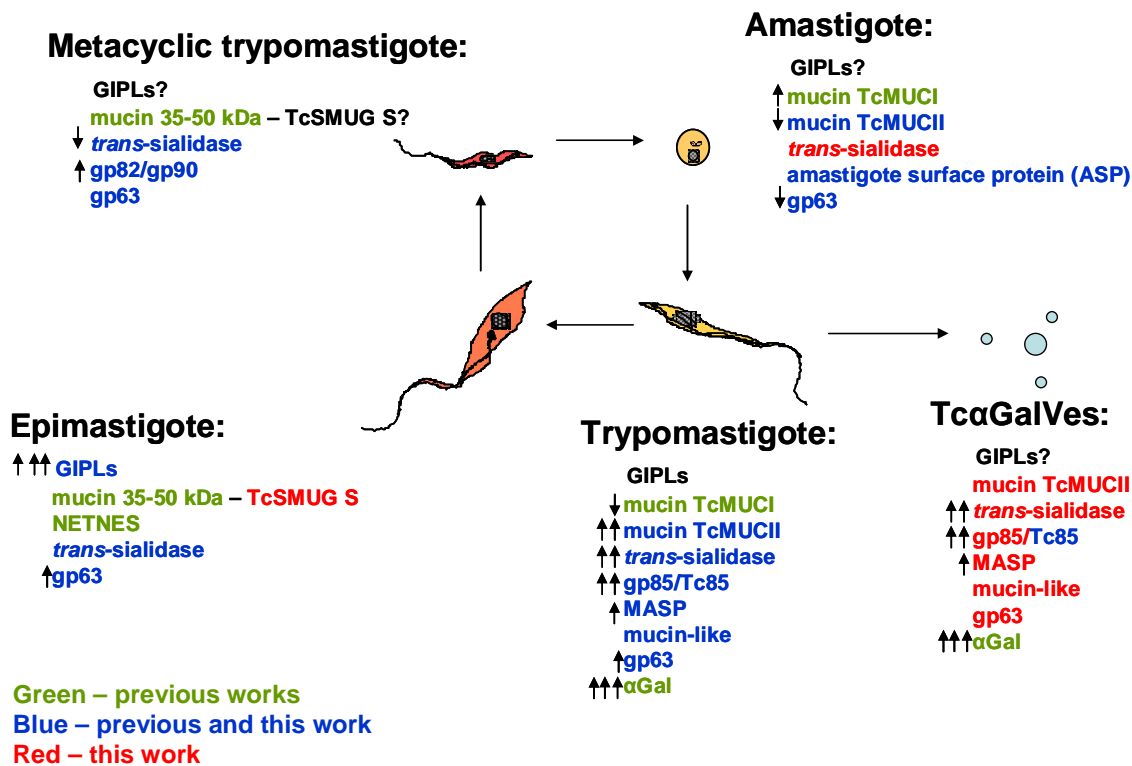


Figure 6.1 Overview of *T. cruzi* cell surface glycoconjugates. Epimastigotes are rich in GIPLs and short highly glycosylated GPI-anchored polypeptides, whereas the metacyclic trypomastigotes also present gp82 and gp90. Little information is known about the amastigote surface molecules, but bloodstream trypomastigotes have been shown to be rich in TcMUCII, *trans*-sialidases, gp85, Tc85, MASP and αGal epitopes. TcαGalVes express the same classes of molecules compared to trypomastigote glycocalyx, but the composition was shown to be different (Chapter 1 and 4).

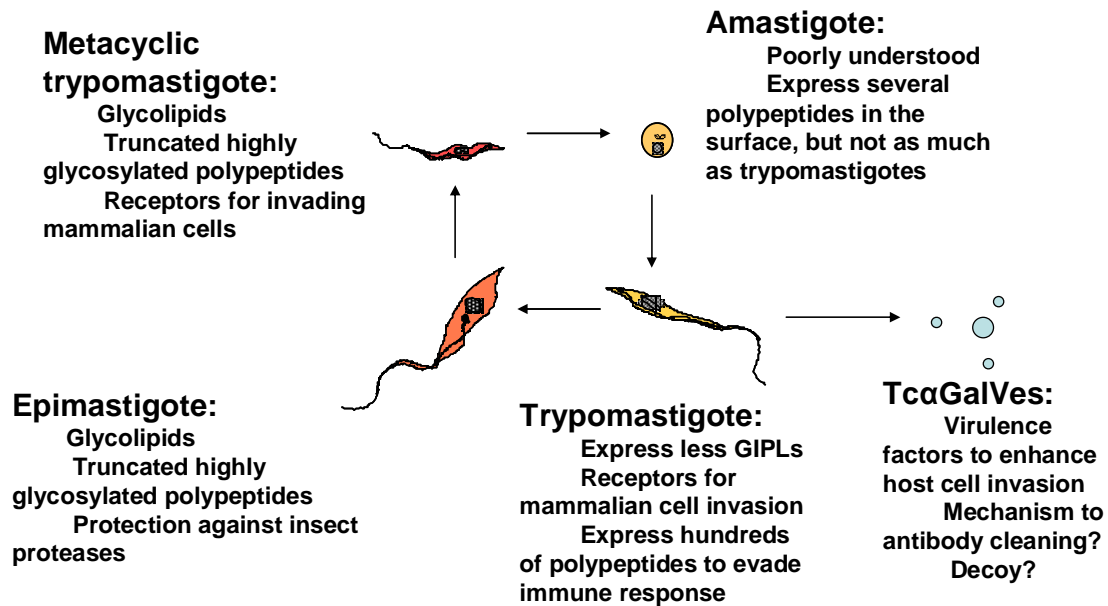


Figure 6.2 Function overview of *T. cruzi* cell surface glycoconjugates. Epimastigotes are rich in GIPLs and short highly glycosylated GPI-anchored polypeptides, which may be important to protect against proteases from the insect midgut. The metacyclic trypomastigotes more than these short glycoconjugates, present receptors involved in the host cell invasion, gp82 and gp90. Little information is known about the amastigote surface molecules, but bloodstream trypomastigotes have been shown to be rich in TcMUCII, *trans*-sialidases, gp85, Tc85, MASP and α Gal epitopes. This highly diverse glycocalyx is involved in the host cell invasion and evasion from the immune response. Tc α GalVes were shown to enhance host cell invasion, but there is also the possibility that they are involved in other processes, such as mechanism of antibody cleaning and act as decoy molecules to direct the immune response to them and not to the whole parasite.

Appendix

1 List of Publications

1. da Cunha, J.P., **Nakayasu, E.S.**, Elias, M.C., Pimenta, D.C., Tellez-Inon, M.T., Rojas, F., Manuel, M., Almeida, I.C., and Schenkman, S. (2005) *Trypanosoma cruzi* histone H1 is phosphorylated in a typical cyclin dependent kinase site accordingly to the cell cycle. ***Mol. Biochem. Parasitol.***, 140, 75-86.
2. Nimrichter, L., Cerqueira, M.D., Leitão, E.A., Miranda, K., **Nakayasu, E.S.**, Almeida, S.R., Almeida, I.C., Alviano, C.S., Barreto-Bergter, E. and Rodrigues, M.L. (2005) Structure, cellular distribution, antigenicity and biological functions of *Fonsecaea pedrosoi* ceramide monohexosides. ***Infect. Immun.*** 73: 7860–7868.
3. Paiva-Silva, G.O., Cruz-Oliveira, C., **Nakayasu, E.S.**, Maya-Monteiro, C.M., Dunkov, B.C. Masuda, H., Almeida, I.C., Oliveira, P.L. (2006) A novel heme degradation pathway in a blood sucking insect. ***Proc. Natl. Acad. Sci. USA***, 103(21):8030-5.
4. Acosta-Serrano, A., Hutchinson, C., **Nakayasu, E.S.**, Almeida, I.C., and Carrington, M. (2006) Comparison and evolution of the surface architecture of trypanosomatid parasites. In ***Trypanosomes - After the Genome***. D. Barry, J. Mottram, R. McCulloch, and A. Acosta-Serrano (Eds.), Horizon Scientific Press, Norwich, UK. Chapter 10.
5. da Cunha, J.P., **Nakayasu, E.S.**, I.C., and Schenkman, S. (2006) Post-translational modifications of *Trypanosoma cruzi* histone H4. ***Mol. Biochem. Parasitol.*** 150 (2):268-77.
6. **Nakayasu, E.S.**, and Almeida, I.C. (2007) Proteomics studies in *Trypanosoma cruzi*. In ***Bioinformatics in Tropical Disease Research: A Practical Approach***. H. Del Portillo and C. Huynh (Eds.), National Center for Biotechnology Information (NCBI), Bethesda, MD. Available at: <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=bioinfo.chapter.B05>

7. Jurado, J.D., Rael, E.D., Lieb, C.S., **Nakayasu, E.S.**, Hayes, W.K., Bush, S.P. and Ross, J.A. (2007) Complement inactivating proteins and intraspecies venom variation in *Crotalus oreganus helleri*, ***Toxicon***, 49 (3): 339-50.
8. Matsuo, A.L., Carmona, A.K., Silva, L.S., Cunha, C.E., **Nakayasu, E.S.**, Almeida, I.C., Juliano, M.A. and Pucia, R. (2007) C-Npys (S-3-nitro-2-pyridinesulfonyl) and peptide derivatives can inhibit a serine-thiol proteinase activity of *Paracoccidioides brasiliensis*. ***Biochem. Biophys. Res. Commun.***, 355(4):1000-5.
9. Rodrigues, M.L., **Nakayasu, E.S.**, Oliveira, D.L., Nimrichter, L., Nosanchuk, J.D., Almeida, I.C., Casadevall, A. (2008) Extracellular vesicles produced by *Cryptococcus neoformans* contain protein components associated with virulence. ***Eukaryot Cell***. 7(1):58-67.
10. Fridberg, A., Olson, C.L., **Nakayasu, E.S.**, Tyler, K.M., Almeida, I.C., Engman, D.M. (2008) Sphingolipid synthesis is necessary for kinetoplast segregation and cytokinesis in *Trypanosoma brucei*. ***J Cell Sci.*** 121(Pt 4):522-35.
11. Albuquerque, P.C., **Nakayasu, E.S.**, Rodrigues, M.L., Frases, S., Casadevall, A., Zancoppe-Oliveira, R.M., Almeida, I.C., Nosanchuk, J.D. (2008) Vesicular transport in *Histoplasma capsulatum*: an effective mechanism for trans-cell wall transfer of proteins and lipids in ascomycetes. ***Cell Microbiol.*** 10(8):1695-710.
12. Looi, K.S., **Nakayasu, E.S.**, Diaz, R.A., Tan, E.M., Almeida, I.C., Zhang, J.Y. (2008) Using proteomic approach to identify tumor-associated antigens as markers in hepatocellular carcinoma. ***J Proteome Res.*** 7(9):4004-12.
13. Dhiman, M., **Nakayasu, E.S.**, Madaiah, Y.H., Reynolds, B.K., Wen, J.J., Almeida, I.C., Garg, N.J. (2008) Enhanced nitrosative stress during *Trypanosoma cruzi* infection causes

- nitrotyrosine modification of host proteins: implications in Chagas' disease. **Am J Pathol.** 173(3):728-40.
14. Sant'Anna, C.*, Nakayasu, E.S.*, Pereira, M.G., Lourenço, D., de Souza, W., Almeida, I.C., Cunha-e-Silva, N.L. (2008) Subcellular proteomics of *Trypanosoma cruzi* reservosomes. Submitted to **Proteomics** (in press).
 15. Yichoy, M.*, Nakayasu, E.S.*, Shpak, M., Aguilar C., Aley, S.B., Almeida, I.C., Das, S. (2008) Lipidomic Analysis Reveals That Phosphatidylglycerol and Phosphatidylethanolamine are Synthesized *de novo* in the Early-Divergent Protozoon *Giardia lamblia*. Submitted to **Mol Biochem Parasitol** (Accepted for publication).
 16. Nakayasu, E.S., Yashunsky, D.V., Nohara, L.L., Torrecilhas, A.C.T., Nikolaev, A.V., Almeida, I.C. (2008) GPIomics: Global Analysis of Glycosylphosphatidylinositol-Anchored Molecules of *Trypanosoma cruzi*. Submitted to **Mol Sys Biol** (under review).
 17. Venkatakrishnan, V., Nakayasu, E.S., Almeida, I.C., Miller, R.T. (2008) Absence of nitric oxide synthase in sequentially purified rat liver mitochondria. Submitted to **J. Biol. Chem.** (under review).
 18. Cordero E.M.*, Nakayasu E.S.*, Gentil L.G., Yoshida N., Almeida I.C., Franco da Silveira J. (2008) Proteomic analysis of the GPI-enriched membrane extracts of epimastigotes and metacyclic trypomastigotes of *Trypanosoma cruzi*. Submitted to **J Proteome Res** (under review)
 19. Nakayasu, E.S.*, Gaynor, M.R.*, Ross, J.A., and Almeida, I.C. (2008) Phosphoproteomic analysis of the human pathogen *Trypanosoma cruzi* at the epimastigote stage. Submitted to **Proteomics** (under review).
 20. Torrecilhas, A.C.T.*, Nakayasu, E.S.*, Nohara, L.L., Gozzo, F.C., Jacysyn, J.F., Maldonado, R.A., Ropert, C., de Souza, W., Cunha e Silva, N., Andrade, D., Scharfstein,

- J., Golenbock, D., Gazzinelli, R.T., Colli, W., Ferguson, M.A.J., Alves, M.J.M., and Almeida, I.C. (2008) *Trypanosoma cruzi* shed vesicles enhance host cell invasion by engaging Toll-like receptor 2. In preparation.
21. Ganiko, L., **Nakayasu, E.S.** and Almeida, I.C. (2008) Role of host galectin-3 and annexin A2-S100A10 heterotetramer in the cell invasion by *Trypanosoma cruzi*. In preparation.
22. **Nakayasu, E.S.**, Sobreira, T.J., Torres Jr., R., Ganiko, L., Oliveira, P.S.L., Marques, A.F. and Almeida, I.C. (2008) Improved proteomic approach for the discovery of potential vaccine targets in *Trypanosoma cruzi*. In preparation.

* These authors contributed equally to the work.

2 List of Abbreviations

α Gal – alpha-galactosyl

AAG – alkylacylglycerol

ACN – acetonitrile

AEP – aminoethylphosphonate

anti-Gal – anti-alpha-galactosyl antibodies

ASP – amastigote surface protein

BCA – bicinchoninic acid

BM – bicarbonate/methanol buffer

BSA – bovine serum albumin

Cer – ceramide

CID – collision-induced dissociation

CL-ELISA – chemiluminescent enzyme-linked immunosorbent assay

CLA – chemiluminescent assay

CRP – complement regulatory protein

DCn – normalized difference between cross-correlation scores

DDA – data-dependent acquisition

DMEM – Dulbecco's modified Eagle's medium

DTT – dithiothreitol

EDTA – ethylenediaminetetraacetic acid

eEF1B – translation elongation factor 1-beta

ELISA – enzyme-linked immunosorbent assay

ESCRT – endosomal sorting complexes required for transport

ESI – electrospray ionization

EtNP – ethanolaminephosphate

FA – formic acid

FBS – fetal bovine serum

FCaBP – flagellar calcium-binding protein

FCS – fetal calf serum

FDR – false discovery rate

GAPDH – glyceraldehyde 3-phosphate dehydrogenase

GC-MS – gas chromatography-mass spectrometry

GIPL – glycoinositolphospholipid

GlcN – glucosamine

GPI – glycosylphosphatidylinositol

GO – gene ontology

Gro - glycerol

Hex – hexose

HexN - hexosamine

HIC – hydrophobic-interaction chromatography

HILIC – hydrophilic-interaction chromatography

HPLC – high-performance liquid chromatography

HSP – heat-shock protein

HTS – high-throughput screening

IA – iodoacetamide

IMAC – immobilized-metal affinity chromatography

InsP – inositolphosphate

IPC – inositolphosphorylceramide

IT – ion trap

LC-MS – liquid chromatography-tandem mass spectrometry

LIT – liver-infusion tryptose medium

LPG – lipophosphoglycan

Man – mannose

MAPKs – mitogen-activated protein kinases

MASP – mucin-associated surface protein

MDLC – multidimensional liquid chromatography

MHC – major histocompatibility complex

MAb – monoclonal antibody

MRM – multiple reaction monitoring

MS – mass spectrometry

MSA – multistage activation

MSn – n^{th} -order mass spectrometry analysis

MVB – multivesicular bodies

MyD88 – myeloid differentiation primary response gene 88

NANA – *N*-acetylneuraminic acid

NMR – nuclear magnetic resonance

NO – nitric oxide

PBS – phosphate-buffered saline solution

PI – phosphatidylinositol

PTFE - polytetrafluoroethylene

PMSF – phenylmethylsulfonyl fluoride

PPG – proteophosphoglycan

PVDF – polyvinylidene fluoride

PTM – post-translational modification

QTOF – quadrupole time-of-flight

RHS – retrotransposon hot spot

RP – reverse phase

RPC – reverse phase chromatography

SAP – serine-, alanine-, and proline-rich protein

SAPA – shed acute-phase antigen

SAX – strong anion-exchange

SCX – strong cation-exchange

SDS-PAGE – sodium dodecyl sulfate-polyacrylamide gel-electrophoresis

T α GalVes – *Trypanosoma cruzi* alpha-galactosyl-rich vesicles

TFA – trifluoroacetic acid

TIC – total-ion chromatogram

TG – trypsin and endoproteinase Glu-C digestion

TLR – Toll-like receptor

TM – trypsin digestion in methanol-containing buffer

TMS – trimethylsilyl

TPR – tetratricopeptide-repeat

TU – trypsin digestion in urea-containing buffer

TS – *trans*-sialidase

TX-114 – triton X-114

VSG – variant surface glycoprotein

Xcorr – cross-correlation score

2D LC-MS/MS – two-dimensional liquid chromatography coupled to tandem mass spectrometry

Vita

Ernesto Satoshi Nakayasu received his degree in Biological Sciences from University of Sao Paulo, Brazil. In 2003, he joined two of Brazil's pioneer proteomic and mass spectrometry groups, led by Drs. Sirlei Daffre and Igor Almeida, at the Department of Parasitology, University of Sao Paulo. In 2004, spent 2 months training at one of most prestigious European proteomic facilities, directed by Prof. Michael Ferguson at the Wellcome Trust Biocentre, University of Dundee, Scotland, UK.

In January 2005 he joined the Biological Sciences PhD program at University of Texas at El Paso (UTEP), El Paso, Texas. He is conducting his thesis under the supervision of Dr. Igor Almeida. His dissertation is focused on the molecular characterization of proteins, glycolipids and protein posttranslational modifications of *Trypanosoma cruzi* (etiological agent of Chagas' disease), which might have a great impact for the rational development of vaccines and other therapies against Chagas' disease.

He has attended national and international conferences, and presented poster at meetings, such as, the Meeting of the American Society for Biochemistry and Molecular Biology (2006) and Gordon Conference in the Biology of Host-Parasite Interactions (2006 and 2008). He also presented his work in form of short talks during the Kinetoplastid Biology (2005), the Rio Grande Branch of American Society for Microbiology (2006) and Brazilian Society for Mass Spectrometry (2007) meetings. He was awarded as the best oral presentation from a PhD student during the Meeting Rio Grande Branch of American Society for Microbiology (2006); and best PhD student poster during the UTEP chapter of Society for Advancement of Chicanos and Native Americans in Science (2007). He also was awarded the George A. Krutilek memorial scholarship from Graduate School, UTEP (2006-2008).

He has already published 13 per-reviewed articles in high impact factor journals, most of them in collaboration with other research groups from Brazil, United States and United Kingdom. He is also the author and co-author of two book chapters.

After defending his thesis, he wishes to apply for a post-doctoral position in a prestigious research center. He wants to keep working with different aspects of the structure of biomolecules and their relationship with biological activities. The overall goal of his career is to become a university professor and keeping in the research area. More importantly, he would like to share his knowledge and skills with other students.

Permanent address: Avenida Nossa Senhora Assuncao, 647, Apto. 25C.

Jardim Bonfiglioli, Sao Paulo, SP, Brazil, CEP 05359-001

This dissertation was typed by Ernesto Satoshi Nakayasu