Published in final edited form as:
Proteomics. 2009 July ; 9(13): 3489–3506. doi:10.1002/pmic.200800874.

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

Ernesto S. Nakayasu[1,*], **Matthew R. Gaynor**[1,*], **Tiago J.P. Sobreira**[2], **Jeremy A. Ross**[1], and **Igor C. Almeida**[1]

[1] The Border Biomedical Research Center, Department of Biological Sciences, University of Texas at El Paso, El Paso, TX, USA

[2] Group of Computational Biology, Laboratory of Genetics and Molecular Cardiology, Heart Institute (InCor). Av. Dr. Eneas de Carvalho Aguiar, 44, Bloco II, 10° andar, 05403-000, Sao Paulo, SP, Brazil

Abstract

Trypanosoma cruzi is the etiologic agent of Chagas disease, which affects millions of people in Latin America and has become a public health concern in the United States and areas of 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 of phosphoproteins from *T. cruzi* epimastigote forms. Our liquid chromatography (LC)-tandem MS (MS/MS, MS/MS/MS, and multistage activation) analysis has identified 237 phosphopeptides from 119 distinct proteins. Furthermore, 220 phosphorylation sites were unambiguously mapped: 148 on serine, 57 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*. We discuss the potential roles of the identified phosphoproteins in parasite physiology and drug development.

Keywords

Cell Signaling; Chagas disease; Drug targets; Phosphoproteome; Trypanosoma cruzi

1 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

Correspondence: Dr. Igor C. Almeida. The Border Biomedical Research Center, 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.
*These authors contributed equally to this work.

CONFLICT OF INTEREST
The authors have declared no conflict of interest.

to 50,000 people may die every year due to complications in the acute or chronic phases of the disease [1,2]. In addition, due to the migration of chronically infected, asymptomatic people from endemic areas and lack of screening in blood banks, Chagas disease has become a public health concern in United States and certain areas of Europe, such as Catalonia [3,4].

Chagas disease is caused by the protozoan parasite *Trypanosoma cruzi*, which is naturally transmitted by hematophagous Reduviidae insects, popular known as kissing bugs. The parasite, however, may also be transmitted congenitally or via blood transfusion or organ transplantation [1]. 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 excreta during a bloodmeal, and reach the host bloodstream through the bite wound or exposed ocular or oral mucosa. Inside the host, the parasite can infect different types of nucleated cells, where they immediately escape the parasitophorous vacuole and differentiate into amastigote forms, which reproduce freely in the cytoplasm by binary fission. After several divisions, amastigotes transform into trypomastigote forms, which are eventually released into the extracellular milieu and reach the bloodstream, where they can be taken up by the infect host cells or an insect vector, thus completing the natural life cycle [5].

The treatment of Chagas disease is currently restricted to two drugs, nifurtimox and benznidazole, which have limited efficacy and cause severe side effects [6,7]. In addition, there is no human vaccine against *T. cruzi* [8,9]. Therefore, there is a critical need to develop new therapeutic strategies for preventing or treating Chagas disease.

In line with the current interest in protein kinases as molecular targets for the treatment of a variety of diseases [10–13], the possibility that kinase inhibitors represent novel anti-parasitic agents is currently being explored [14]. 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 [15]. 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 [15,16]. Of these kinases, approximately 12% are unique to trypanosomatids [15,16]. 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 [17]. Taken all 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 [18]. Another proteomic study in *L. donovani* was able to identify 18 phosphosites from 16 distinct phosphopeptides [19]. However, to our knowledge the only phosphorylation site thus far determined in a *T. cruzi* protein is serine phosphorylation of the linker histone H1 [20]. In order to gain insight into the signaling networks that govern parasite function, we analyzed the phosphoproteome of *T. cruzi* at the non-infective, insect-derived epimastigote stage. We have chosen to initially perform the analysis of this stage, instead of the three other stages (i.e., metacyclic trypomastigote, amastigote, and trypomastigote), to better standardize the *T. cruzi* phosphoproteomic methodology, using a parasite form that can be obtained in large amounts in cell-free medium. 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 tandem 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.

2 Materials and Methods

2.1 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 [21]. 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 light. Protein content was measured by the Micro BCA assay (Pierce), according to the manufacturer's protocol.

2.2 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% trifluoroacetic acid (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).

2.3 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 μl 0.05% TFA, desalted in POROS R2 50 ziptips as described by Jurado *et al*. [22], and dried again.

2.4 Immobilized Metal-Affinity Chromatography (IMAC)

Fifty microliters of IMAC resin (PHOS Select Iron Affinity Gel, Sigma-Aldrich) was washed 2X with 450 μl 0.25 M acetic acid/30% ACN (IMAC buffer). Each fraction from SCX chromatography was redissolved in 100 μl 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 μl IMAC buffer and eluted 2X with 100 μl 5% TFA/45% ACN. Samples were dried in a vacuum centrifuge prior to LC-MS analysis.

2.5 Liquid Chromatography-Mass Spectrometry (LC-MS)

Following IMAC purification each fraction was redissolved in 30 μl 0.05% TFA and 8 μl were used for each injection. The peptides were loaded in a C18 trap column (0.25 μl, Opti-Pak, Optimized Technologies) coupled to a nanoHPLC system (NanoLC-1DPlus, Eksigent). The separation was carried in a capillary reverse phase (RP) column (Acclaim, 3 μm C18, 75 μm × 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 collision-induced dissociation (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 MS^2 scan event.

2.6 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*, bovine, human keratin, and porcine trypsin sequences (total 191,762 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 ≥ 0.05, peptide probability ≤ 0.1, and Xcorr of 1.5, 2.2, and 2.7 (2.8 only for MS/MS/MS data) 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 all possible modified amino acid residues were 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/) [23]. All valid phosphoproteins were submitted to Gene Ontology (GO), Blast (*e*-value ≤ 1e-5) and InterPro annotations using Blast2go algorithm (February 11th, 2009, http://www.blast2go.de/) [24].

2.7 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 $Na_4P_2O_7$, 50 mM NaF, 1 mM Na_3VO_4, 1% Triton X-100) containing 1 mM phenylmethylsulfonyl fluoride, 5 μg/ml aprotinin, 2 μg/ml leupeptin, 1 μg/ml pepstatin A, and clarified by centrifugation (16,000 × 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 with protein A-Sepharose beads (Rockland Immunochemicals), for 1 h at 4°C. 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 enhancedchemiluminescence reagent (ECL) and X-ray film (Phenix Research Products).

3 Results and Discussion

3.1 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 fragmentation of phosphopeptides results in a strong neutral loss of the phosphate group [25]. 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 [26] and MSA (also known as pseudo-MS^3) [27] 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 1). However, the MSA approach led to the identification of more phosphopeptides (n=160) as compared to the MS/MS (n=51) or MS/MS/MS (n=84) approach (Table 1,Supplementary Figures 1–3). In addition, only 17 peptides were sequenced using all three MSn acquisition regimens (Fig. 1), indicating that these regimens are complementary. It is noteworthy that the MS/MS/MS regimen led to an increased false-discovery rate (FDR), thus the Xcorr threshold had to be increased to validate the sequences. Two recent reports have compared the use of MS/MS, MS/MS/MS, and MSA for the analysis of phosphopeptides using high resolution mass spectrometers (Orbitrap-MS or FT-ICR-MS) for the first MS stage [28,29]. On the one hand, Villen *et al.* showed no significant differences in the number of phosphopeptides identified and performance to locate the modification sites using all three methods of fragmentation [29]. On the other hand, Ulintz *et al.* reported that the MSA method outperformed the two other strategies [28]. The results shown herein indicate that indeed MSA had a superior performance compared to the other MSn strategies. These results suggest that the MSA-provided higher quality spectra [27] are more crucial for the analysis of phosphopeptides when lower resolution data is recorded in the first MS stage.

In our analyses, 237 phosphopeptides were sequenced from 221 phosphorylation sites in 119 proteins from *T. cruzi* (Table 1, Supplementary Table 1). Of those 221 newly identified phosphorylation sites, 148 (65.5%) were on serine, 57 (25.2%) on threonine, and 8 (3.5%) on tyrosine (Table 1). Eight phosphorylation sites could not be assigned to specific amino acid residues, 7 on serine or threonine and 1 on serine or tyrosine (Table 1). From the identified phosphoproteins, approximately 40% had multiple phosphorylation sites (Table 1, Supplementary Table 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.

3.2 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 potential 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 (13.9%), CK1 (11.1%), PKA (8.71%), CK2 (7.7%), GSK3 (7.0%), ERK (5.9%), and CDK1 (4.2%) (Table 2). Fifty-seven (19.9%) 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 [15,16]. As expected, there was a significant amount of redundancy among the identified motifs; therefore, more motifs (n=287) were present compared to phosphorylation sites (n=221). 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 2). A key difference between host and parasite kinomes is the absence of receptor-linked and cytoplasmic tyrosine kinases in trypanosomatids [15]; however, the presence of protein-tyrosine phosphatases indicates that tyrosine phosphorylation is a key regulatory mechanism in *T. cruzi* [17]. 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 [15].

To determine whether any motif was over- or under-represented, the distribution of motifs of epimastigote phosphoproteome was compared to that of the entire *T. cruzi* database. The CAMK2, ERK, and CDK1 motifs were significantly over-represented ($p < 0.01$ by Fisher's exact test), whereas the CK1, NEK6, PLK1, and ALK motifs were under-represented ($p < 0.01$) (Table 2).

3.3 Functional categorization of identified phosphoproteins

Of 119 sequenced phosphoproteins, 68 (57%) were annotated in the GenBank database as hypothetical proteins, making it difficult to infer their biological function (Table 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 [30], new entries in the GenBank database could help in the annotation of the identified sequences. Indeed, Blast analysis resulted in the annotation of 21 sequences previously described as hypothetical proteins (Table 3, Supplementary Table 2). In addition, InterPro analysis resulted in 74 entries, 25 of which were previously annotated as hypothetical proteins (Table 3, Supplementary Table 3). Finally, GO analysis generated 69 annotations, 22 of which were previously annotated as hypothetical proteins (Table 3, Supplementary Table 4). Taken together, Blast, InterPro and GO analyses, assigned 35 (51%) of the 68 hypothetical sequences to a predicted biological function (Table 3). The overall function of the identified phosphoproteins is discussed in more detail in the following subsections.

3.3.1 Gene ontology analysis—The overall functional distribution of the identified phosphoproteins, as determined by GO analysis (Supplementary Tables 4–5), is shown in Figure 2. 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. 2, Supplementary Tables 4–5). The most abundant cellular components were intracellular, membrane bound, and localized in organelles (Fig. 2, Supplementary Tables 4–5). 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. 2, Supplementary Tables 4–5). Overall, GO analysis further illustrates that protein phosphorylation is a significant regulatory mechanism governing a variety of *T. cruzi* biological functions.

3.3.2 Signaling transduction-related proteins—Signal transduction pathways are highly dynamic protein networks that integrate information from various stimuli. Sixteen of the identified proteins are proposed to be cell signaling modulators. Of these, six (37.5%) are classified as either protein kinases or phosphatases by Blast, InterPro, or GO analysis (Table 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, 13.9% of the phosphorylation motifs were characterized as CAMK2 consensus sites, indicating Ca^{2+} might play 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 [31]. 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 [32]. WD40 repeats are also present in several proteins from *Chlamydomonas reinhardtii* intraflagellar transport system [31], 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.

3.3.3 Cytoskeleton, flagellum, and trafficking proteins—Fifteen of the identified phosphoproteins were annotated as cytoskeleton, flagellum, and trafficking proteins (Table 3). In mammalian sperm, hyperactivated motility is mediated by Ca^{2+} [33]; 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 [33–35]. We found that one paraflagellar rod protein (PAR1b) and one protein 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. In agreement with this data, tyrosine phosphorylation was recently shown to be enriched in *T. brucei* flagellum [36].

3.3.4 Transporters—We found 8 transporters in our analysis, 3 ABC, 1 acetate- and 4 ion-transporters. The ABC transporters are widely correlated with multidrug resistance from bacteria to humans [37]. The *T. cruzi* genome has 33 ABC transporter genes [38]. 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 [37]. Lara *et al.* proposed that the heme uptake by epimastigotes was dependent on ABC transporters [39]. 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.

3.3.5 DNA, RNA, and protein turnover—We have found two phosphoproteins with functions corresponding to nucleosome and chromatin assembly (i.e., histone H2B, nucleosome assembly protein). Marques Porto et al. showed by ^{32}P-labeling the phosphorylation of *T. cruzi* histone H1 and H2B [40]. 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 [20], 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 calpain indicates that phosphorylation may play a role in the regulation of these processes in *T. cruzi*.

3.3.6 Metabolism—Among the metabolic proteins, we identified enzymes involved in the late stages of the glycolytic pathway (i.e., pyruvate phosphate dikinase and pyruvate dehydrogenase E1 alpha subunit) (Table 3). Two phosphoproteins involved in nucleotide synthesis (i.e., ribose-phosphate pyrophosphokinase and phosphoribosylpyrophosphate synthetase) were also identified. These results suggest that these highly active metabolic pathways may be controlled at least in part by phosphorylation, thus feeding the machinery responsible for parasite cell growth and proliferation.

3.3.7 Pathogenesis—Although epimastigotes are noninfective forms, we found several proteins known to be related to pathogenesis of *T. cruzi*. For instance, five phosphoproteins belonging to two protein families (i.e., *trans*-sialidase (TS) and dispersed gene family (DGF)) were identified (Table 3). TS is a well known virulence factor [41] and comprises a protein superfamily (TS/gp85) encoded by about 1400 genes in *T. cruzi* genome [30]. 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 [42]. Other members have been related to host cell recognition and invasion to protect the parasite from the host complement system [43]. DGF protein 1 is a multigene family that was recently shown to be expressed on the surface of bloodstream trypomastigotes [44]. DGF protein 1 was also reported to be *N*-glycosylated [45]. 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 3).

3.3.8 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 [46,47]. The finding that several members of RHS family are phosphorylated suggests that these gene products may be highly active, which accelerates gene duplication and evolution [46,47]. This might explain the high divergence and differences in virulence between *T. cruzi* strains and phylogenetic lineages [47,48].

3.4 Analysis of phosphotyrosine proteome

It has previously been estimated that 30% of all eukaryotic proteins contain covalently bound phosphate at any given time [49]. Furthermore, the proportion of pSer, pThr, and pTyr was recently reported in HeLa cells to be approximately 86.4%, 11.8%, and 1.8%, respectively [50]. 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 [51,52]. Das et al. showed that the major *T. brucei* tyrosine-phosphorylated protein is a nuclear RNA-binding protein (Nopp44/46) [51]. Recently, Nett et al. described the tyrosine phosphoproteome of *T. brucei* procyclic forms, where they mapped the tyrosine-phosphorylation sites of 34 proteins. Using anti-phosphotyrosine antibodies and immunofluorescence microscopy these authors showed that the tyrosine phosphorylated proteins were mostly localized at the basal body, flagellum, and nucleolus, indicating that tyrosine phosphorylation may play a central role in guiding signaling molecules to specific parasite locations. 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 [52].

As described above, our MS/MS analysis identified 8 tyrosine phosphorylation sites from 6 distinct proteins: protein kinase (glycogen synthase kinase 3, GSK3); tyrosine aminotransferase; mitochondrial DNA polymerase I protein D; and three hypothetical proteins (EAN84292.1, EAN94368.1, and EAN95748.1) (Table 1 and Supplementary Table 1). The Figure 3A illustrates the identification of the tyrosine phosphorylated peptide LSPSEPNVAYpICSR from protein kinase GSK3. Initially, we observed an intense fragment corresponding to the neutral loss of phosphoric acid at *m/z* 788.38 ($[M - H_3PO_4 + 2H]^{2+}$). After detailed analysis of remainder MS/MS fragments, we unambiguously mapped the phosphorylation to the Tyr10. Phosphorylation of Ser2, Ser4, or Ser13 was discarded based on the identification of fragment series $b_{3–5}$ and $b_{8–9}$ (for Ser2 and Ser4) and $y_{2–4}$ (for Ser13). Furthermore, the presence of fragment series $y_{5–10}$ corroborates the phosphorylation on Tyr10. Contrary to what has been reported for triple-quadrupole [53] and Q-TOF [54] mass

spectrometers, intense neutral loss of the phosphate group from tyrosine phosphopeptides was previously observed by positive-ion mode MS/MS fragmentation in ion-trap instrument [25].

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 masses (Fig. 3B). Interestingly, except for the p250, the predicted molecular masses of the tyrosine phosphorylated proteins identified by MS/MS seemed to closely correspond to those of the phosphoproteins detected by Western blotting: protein kinase (GSK3) (40.4 kDa), tyrosine aminotransferase (46.1 kDa), mitochondrial DNA polymerase I protein D (26.1 kDa), EAN84292.1 (103 kDa), EAN94368.1 (61.3 kDa), and EAN95748.1 (60 kDa). However, since other PTMs could alter the relative mobility of proteins on SDS-PAGE, the identity of the phosphorylated proteins observed on Western blot could not be inferred. Also, the low abundance of these proteins on the SDS-PAGE (as assessed by Coomassie blue staining) precluded their identification by LC-MS/MS.

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) (Supplementary Fig. 4). The amino acid sequence alignment also shows that *T. cruzi* GSK3 Y187 (arrow) is conserved in each of the included species (Supplementary Fig. 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 [55, 56]. The significant proportion of GSK3 consensus phosphorylation sites (Table 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 [57, 58].

4 Concluding remarks

In this study, we have successfully mapped 237 phosphopeptides from 119 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, and survival. 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by grants from the National Institutes of Health (2S06GM008012-37 and 5G12RR008124). E.S.N. was partially supported by the George A. Krutilek memorial graduate scholarship from the Graduate School, UTEP. M.R.G. was supported by the Research Experience for Undergraduates (REU) Program/UTEP (National Science Foundation grant # DBI-0353887) and the University of Texas System Louis Stokes Alliance for Minority Participation (LSAMP) (NSF grant # HRD-0832951). T.J.P.S was supported by CAPES (Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior, Brazil). 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.

Abbreviations

BSA
bovine serum albumin

CID
collision-induced dissociation

ECL
enhanced chemiluminescent reagent

EDTA
ethylenediaminetetraacetic acid

FA
formic acid

FCaBP
flagellar calcium-binding protein

GO
gene ontology

IMAC
immobilized metal-affinity chromatography

LC-MS
liquid chromatography-mass spectrometry

MS/MS
tandem mass spectrometry

MS/MS/MS
dual-stage fragmentation

MSA
multistage activation

PAR
paraflagellar rod protein

PBS
phosphate-buffered saline solution

PVDF
polyvinylidene fluoride

PTM
post-translational modification

RP
reverse phase

SCX
strong-cation exchange

SDS-PAGE

sodium dodecyl sulfate-polyacrylamide gel electrophoresis

TFA

trifluoroacetic acid

Th

Thompson

References

1. Barrett MP, Burchmore RJ, Stich A, Lazzari JO, et al. The trypanosomiases. Lancet 2003;362:1469–1480. [PubMed: 14602444]
2. Dias JC, Silveira AC, Schofield CJ. The impact of Chagas disease control in Latin America: a review. Mem Inst Oswaldo Cruz 2002;97:603–612. [PubMed: 12219120]
3. Bern C, Montgomery SP, Herwaldt BL, Rassi A Jr, et al. Evaluation and treatment of Chagas disease in the United States: a systematic review. JAMA 2007;298:2171–2181. [PubMed: 18000201]
4. Piron M, Verges M, Munoz J, Casamitjana N, et al. Seroprevalence of Trypanosoma cruzi infection in at-risk blood donors in Catalonia (Spain). Transfusion 2008;48:1862–1868. [PubMed: 18522707]
5. Tyler KM, Engman DM. The life cycle of Trypanosoma cruzi revisited. Int J Parasitol 2001;31:472–481. [PubMed: 11334932]
6. Urbina JA, Docampo R. Specific chemotherapy of Chagas disease: controversies and advances. Trends Parasitol 2003;19:495–501. [PubMed: 14580960]
7. Wilkinson SR, Taylor MC, Horn D, Kelly JM, Cheeseman I. A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. Proc Natl Acad Sci U S A 2008;105:5022–5027. [PubMed: 18367671]
8. Dumonteil E. DNA Vaccines against Protozoan Parasites: Advances and Challenges. J Biomed Biotechnol 2007;2007:90520. [PubMed: 17710244]
9. Garg N, Bhatia V. Current status and future prospects for a vaccine against American trypanosomiasis. Expert Rev Vaccines 2005;4:867–880. [PubMed: 16372882]
10. Arslan MA, Kutuk O, Basaga H. Protein kinases as drug targets in cancer. Curr Cancer Drug Targets 2006;6:623–634. [PubMed: 17100568]
11. Cohen P, Goedert M. GSK3 inhibitors: development and therapeutic potential. Nat Rev Drug Discov 2004;3:479–487. [PubMed: 15173837]
12. Force T, Kuida K, Namchuk M, Parang K, Kyriakis JM. Inhibitors of protein kinase signaling pathways: emerging therapies for cardiovascular disease. Circulation 2004;109:1196–1205. [PubMed: 15023894]
13. Kumar S, Boehm J, Lee JC. p38 MAP kinases: key signalling molecules as therapeutic targets for inflammatory diseases. Nat Rev Drug Discov 2003;2:717–726. [PubMed: 12951578]
14. Doerig C. Protein kinases as targets for anti-parasitic chemotherapy. Biochim Biophys Acta 2004;1697:155–168. [PubMed: 15023358]
15. Parsons M, Worthey EA, Ward PN, Mottram JC. Comparative analysis of the kinomes of three pathogenic trypanosomatids: Leishmania major, Trypanosoma brucei and Trypanosoma cruzi. BMC Genomics 2005;6:127. [PubMed: 16164760]
16. Naula C, Parsons M, Mottram JC. Protein kinases as drug targets in trypanosomes and Leishmania. Biochim Biophys Acta 2005;1754:151–159. [PubMed: 16198642]
17. Brenchley R, Tariq H, McElhinney H, Szoor B, et al. The TriTryp phosphatome: analysis of the protein phosphatase catalytic domains. BMC Genomics 2007;8:434. [PubMed: 18039372]
18. Morales MA, Watanabe R, Laurent C, Lenormand P, et al. Phosphoproteomic analysis of Leishmania donovani pro- and amastigote stages. Proteomics 2008;8:350–363. [PubMed: 18203260]
19. Rosenzweig D, Smith D, Myler PJ, Olafson RW, Zilberstein D. Post-translational modification of cellular proteins during Leishmania donovani differentiation. Proteomics 2008;8:1843–1850. [PubMed: 18398879]

20. da Cunha JP, Nakayasu ES, Elias MC, Pimenta DC, et al. Trypanosoma cruzi histone H1 is phosphorylated in a typical cyclin dependent kinase site accordingly to the cell cycle. Mol Biochem Parasitol 2005;140:75–86. [PubMed: 15694489]
21. Camargo EP. Growth and differentiation In *Trypanosoma cruzi*. I. Origin of metacyclic trypanosomes in liquid media. Rev Inst Med Trop Sao Paulo 1964;12:93–100. [PubMed: 14177814]
22. Jurado JD, Rael ED, Lieb CS, Nakayasu E, et al. Complement inactivating proteins and intraspecies venom variation in Crotalus oreganus helleri. Toxicon 2007;49:339–350. [PubMed: 17134729]
23. Gnad F, Ren S, Cox J, Olsen JV, et al. PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. Genome Biol 2007;8:R250. [PubMed: 18039369]
24. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 2008;36:3420–3435. [PubMed: 18445632]
25. DeGnore JP, Qin J. Fragmentation of phosphopeptides in an ion trap mass spectrometer. J Am Soc Mass Spectrom 1998;9:1175–1188. [PubMed: 9794085]
26. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A 2004;101:12130–12135. [PubMed: 15302935]
27. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ. A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. Anal Chem 2004;76:3590–3598. [PubMed: 15228329]
28. Ulintz PJ, Yocum AK, Bodenmiller B, Aebersold R, et al. Comparison of MS(2)-Only, MSA, and MS(2)/MS(3) Methodologies for Phosphopeptide Identification. J Proteome Res 2009;8:887–899. [PubMed: 19072539]
29. Villen J, Beausoleil SA, Gygi SP. Evaluation of the utility of neutral-loss-dependent MS3 strategies in large-scale phosphorylation analysis. Proteomics 2008;8:4444–4452. [PubMed: 18972524]
30. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, et al. The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science 2005;309:409–415. [PubMed: 16020725]
31. Cole DG. The intraflagellar transport machinery of Chlamydomonas reinhardtii. Traffic 2003;4:435–442. [PubMed: 12795688]
32. Yaffe MB, Elia AE. Phosphoserine/threonine-binding domains. Curr Opin Cell Biol 2001;13:131–138. [PubMed: 11248545]
33. Hill KL. Biology and mechanism of trypanosome cell motility. Eukaryot Cell 2003;2:200–208. [PubMed: 12684369]
34. Broadhead R, Dawe HR, Farr H, Griffiths S, et al. Flagellar motility is required for the viability of the bloodstream trypanosome. Nature 2006;440:224–227. [PubMed: 16525475]
35. Maldonado RA, Mirzoeva S, Godsel LM, Lukas TJ, et al. Identification of calcium binding sites in the trypanosome flagellar calcium-acyl switch protein. Mol Biochem Parasitol 1999;101:61–70. [PubMed: 10413043]
36. Nett IR, Davidson L, Lamont D, Ferguson MA. Trypanosoma brucei: Identification and specific localization of tyrosine phosphorylated proteins. Eukaryot Cell. 2009
37. Jones PM, George AM. Multidrug resistance in parasites: ABC transporters, P-glycoproteins and molecular modelling. Int J Parasitol 2005;35:555–566. [PubMed: 15826647]
38. Aguero F, Zheng W, Weatherly DB, Mendes P, Kissinger JC. TcruziDB: an integrated, post-genomics community resource for Trypanosoma cruzi. Nucleic Acids Res 2006;34:D428–431. [PubMed: 16381904]
39. Lara FA, Sant'anna C, Lemos D, Laranja GA, et al. Heme requirement and intracellular trafficking in Trypanosoma cruzi epimastigotes. Biochem Biophys Res Commun 2007;355:16–22. [PubMed: 17292866]
40. Marques Porto R, Amino R, Elias MC, Faria M, Schenkman S. Histone H1 is phosphorylated in non-replicating and infective forms of Trypanosoma cruzi. Mol Biochem Parasitol 2002;119:265–271. [PubMed: 11814578]
41. Risso MG, Garbarino GB, Mocetti E, Campetella O, et al. Differential expression of a virulence factor, the trans-sialidase, by the main Trypanosoma cruzi phylogenetic lineages. J Infect Dis 2004;189:2250–2259. [PubMed: 15181573]

42. Pereira-Chioccola VL, Acosta-Serrano A, Correia de Almeida I, Ferguson MA, et al. Mucin-like molecules form a negatively charged coat that protects Trypanosoma cruzi trypomastigotes from killing by human anti-alpha-galactosyl antibodies. J Cell Sci 2000;113(Pt 7):1299–1307. [PubMed: 10704380]
43. Acosta-Serrano, A.; Hutchinson, C.; Nakayasu, ES.; Almeida, IC.; Carrington, M. Trypanosomes: After the genome. Barry, JD.; Mottram, JC.; McCulloch, R.; Acosta-Serrano, A., editors. Horizon Scientific Press; Norwich, UK: 2007. p. 319-337.
44. Kawashita SY, da Silva CV, Mortara RA, Burleigh BA, Briones MRS. Homology, paralogy and function of DGF-1, a hihgly dispersed *Trypanosoma cruzi* specific gene family and its implications for information entropy if its encoded proteins. Mol Biochem Parasitol. 2009Accepted
45. Atwood JA 3rd, Minning T, Ludolf F, Nuccio A, et al. Glycoproteomics of Trypanosoma cruzi trypomastigotes using subcellular fractionation, lectin affinity, and stable isotope labeling. J Proteome Res 2006;5:3376–3384. [PubMed: 17137339]
46. Bringaud F, Biteau N, Melville SE, Hez S, et al. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of Trypanosoma brucei. Eukaryot Cell 2002;1:137–151. [PubMed: 12455980]
47. Kim D, Chiurillo MA, El-Sayed N, Jones K, et al. 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 2005;346:153–161. [PubMed: 15716016]
48. Buscaglia CA, Di Noia JM. Trypanosoma cruzi clonal diversity and the epidemiology of Chagas' disease. Microbes Infect 2003;5:419–427. [PubMed: 12737998]
49. Venter JC, Adams MD, Myers EW, Li PW, et al. The sequence of the human genome. Science 2001;291:1304–1351. [PubMed: 11181995]
50. Olsen JV, Blagoev B, Gnad F, Macek B, et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 2006;127:635–648. [PubMed: 17081983]
51. Das A, Peterson GC, Kanner SB, Frevert U, Parsons M. A major tyrosine-phosphorylated protein of Trypanosoma brucei is a nucleolar RNA-binding protein. J Biol Chem 1996;271:15675–15681. [PubMed: 8663171]
52. Favoreto S Jr, Dorta ML, Yoshida N. Trypanosoma cruzi 175-kDa protein tyrosine phosphorylation is associated with host cell invasion. Exp Parasitol 1998;89:188–194. [PubMed: 9635442]
53. Tholey A, Reed J, Lehmann WD. Electrospray tandem mass spectrometric studies of phosphopeptides and phosphopeptide analogues. J Mass Spectrom 1999;34:117–123. [PubMed: 12440389]
54. Steen H, Kuster B, Fernandez M, Pandey A, Mann M. Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. Anal Chem 2001;73:1440–1448. [PubMed: 11321292]
55. Hughes K, Nikolakaki E, Plyte SE, Totty NF, Woodgett JR. Modulation of the glycogen synthase kinase-3 family by tyrosine phosphorylation. Embo J 1993;12:803–808. [PubMed: 8382613]
56. Wang QM, Fiol CJ, DePaoli-Roach AA, Roach PJ. Glycogen synthase kinase-3 beta is a dual specificity kinase differentially regulated by tyrosine and serine/threonine phosphorylation. J Biol Chem 1994;269:14566–14574. [PubMed: 7514173]
57. Droucheau E, Primot A, Thomas V, Mattei D, et al. Plasmodium falciparum glycogen synthase kinase-3: molecular model, expression, intracellular localisation and selective inhibitors. Biochim Biophys Acta 2004;1697:181–196. [PubMed: 15023360]
58. Knockaert M, Wieking K, Schmitt S, Leost M, et al. Intracellular Targets of Paullones. Identification following affinity purification on immobilized inhibitor. J Biol Chem 2002;277:25493–25501. [PubMed: 11964410]

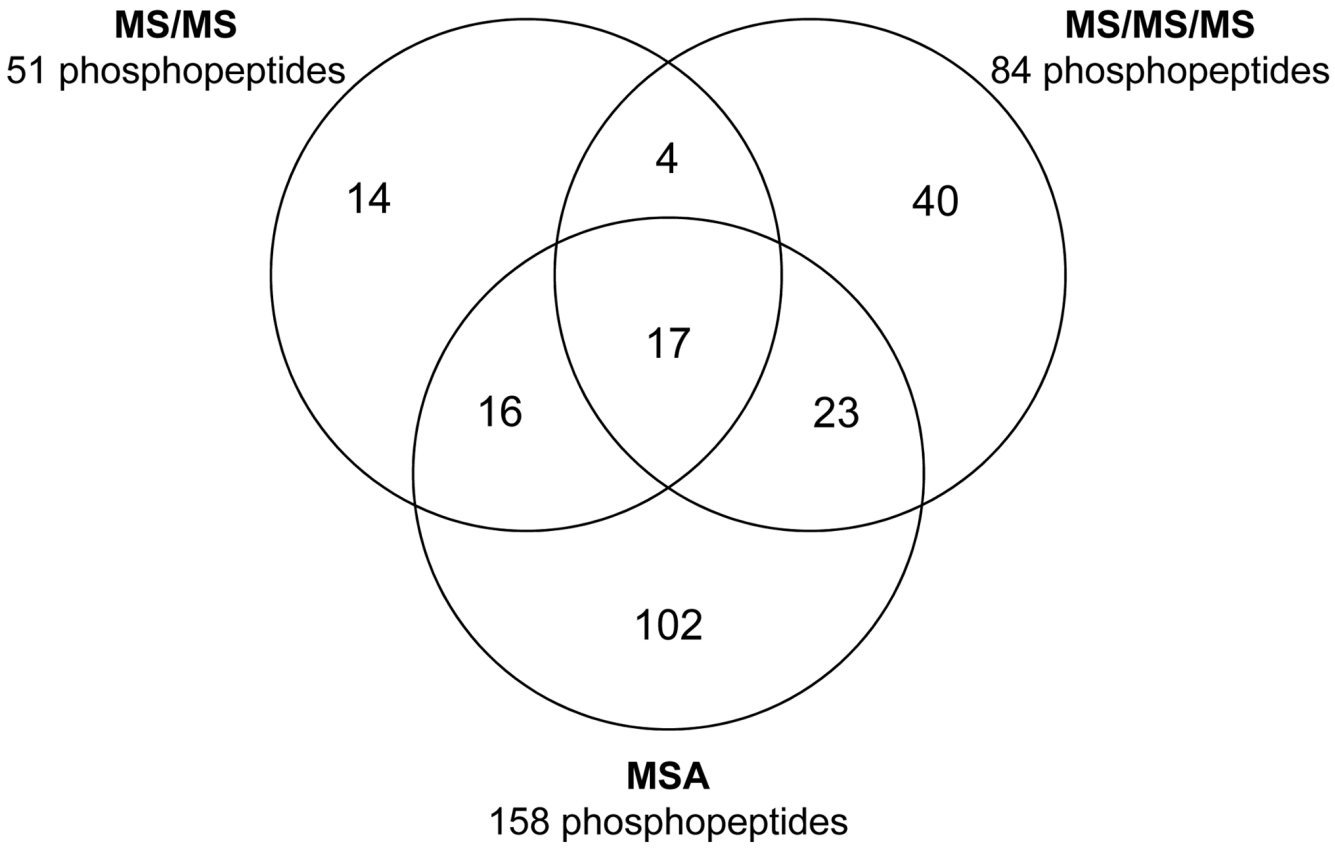


Figure 1.
Venn diagram of *T. cruzi* phosphopeptides identified by MS/MS, MS/MS/MS, and MSA analyses.

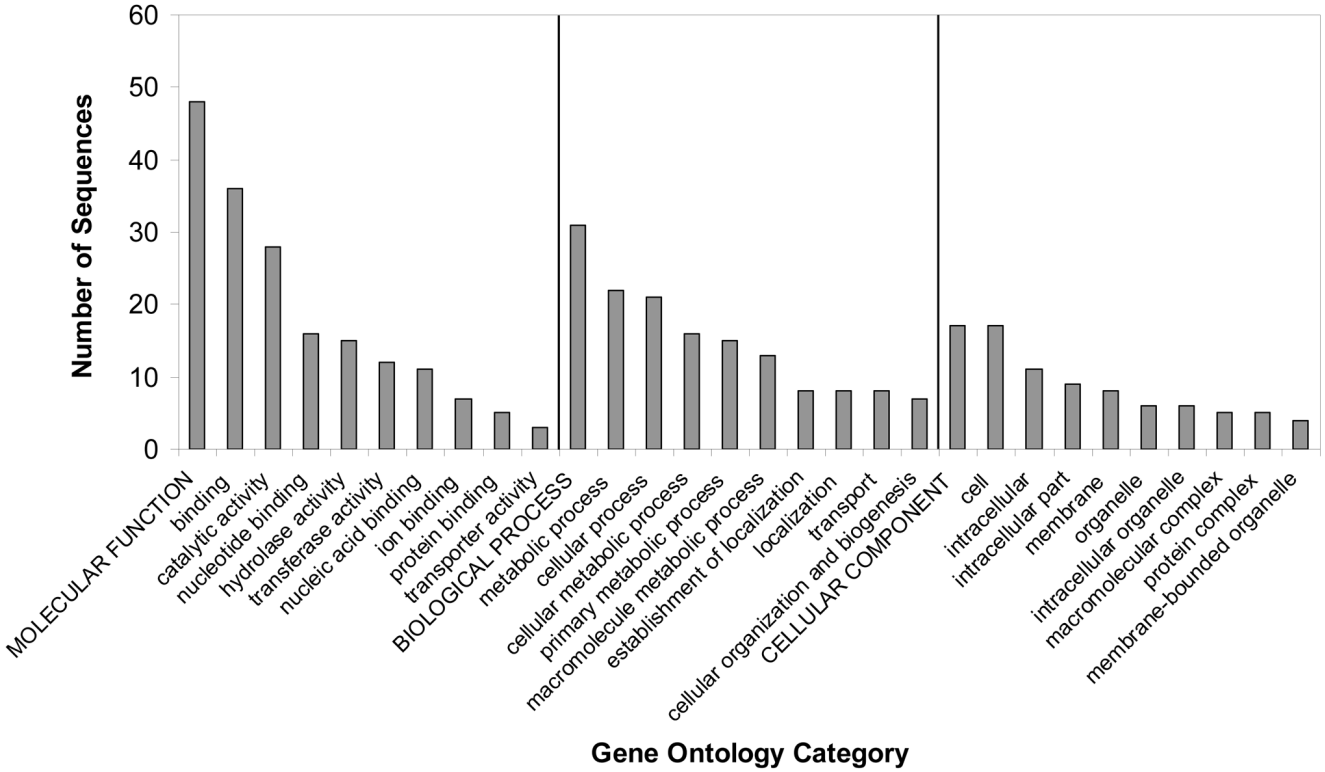
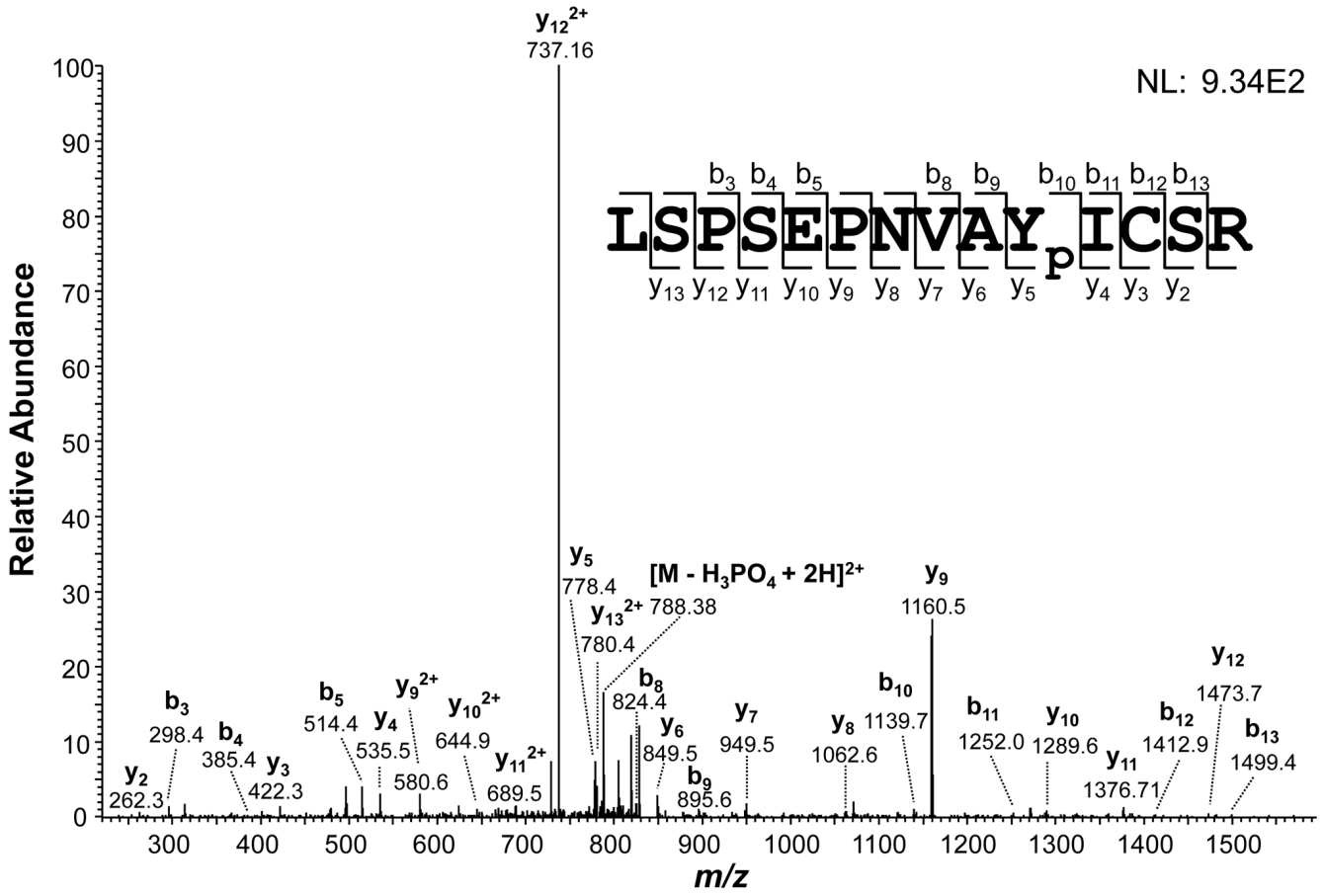


Figure 2.

Gene ontology analysis. GO analysis was performed using Blast2GO algorithm. The representative categories for molecular function, cellular component, and biological process are shown in the graph. For the complete list, see Supplementary Tables 4 and 5.



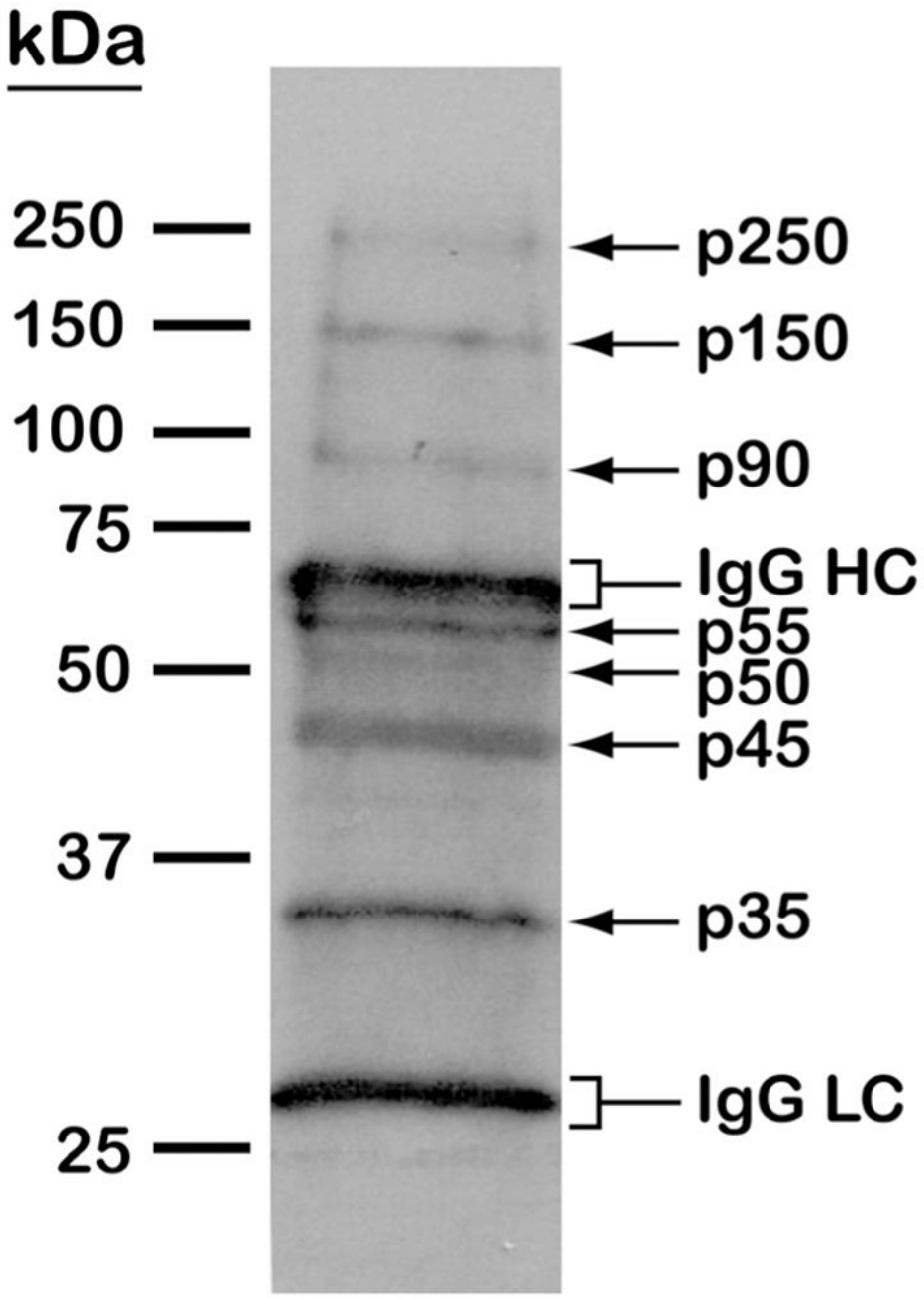


Figure 3.

(A) MS/MS spectrum of the tyrosine-phosphorylated peptide LSPSEPNVAYpICSR from glycogen synthase kinase 3 (GSK3). Matched b and y fragments are indicated. **(B)** 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 chemiluminescent reagent. Arrows denote *T. cruzi* tyrosine phosphorylated proteins. Brackets denote immunoglobulin G heavy (IgG HC) and light chains (IgG LC).

Table 1

Statistics of phosphoproteome analysis of Trypanosoma cruzi epimastigote forms.

		Phosphoproteins					
MS approach	**Non-phosphorylated proteins**	**1P**	**2P**	**3P**	**4P**	**>4P**	**Total phosphoproteins**
MS/MS	111	25	8	4	0	0	37
MS/MS/MS	110	46	14	3	1	2	65
MSA	80	60	22	11	5	1	99
Total	156	74	30	18	5	4	131
T. cruzi	139	72	24	16	3	4	119

			Phosphopeptides				
	Non-phosphorylated peptides	Phosphorylation sites	1P	2P	>2P	Total phosphopeptides	Reverse	FDR (%)
MS/MS	146	52	47	3	1	51	5	2.0
MS/MS/MS	124	65	75	5	4	84	9	4.1
MSA	96	162	146	9	5	160	9	3.4
Total	206	226	231	11	8	250	21	4.6
T. cruzi	196	221	227	5	5	237	-	-

	Distribution of *T. cruzi* phosphorylated sites				
	Phosphoserine	Phosphothreonine	Phosphotyrosine	Ser or Thr	Ser or Tyr
# of sites	148	57	8	7	1
% of Total	65.5	25.2	3.5	3.1	0.4

Table 2

Distribution of kinase specific motifs in *T. cruzi*.

	Phosphoproteome		All *T. cruzi* Sequences	
Kinase motif	**Number**	**% of total**	**Number**	**% of total**
CAMK2 (R-X-X-**S/T**)	40	13.94*	170873	8.53
CK1 (S-X-X-**S/T,** S/T-X-X-X-**S**)	32	11.15*	450808	22.49
PKA (R-X-**S/T**, R-R/K-X-**S/T**)	25	8.71	193369	9.65
CK2 (**S/T**-X-X-E)	22	7.67	183303	9.15
GSK3 (**S**-X-X-X-S)	20	6.97	134201	6.7
ERK (V-X-**S/T**-P, P-X-**S/T**-P)	17	5.92*	23114	1.15
CDK1 (**S/T**-P-K/R, **S/T**-P-X-K/R)	12	4.18*	30468	1.52
NEK6 (L-X-X-**S/T**)	10	3.48*	250587	12.5
CHK1 (M/I/L/V-X-R/K-X-X-**S/T**)	10	3.48	81145	4.05
PKD (L/V/I-X-R/K-X-X-**S/T**)	9	3.14	57644	2.88
Aurora (R/K-X-**S/T**-I/L/V)	8	2.79	57021	2.85
CDK2 (**S/T**-P-X-K/R)	6	2.09	16244	0.81
AKT (R-R/S/T-X-**S/T**-X-S/T, R-X-R-X-X-**S/T**)	5	1.74	20994	1.05
Aurora-A (R/K/N-R-X-**S/T**-M/L/V/I)	4	1.39	6121	0.31
EGFR (D/P/S/A/E/N-X-**Y**-V/L/D/E/I/N/P)	3	1.05	69975	3.49
ALK (**Y**-X-X-I/L/V/M)	3	1.05*	112032	5.59
PLK1 (E/D-X-**S/T**-F/L/I/Y/W/V/M)	2	0.70*	91008	4.54
SRC (E/D-X-X-**Y**-X-X-D/E/A/G/S/T)	1	0.35	18509	0.92
CHK1/2 (L-X-R-X-X-**S/T**)	1	0.35	16716	0.83
Unknown motif	57	19.86	-	-
Other motifs	-	-	19945	1
Total	287	100	2004077	100

* $p < 0.01$ by Fisher's exact test.

Table 3

Functional categorization of identified phosphoproteins of Trypanosoma cruzi epimastigotes.

Hit #	Protein description	Accession #	P sites	Blast Analysis: Sequence description	Blast Analysis: min. e-value	InterPro	Gene Ontology
	Signal transduction						
1	regulatory subunit of protein kinase a-like protein	EAN92816.1	5	regulatory subunit of protein kinase a-like protein	1.0E-0.0	Cyclic nucleotide-binding, RmlC-like jelly roll fold	cAMP-dependent protein kinase complex
2	phosphoprotein phosphatase	EAN90617.1	2	serine threonine protein	1.0E-0.0	Serine/threonine-specific protein phosphatase and bis(5-nucleosyl)-tetraphosphatase	hydrolase activity
3	rac serine-threonine kinase	EAN95526.1	2	serine threonine protein kinase	1.0E-0.0	Zinc finger, FYVE-type	ATP binding, protein amino acid phosphorylation, protein serine/threonine kinase activity, zinc ion binding
4	hypothetical protein	EAN95573.1	2	WD repeat domain 17	1.0E-0.0	WD40/YVTN repeat-like	none
5	protein kinase-A catalytic subunit	AAL17691.2	1	protein kinase a catalytic subunit	1.0E-0.0	Serine/threonine protein kinase	protein serine/threonine kinase activity
6	hypothetical protein	EAN96688.1	1	hypothetical protein	1.0E-0.0	WD40 repeat-like	none
7	hypothetical protein	EAN97518.1	1	hypothetical protein	1.0E-0.0	WD40/YVTN repeat-like	none
8	hypothetical protein	EAN88571.1	1	WD-repeat proteinexpressed	1.0E-0.0	WD40/YVTN repeat-like	nucleotide binding
9	calmodulin	EAN86242.1	1	calmodulin	1.00E-80	Calcium-binding EF-hand	calcium ion binding
10	protein kinase (GSK3 beta)	EAN98751.1	1	protein kinase	1.0E-0.0	Serine/threonine protein kinase-related	ATP binding, protein amino acid phosphorylation, protein serine/threonine kinase activity
11	hypothetical protein	EAN99395.1	1	EF-hand family protein	1.0E-0.0	Calcium-binding EF-hand	calcium ion binding
12	hypothetical protein	EAN93781.1	2	hypothetical protein	1.0E-0.0	Forkhead-associated	none
13	hypothetical protein	EAN91302.1	2	chromosome 6 open reading frame 224	1.0E-0.0	Shikimate kinase, Adenylate kinase, YHS	shikimate kinase activity, ATP binding
14	hypothetical protein	EAN97116.1	1	calmodulin-related protein	1.00E-180	none	calcium ion binding
15	hypothetical protein	EAN83100.1	1	EF-hand domain	1.0E-0.0	Region of unknown function DM10, Protein of unknown function DUF1126	calcium ion binding, flagellum,
16	hypothetical protein	EAN83235.1	1	hypothetical protein	1.0E-0.0	C2 calcium-dependent membrane targeting	none
	Cytoskeleton, flagellum and traffic proteins						
17	paraflagellar rod component Par1b	AAC32103.1	3	69 kda paraflagellar rod protein	1.0E-0.0	Paraflagellar rod	calmodulin binding, microtubule-based flagellum
18	hypothetical protein	EAN99904.1	2	hypothetical protein	1.00E-134	Major sperm protein, PapD-like	structural molecule activity
19	Beta tubulin 2.3	AAL75957.1	1	beta-tubulin	1.0E-0.0	Tubulin/FtsZ, GTPase	structural molecule activity
20	kinesin	EAN86063.1	1	kinesin family member 14	1.0E-0.0	Kinesin, motor region	microtubule motor activity, ATP binding, microtubule associated complex, microtubule-based movement

				Blast Analysis			
Hit #	**Protein description**	**Accession #**	**P sites**	**Sequence description**	**min. e-value**	**InterPro**	**Gene Ontology**
21	hypothetical protein	EAN84227.1	1	dc2-related axonemal dynein intermediate chain	1.0E-0.0	none	none
22	dynein heavy chain	EAN97081.1	1	dynein heavy chain	1.0E-0.0	ATPase associated with various cellular activities, AAA-5	ATP binding, ATPase activity
23	clathrin coat assembly protein	EAN98227.1	1	clathrin coat assembly protein	1.0E-0.0	ENTH/VHS, Epsin-like, N-terminal	phosphatidylinositol binding, clathrin coat, clathrin binding, clathrin cage assembly
24	hypothetical protein	EAN96083.1	1	flagellar protein	1.00E-154	none	none
25	hypothetical protein	EAN88130.1	1	Tctex1 domain	1.00E-68	Tctex-1	flagellum, microtubule associated complex, microtubule-based process
26	hypothetical protein	EAN87422.1	1	neurofilament heavy polypeptide (NF-H)	1.0E-0.0	none	structural molecule activity
	Transporters						
27	ABC transporter	EAN96763.1	6	ABC transporter	1.0E-0.0	ABC-2 type transporter	ATP binding, ATPase activity
28	ABC transporter	EAN92058.1	3	ABC transporter	1.0E-0.0	ABC transporter-like	ATP binding, ATPase activity
29	ABC transporter	EAN89676.1	2	ABC transporter	1.0E-0.0	ABC-2 type transporter	ATP binding, ATPase activity
30	metal-ion transporter	EAN85275.1	1	zinc transporter-like protein	1.0E-0.0	Cation efflux protein	cation transport, cation transmembrane transporter activity, membrane
31	chloride channel protein	EAN94356.1	1	chloride channel 7	1.0E-0.0	Cystathionine beta-synthase, Chloride channel, voltage gated	voltage-gated chloride channel activity, chloride transport, membrane
32	calcium channel protein	EAN97848.1	1	calcium channel protein	1.0E-0.0	Ion transport	ion channel activity, ion transport, membrane
33	GPR1/FUN34/yaaH family	EAN82352.1	1	GPR1/FUN34/yaaH family (Acetate transporter)	3.44E-135	GPR1/FUN34/yaaH	membrane
34	Na+-ATPase	BAC98847.1	1	calcium motive p-type ATPase	1.0E-0.0	ATPase, P-type, K/Mg/Cd/Cu/Zn/Na/Ca/Na/H-transporter	ATP binding, membrane, cation transport, catalytic activity, metabolic process
	DNA, RNA and protein turnover						
35	GTP-binding elongation factor Tu family	EAN90873.1	4	GTP binding protein 1	1.0E-0.0	Translation elongation factor EF1A/initiation factor IF2gamma, C-terminal	GTPase activity, GTP binding
36	heterogeneous nuclear ribonucleoprotein H/F	EAN91678.1	4	heterogeneous nuclear ribonucleoprotein h	1.0E-0.0	RNA recognition motif, RNP-1	nucleic acid binding, nucleotide binding
37	mitochondrial DNA polymerase I protein D	EAN86491.1	3	mitochondrial dna polymerase i protein	1.00E-135	DNA polymerase A	DNA binding, DNA-directed DNA polymerase activity, DNA replication
38	hypothetical protein	EAN95162.1	3	hypothetical protein	1.0E-0.0	RNA recognition motif, RNP-1	nucleic acid binding
39	eukaryotic translation initiation factor	EAN92887.1	2	eukaryotic translation initiation factor	1.0E-0.0	Eukaryotic translation initiation factor 4E (eIF-4E)	RNA binding, translation initiation factor activity, cytoplasm, translational initiation

				Blast Analysis			
Hit #	**Protein description**	**Accession #**	**P sites**	**Sequence description**	**min. e-value**	**InterPro**	**Gene Ontology**
40	25 kDa translation elongation factor 1-beta	EAN82455.1	1	translation elongation factor 1-beta	2.16E-79	Translation elongation factor EF1B, beta and delta chains, guanine nucleotide	translation elongation factor activity
41	Tcc1a22.3	AAL82703.1	1	elongation factor 1-gamma	1.0E-0.0	Glutathione S-transferase, Thioredoxin fold	translation elongation, protein biosynthesis,
42	hypothetical protein	EAN84292.1	2	hypothetical protein	1.0E-0.0	Signal recognition particle, SRP54 subunit, GTPase	GTP binding, SRP-dependent cotranslational protein targeting to membrane, 7S RNA binding, membrane
43	histone H2B	EAN83534.1	1	histone H2B	4.12E-58	Histone H2B	nucleosome, DNA binding, nucleus, nucleosome assembly
44	hypothetical protein	EAN86654.1	1	PAB1 binding protein	1.0E-0.0	Ataxin-2, N-terminal	mitochondrion, mRNA polyadenylation, cytoplasm, positive regulation of protein biosynthesis, polysome, nucleus
45	mitochondrial DNA polymerase I protein C	EAN88565.1	1	mitochondrial dna polymerase i protein	1.0E-0.0	DNA-directed DNA polymerase, family A	DNA binding, DNA-directed DNA polymerase activity, DNA replication
46	ATP-dependent RNA helicase	EAN96171.1	1	ATP-dependent rna helicase	1.0E-0.0	DNA/RNA helicase, DEAD/DEAH box type, N-terminal	nucleic acid binding, ATP binding, ATP-dependent helicase activity
47	tyrosyl or methionyl-tRNA synthetase	EAN96489.1	1	tyrosyl-tRNA synthetase	1.0–128	Nucleic acid-binding, OB-fold	tRNA binding
48	glutamamyl carboxypeptidase	EAN85499.1	1	acetylornithine deacetylase	1.00E-152	Peptidase M20, dimerisation	proteolysis, metallopeptidase activity, protein dimerization activity
49	calpain cysteine peptidase	EAN83138.1	1	calpain-like cysteine peptidase	1.0E-0.0	Peptidase C2, calpain	calcium-dependent cysteine-type endopeptidase activity, intracellular, proteolysis
50	calpain-like cysteine peptidase	EAN88142.1	1	calpain-like cysteine peptidase	1.0E-0.0	Peptidase C2, calpain	calcium-dependent cysteine-type endopeptidase activity, intracellular, proteolysis
51	calpain-like cysteine peptidase	EAN92790.1	1	calpain-like cysteine peptidase	1.00E-139	Region of unknown function DUF1935	proteolysis, calcium-dependent cysteine-type endopeptidase activity, intracellular
52	hypothetical protein	EAN99674.1	2	hypothetical protein	1.0E-0.0	Zinc finger, CCCH-type	nucleic acid binding, zinc ion binding
53	hypothetical protein	EAN96152.1	3	cysteine clan family c2	1.0E-0.0	none	proteolysis, calcium-dependent cysteine-type endopeptidase activity, intracellular
54	nucleosome assembly protein	EAN95410.1	1	nucleosome assembly protein	1.00E-158	Nucleosome assembly protein (NAP)	nucleus, nucleosome assembly
	Metabolism						
55	Tyrosine Aminotransferase	1BW0	3	tyrosine aminotransferase	1.0E-0.0	1-aminocyclopropane-1-carboxylate synthase	tyrosine transaminase activity
56	ribose-phosphate pyrophosphokinase	EAN92422.1	3	ribose-phosphate pyrophosphokinase	1.0E-0.0	Phosphoribosyl pyrophosphokinase	nucleoside metabolic process, magnesium ion binding, ribose phosphate diphosphokinase activity, nucleotide biosynthetic process

				Blast Analysis			
Hit #	**Protein description**	**Accession #**	**P sites**	**Sequence description**	**min. e-value**	**InterPro**	**Gene Ontology**
57	pyruvate dehydrogenase E1 alpha subunit	AAD11551.1	2	pyruvate dehydrogenase e1 component alpha	1.0E-0.0	Pyruvate dehydrogenase (acetyl-transferring) E1 component, alpha subunit,	metabolic process
58	pyruvate phosphate dikinase 2	AAG12986.1	1	pyruvate phosphate dikinase	1.0E-0.0	Pyruvate/Phosphoenolpyruvate kinase, catalytic core	kinase activity
59	phosphoribosylpyrophosphate synthetase	EAN97393.1	1	phosphoribosylpyrophosphate synthetase	1.0E-0.0	Phosphoribosyl pyrophosphokinase	nucleoside metabolic process, magnesium ion binding, ribose phosphate diphosphokinase activity, nucleotide biosynthetic process
60	ribose-phosphate pyrophosphokinase	EAN95566.1	1	ribose-phosphate pyrophosphokinase	1.0E-0.0	Phosphoribosyl pyrophosphokinase	nucleoside metabolic process, magnesium ion binding, ribose phosphate diphosphokinase activity, nucleotide biosynthetic process
61	hypothetical protein	EAN95143.1	1	hypothetical protein	1.0E-0.0	Methyltransferase type 11	metabolic process, methyltransferase activity
62	hypothetical protein	EAN95748.1	1	had superfamily	1.0E-0.0	HAD-superfamily hydrolase, subfamily IIA, CECR5	metabolic process, hydrolase activity
	Pathogenesis						
63	trans-sialidase	EAN91292.1	3	trans-sialidase	1.0E-0.0	Concanavalin A-like lectin/glucanase, subgroup	exo-alpha-sialidase activity, pathogenesis
64	trans-sialidase	EAN83746.1	2	trans-sialidase	1.0E-0.0	Neuraminidase	sialidase activity, pathogenesis
65	trans-sialidase	EAN96241.1	2	trans-sialidase	1.0E-0.0	Concanavalin A-like lectin/glucanase, subgroup	exo-alpha-sialidase activity, pathogenesis
66	trans-sialidase	EAN99282.1	1	trans-sialidase	1.0E-0.0	Concanavalin A-like lectin/glucanase	exo-alpha-sialidase activity, pathogenesis
67	dispersed gene family protein 1 (DGF-1)	EAN86627.1	1	dispersed gene family protein 1 (DGF-1)	1.0E-0.0	Parallel beta-helix repeat, Pectin lyase fold/virulence factor	none
	Binding						
68	hypothetical protein	EAN97824.1	3	c11orf60 protein	1.0E-0.0	protein binding	cytoplasm, nucleus
69	hypothetical protein	EAO00106.1	3	hypothetical protein	1.0E-0.0	none	protein binding
70	hypothetical protein	EAN92411.1	3	p25-alpha domain-containing	1.00E-74	P25-alpha	none
71	hypothetical protein	EAN89957.1	3	hypothetical protein	1.0E-0.0	Leucine-rich repeat	protein binding
72	hypothetical protein	EAN85866.1	1	hypothetical protein	1.0E-0.0	Tetratricopeptide-like helical, Sel1-like, CS domain	binding
73	hypothetical protein	EAN88376.1	1	hypothetical protein	1.0E-0.0	C2 calcium-dependent membrane targeting	none
74	hypothetical protein	EAN96978.1	1	hypothetical protein	1.0E-0.0	none	nucleic acid binding, zinc ion binding
75	hypothetical protein	EAN92548.1	1	T-complex-associated testis expressed 1	1.0E-0.0	Leucine-rich repeat	protein binding
76	I/6 autoantigen	EAN91318.1	1	I/6 autoantigen	1.00E-83	calcium ion binding	structural constituent of cytoskeleton, cytoplasm, calcium ion binding, cytoskeleton, microtubule
77	I/6 autoantigen	EAN94637.1	1	I/6 autoantigen	1.00E-82	calcium ion binding	structural constituent of cytoskeleton, cytoplasm, calcium ion binding, cytoskeleton, microtubule

Hit #	Protein description	Accession #	P sites	Blast Analysis: Sequence description	Blast Analysis: min. e-value	InterPro	Gene Ontology
	Other functions						
78	retrotransposon hot spot (RHS) protein	EAN86249.1	2	retrotransposon hot spotprotein	1.0E-0.0	SMAD/FHA domain	none
79	retrotransposon hot spot (RHS) protein	EAN99306.1	2	retrotransposon hot spotprotein	1.0E-0.0	SMAD/FHA domain	none
80	retrotransposon hot spot (RHS) protein	EAN94997.1	1	retrotransposon hot spotprotein	1.0E-0.0	SMAD/FHA domain	none
81	hypothetical protein	EAN88126.1	2	isoform c	1.0E-0.0	CCR4-Not complex component, Not1	muscle development, dendrite morphogenesis, sugar binding
82	hypothetical protein	EAN90373.1	2	Mucoid inhibitor A	1.0E-0.0	Conserved hypothetical protein CHP02231	none
83	hypothetical protein	EAN85046.1	1	hypothetical multipass transmembrane protein	1.0E-0.0	none	none
84	peroxin 14	EAN89658.1	1	peroxin 14	1.0E-0.0	Peroxisome membrane anchor protein Pex14p, N-terminal	peroxisome, membrane
85	hypothetical protein	EAN98546.1	1	cg17097-isoform b	1.0E-0.0	none	none
86	hypothetical protein	EAN93909.1	1	hypothetical multipass transmembrane protein	1.0E-0.0	none	none
	Proteins without known function or domain						
87	hypothetical protein	EAN90722.1	5	hypothetical protein	1.0E-0.0	none	none
88	hypothetical protein	EAN94368.1	5	hypothetical protein	1.0E-0.0	none	none
89	hypothetical protein	EAN88585.1	4	hypothetical protein	1.0E-0.0	none	none
90	hypothetical protein	EAN91358.1	3	hypothetical protein	1.00E-49	none	none
91	hypothetical protein	EAN91901.1	3	hypothetical protein	1.0E-0.0	none	none
92	hypothetical protein	EAN94097.1	3	hypothetical protein	1.00E-105	none	none
93	hypothetical protein	EAN94312.1	3	hypothetical protein	1.0E-0.0	none	none
94	hypothetical protein	EAN83060.1	2	hypothetical protein	2.83E-114	none	none
95	hypothetical protein	EAN85364.1	2	hypothetical protein	2.46E-121	none	none
96	hypothetical protein	EAN93581.1	2	hypothetical protein	1.00E-159	none	none
97	hypothetical protein	EAN95093.1	2	hypothetical protein	1.0E-0.0	none	none
98	hypothetical protein	EAN96967.1	2	hypothetical protein	1.00E-124	none	none
99	hypothetical protein	EAN99066.1	2	hypothetical protein	1.0E-0.0	none	none
100	hypothetical protein	EAN99070.1	2	hypothetical protein	1.00E-178	none	none
101	hypothetical protein	EAN86750.1	1	hypothetical protein	1.0E-0.0	none	none
102	hypothetical protein	EAN87778.1	1	hypothetical protein	1.0E-0.0	none	none
103	hypothetical protein	EAN88004.1	1	hypothetical protein	1.0E-0.0	none	none

				Blast Analysis			
Hit #	**Protein description**	**Accession #**	**P sites**	**Sequence description**	**min. e-value**	**InterPro**	**Gene Ontology**
104	hypothetical protein	EAN90456.1	1	hypothetical protein	1.0E-0.0	none	none
105	hypothetical protein	EAN93004.1	1	hypothetical protein	1.0E-0.0	none	none
106	hypothetical protein	EAN93058.1	1	hypothetical protein	1.00E-70	none	none
107	hypothetical protein	EAN82852.1	1	hypothetical protein	1.0E-0.0	none	none
108	hypothetical protein	EAN94518.1	1	hypothetical protein	1.0E-0.0	none	none
109	hypothetical protein	EAN94661.1	1	hypothetical protein	1.0E-0.0	none	none
110	hypothetical protein	EAN94862.1	1	hypothetical protein	1.0E-0.0	none	none
111	hypothetical protein	EAN98665.1	1	hypothetical protein	1.0E-0.0	none	none
112	hypothetical protein	EAN99075.1	1	hypothetical protein	1.0E-0.0	none	none
113	hypothetical protein	EAN99719.1	1	hypothetical protein	1.0E-0.0	none	none
114	hypothetical protein	EAN99770.1	1	hypothetical protein	1.0E-0.0	none	none
115	hypothetical protein	EAN99804.1	1	hypothetical protein	1.00E-176	none	none
116	hypothetical protein	EAN96262.1	1	hypothetical protein	1.60E-93	none	none
117	hypothetical protein	EAN80998.1	1	hypothetical protein	1.0E-0.0	none	none
118	hypothetical protein	EAN86037.1	1	hypothetical protein	1.00E-67	none	none
119	hypothetical protein	EAN91356.1	1	hypothetical protein	1.0E-0.0	none	none