

Article

The Experimental Proteome of *Leishmania infantum* Promastigote and Its Usefulness for Improving Gene Annotations

África Sanchiz, Esperanza Morato, Alberto Rastrojo, Esther Camacho, Sandra González-de la Fuente, Anabel Marina, Begoña Aguado and Jose M. Requena *

Centro de Biología Molecular “Severo Ochoa” (CBMSO, CSIC-UAM) Campus de Excelencia Internacional (CEI) UAM+CSIC, Universidad Autónoma de Madrid, 28049 Madrid, Spain; africa.sanchiz@gmail.com (Á.S.); emorato@cbm.csic.es (E.M.); arastrojo@cbm.csic.es (A.R.); ecamacho@cbm.csic.es (E.C.); sandra.g@cbm.csic.es (S.G.-d.l.F.); amarina@cbm.csic.es (A.M.); baguado@cbm.csic.es (B.A.)

* Correspondence: jmrequena@cbm.csic.es

Received: 27 July 2020; Accepted: 28 August 2020; Published: 2 September 2020

Abstract: *Leishmania infantum* causes visceral leishmaniasis (kala-azar), the most severe form of leishmaniasis, which is lethal if untreated. A few years ago, the re-sequencing and de novo assembling of the *L. infantum* (JPCM5 strain) genome was accomplished, and now we aimed to describe and characterize the experimental proteome of this species. In this work, we performed a proteomic analysis from axenic cultured promastigotes and carried out a detailed comparison with other *Leishmania* experimental proteomes published to date. We identified 2352 proteins based on a search of mass spectrometry data against a database built from the six-frame translated genome sequence of *L. infantum*. We detected many proteins belonging to organelles such as glycosomes, mitochondria, or flagellum, as well as many metabolic enzymes and many putative RNA binding proteins and molecular chaperones. Moreover, we listed some proteins presenting post-translational modifications, such as phosphorylations, acetylations, and methylations. On the other hand, the identification of peptides mapping to genomic regions previously annotated as non-coding allowed for the correction of annotations, leading to the N-terminal extension of protein sequences and the uncovering of eight novel protein-coding genes. The alliance of proteomics, genomics, and transcriptomics has resulted in a powerful combination for improving the annotation of the *L. infantum* reference genome.

Keywords: *Leishmania infantum*; proteome; post-translational modifications (PTMs); proteogenomics; mass spectrometry

1. Introduction

The genus *Leishmania* belongs to the order Trypanosomatida and includes protozoan parasites that are responsible for a complex of diseases named leishmaniasis, which is the second most common cause of mortality among tropical infectious diseases, after malaria [1]. Some species of *Leishmania* are human pathogens that cause different clinical manifestations as cutaneous (CL), mucocutaneous (MCL), or visceral (VL) leishmaniasis. Old World species *Leishmania infantum* and *Leishmania donovani* cause VL or kala-azar, which is often lethal if untreated, whereas *Leishmania major* causes CL and *Leishmania braziliensis* is associated with MCL. The VL-causative species are genetically almost identical, although they differ in geographic distribution: *L. donovani* is found in the Indian subcontinent and East Africa, while *L. infantum* is endemic in the countries around the Mediterranean basin, Latin America, and China [2]. Two stages, promastigote and amastigote, alternate in the *Leishmania* life cycle. Promastigotes are

flagellated and motile forms develop extracellularly in the gut of their sand-fly vector. The infection of the mammal host takes place during the sand-fly blood meal; afterwards, parasites are phagocytized by macrophages, and amastigote forms develop inside these host cells.

Within the *Leishmania* genus, the *L. major* genome was the first to be sequenced [3], followed by the *L. infantum* and *L. braziliensis* ones [4]. During the last decade, the extraordinary progress in DNA sequencing methodologies has allowed for the drafting of the genomes for many other *Leishmania* species [5–12] and for the improvement of the assemblies of the first sequenced genomes [13–15].

The availability of a complete and well-annotated genome provides the ultimate resource for genome-wide scale approaches, such as transcriptome and proteome analyses [16]. In parallel to the advances in sequencing technologies, proteomics methodologies are achieving unprecedented levels of sensitivity, and novel MS-based experimental approaches have become the method of choice for the analysis of complex protein mixtures such as cells, tissues, and even whole organisms. In particular, several proteomic technologies are being used to study diverse aspects of *Leishmania* biology such as parasite development, virulence, and drug resistance [17]. Thus, proteomics approaches have been used to determine differential patterns of protein expression between the promastigote and amastigote stages in *Leishmania mexicana* [18], *L. infantum* [19], and *L. donovani* [20], among others. Other studies have been aimed to ascertain specific proteomes by means of organelle fractionation to obtain enriched fractions of mitochondria, flagella, or glycosomes [21,22]. The identification and mapping of protein post-translational modifications (PTMs) provide additional information about the activation of specific pathways in a given growing condition, thus improving the knowledge on protein interactions in complex networks.

Here, we present a wide and detailed proteome of the *L. infantum* JPCM5 strain, based on axenically grown promastigotes in the logarithmic growth phase. A careful comparative analysis with other published proteomes from different Old and New World *Leishmania* species has also been carried out. Additionally, the MS data allowed for the identification of PTMs (phosphorylation, acetylation, methylation, formylation, and glycosylation) at specific protein sites that might have regulatory functions. Furthermore, we showed how the integration of proteomics with genomic and transcriptomic data represents a powerful and complementary strategy for gene annotation, as demonstrated before in a plethora of species [23]. Hence, by applying this proteogenomic approach, it was possible to improve the annotations for several *L. infantum* genes, as well as the identification of eight novel genes.

2. Materials and Methods

2.1. *Leishmania Infantum* Culture and Protein Extraction

L. infantum JPCM5 strain parasites were grown at 26 °C in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 15% of heat inactivated fetal calf serum (Biowest SAS, Nuaille, France). Promastigote cultures were initiated at 1×10^6 parasites/mL and harvested at the mid-logarithmic phase ($1\text{--}2 \times 10^7$ parasites/mL). Around $1\text{--}2 \times 10^8$ parasites were collected and washed twice with phosphate-buffered saline (PBS); finally, parasites were suspended by pipetting in 300 µL of a RIPA (RadioImmunoPrecipitation Assay) lysis buffer (Thermo Fisher Scientific, Rockford, IL, USA) in the presence of EDTA-free Easy Pack Protease inhibitor (Roche, Diagnostics, Mannheim, Germany). After 6 cycles (30 s pulse/30 s pause) of sonication in a bath at 4 °C, samples were incubated for 90 min at 4 °C, and, afterwards, protein lysates were centrifuged at 14,000 g for 30 min. The supernatant was collected and used for proteomics analyses.

2.2. In-Gel and In-Solution Digestion of Samples by Trypsin and Chymotrypsin

For the in-gel digestion of proteins, samples were mixed with an equal volume of a 2× Laemmli buffer and loaded onto 1.2-cm wide wells of a conventional SDS-PAGE gel (0.75 mm-thick, 4% polyacrylamide stacking-gel, and 10% polyacrylamide resolving-gel). The electrophoresis was stopped as soon as the electrophoretic front entered 3 mm into the resolving gel, so the proteins became concentrated in the stacking/resolving gel interface. After Coomassie staining, the protein-

containing gel was cut into small pieces (2×2 mm cubes) and placed into a microcentrifuge tube, as described elsewhere [24]. The gel pieces were destained in acetonitrile:water (ACN:H₂O, 1:1), reduced and alkylated (disulfide bonds from cysteinyl residues were reduced with 10 mM dithiothreitol (DTT) for 1 h at 56 °C, and then thiol groups were alkylated with 10 mM iodoacetamide for 1 h at room temperature in darkness), and digested in situ with sequencing grade trypsin (Promega, Madison, WI) or chymotrypsin (Roche Diagnostics), as described by Shevchenko et al. [25], with minor modifications. The gel pieces were shrunk by removing all liquid using sufficient ACN. Acetonitrile was pipetted out, and the gel pieces were dried in a speedvac. The dried gel pieces were re-swollen in 100 mM Tris-HCl and 10 mM CaCl₂ at pH 8 with 60 ng/μL trypsin or chymotrypsin at a 5:1 protein:enzyme (*w/w*) ratio. The tubes were kept on ice for 2 h and incubated at 37 °C (trypsin) or 25 °C (chymotrypsin) for 12 h. Digestion was stopped by the addition of 1% trifluoroacetic acid (TFA). Whole supernatants were dried down and then desalted onto OMIX C18 pipette tips (Agilent Technologies, Santa Clara, CA, USA) before the MS analysis.

Additionally, in-solution digestion was performed as described elsewhere [26]. After the denaturation of proteins with an 8 M urea, the protein sample was reduced and alkylated: disulfide bonds from cysteinyl residues were reduced with 10 mM DTT for 1 h at 37 °C, and then thiol groups were alkylated with 50 mM iodoacetamide for 1 h at room temperature in darkness. The sample was diluted to reduce urea concentration below 1.4 M and digested using sequencing-grade trypsin (Promega, Madison, WI, USA) or chymotrypsin (Roche Diagnostics) overnight at 37 °C (trypsin) or 25 °C (chymotrypsin) using a 1:20 (*w/w*) enzyme/protein ratio. Digestion was stopped by the addition of 1% TFA. Whole supernatants were dried down and then desalted onto OMIX C18 pipette tips (Agilent Technologies) before the MS analysis.

2.3. Reverse Phase-Liquid Chromatography Mass Spectrometry Analysis (RP-LC-MS/MS)

The digested protein samples (above) were resuspended in 10 μL of 0.1% formic acid and analyzed by RP-LC-MS/MS in an Easy-nLC II system coupled to an ion trap LTQ-Orbitrap-Velos-Pro hybrid mass spectrometer (Thermo Fisher Scientific). The peptides were concentrated (on-line) by reverse phase chromatography using a 0.1×20 mm C18 RP precolumn (Thermo Fisher Scientific) and then separated using a 0.075×250 mm C18 RP column (Thermo Fisher Scientific) operating at 0.3 μL/min. Peptides were eluted using a 180-min dual gradient. The gradient profile was set as follows: 5–25% solvent B for 135 min, 25–40% solvent B for 45 min, 40–100% solvent B for 2 min, and 100% solvent B for 18 min (Solvent A: 0.1% formic acid in water; solvent B: 0.1% formic acid and 80% acetonitrile in water). ElectroSpray ionization (ESI) was done using a nano-bore emitter stainless steel ID 30 μm (Proxeon) interface. The Orbitrap resolution was set at 30,000. Peptides were detected in survey scans from 400 to 1600 amu (1 μscan), followed by twenty data-dependent MS/MS scans (Top 20) using an isolation width of 2 u (in mass-to-charge ratio units), a normalized collision energy of 35%, and a dynamic exclusion that was applied during 60 s periods.

2.4. Data Analysis

Peptide identification from raw data was carried out using the PEAKS Studio X search engine (Bioinformatics Solutions Inc, Waterloo, ON, Canada). A custom Python script was used to create a database comprising all possible open reading frames (ORF) coding for protein sequences of ≥ 20 amino acids existing in any of the six-frames in the *L. infantum* JPCM5 strain genome sequence [13]. This database (named LINF-all-ORFs) consisted of 294,654 entries. In parallel, a fusion-database, created by merging the *L. infantum* protein sequences annotated in UniProt and the LINF-all-ORFs entries, was also used by the search engine. Finally, a search against a decoy database (decoy fusion-database) was also performed. The following constraints were used for the searches: tryptic or chymotryptic cleavage (semispecific), up to two missed cleavage sites, tolerances of 20 ppm for precursor ions and 0.6 Da for MS/MS fragment ions, and optional Met oxidation and Cys carbamidomethylation were allowed. The false discovery rates (FDRs) for peptide spectrum matches (PSMs) were limited to 0.01 or lower. Those proteins that were identified with at least two distinct peptides were considered for further analysis [27–29].

The LINF-all-ORFs entries with mapped peptides were compared with the annotated *L. infantum* proteins (UniProt database) using in-house Python scripts in order to identify both the misannotations and novel proteins. Additionally, Python scripts were used to ascribe post-translational modifications to particular protein entries.

Functional categories and enzymatic pathways using the DAVID program (Functional Annotation Tool, DAVID Bioinformatics Resources 6.8) and the KEGG Pathway (Kyoto Encyclopedia of Genes and Genomes) were used for the classification of the proteins identified by MS.

3. Results and Discussion

3.1. Protein Identification from the LC-MS/MS Peptide Spectra

The main objective of this study was to obtain the experimental proteome of the *L. infantum* promastigote stage with the additional aim of improving current genomic annotations. For this purpose, a recently published re-sequenced genome [13] was used. However, we did not restrict the search of peptide spectra on currently annotated protein-coding genes; instead, a database consisting of all possible polypeptides (equal or larger than 20 amino acids) was used (see Materials and Methods for further details). The workflow used for sample preparation and proteomics is shown in Figure 1. From the MS/MS data, it could be seen that only those associated with peptides longer than seven amino acids were considered for protein identification. Among the identified proteins, 2344 proteins matched with previously annotated proteins [13]. Moreover, eight novel proteins were uncovered, thus legitimating the search strategy. In addition, some ORFs had to be extended to accommodate the MS-identified peptides (see below for further details about these findings). Most of the proteins—70.5% (1659 out of 2352)—were identified by three or more unique peptides per protein, 14.5% (341) of the protein identifications were supported by two unique peptides, and only 15% (352) of the identifications were done by a single unique peptide. Currently, 3482 out of 8590 annotated proteins (around 40%) in the *L. infantum* genome have the status of hypothetical proteins; the MS spectra obtained in this work provided experimental evidence of the real existence for 456 of those hypothetical proteins.

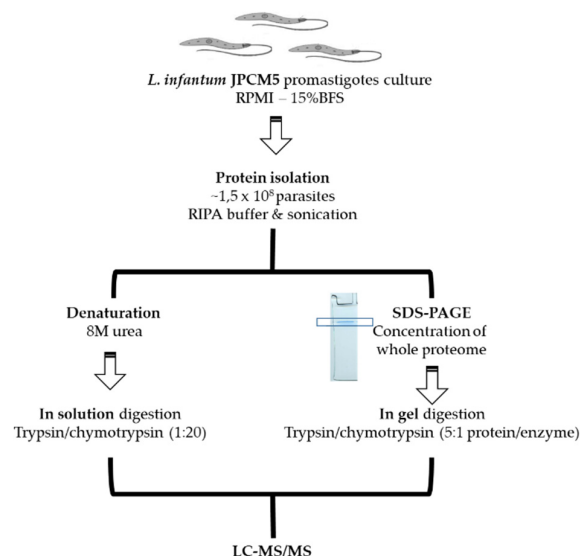


Figure 1. Workflow for protein extraction and proteomic analyses of *Leishmania infantum* promastigotes. The experimental MS data were searched against the UniProt protein database and a database consisting of all possible polypeptides encoded in the six-frames of the *L. infantum* genome (based on v2/2018; www.leish-esp.cbm.uam.es; [13]).

The first comprehensive study aimed to characterize the *L. infantum* proteome was carried out by the Papadopoulos's group [30]. Using two-dimensional (2D) gel electrophoresis, these authors visualized 2261 protein spots in promastigote samples and 2273 spots in amastigote ones. However, after MS analysis, only 168 protein spots, derived from 71 different genes, could be identified [30]. A better proteome resolution was attained after a fractionation step including digitonin extraction; hence, 153 *L. infantum* proteins were identified by MS analysis of selected spots [31]. The combination of two-dimensional liquid chromatography (2DLC), electrospray ionization mass spectrometry (2DLC-ESI-MS), and 2DLC-matrix-assisted laser desorption/ionization mass spectrometry (2DLC-MALDI-MS) allowed Leifso and co-workers to identify 91 *L. infantum* proteins [19]. An enrichment for basic proteins using the technique of free-flow electrophoresis prior to separation by 2D gel electrophoresis led to the identification of around 200 *L. infantum* proteins [32]. Alcolea and coworkers [33] identified 28 proteins in a proteomic study aimed to uncover differentially expressed proteins between the early-logarithmic and the stationary phases during the culturing of *L. infantum* promastigotes. In two different studies using MS analysis of the exoproteome derived from *L. infantum* promastigote cultures, a total of 102 [34] and 494 [35] proteins were identified. Therefore, our work provides the most complete, to date, experimentally evidenced proteome for *L. infantum*.

Outstanding studies on proteome identification have been performed in both *L. donovani* and *L. major*. In 2008, Rosenzweig and collaborators reported the identification of 1713 proteins in *L. donovani* [20]. A comparison between the proteins identified in our work (*L. infantum* JPCM5) and those identified in *L. donovani* showed that 1218 proteins were common (orthologs) in both studies (Figure 2). We failed to identify 207 proteins of those reported in *L. donovani*, whereas we found 1130 proteins that are absent from the *L. donovani* proteome reported by Rosenzweig et al. [20].

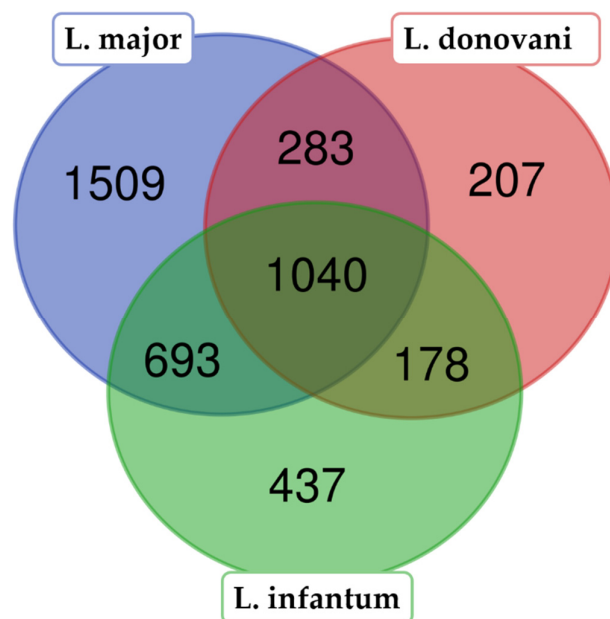


Figure 2. Comparison (Venn diagram) between the identified proteins in this work (*L. infantum*, in green) and those identified in two previous studies [20,36] performing proteomic analysis in *Leishmania donovani* (in red) and *Leishmania major* (in blue). The Venn diagram was created by the tool available at bioinformatics.psb.ugent.be/webtools/. Note—the discrepancy between the number of proteins identified by Rosenzweig et al. [20] (see text) and those represented in the Venn diagram (1713 vs. 1708) was due to 5 gene duplications that were corrected after re-assembling of the *L. infantum* genome [13].

More recently, Pandey and coworkers reported the identification of 3386 different proteins in *L. donovani* promastigote and amastigote stages [37,38]. After comparing their data and the proteins identified in this study, 1650 of the proteins observed in *L. infantum* promastigotes were found to be

present (their orthologues) in the *L. donovani* promastigote proteome. However, among the 613 proteins that Nirujogi et al. [37] reported to be exclusively expressed in *L. donovani* amastigotes, 126 proteins were also identified in our proteomics study, thus indicating that these proteins are also being expressed in the promastigotes stage, at least in *L. infantum* (see Supplementary File, Table S1). Most of them were annotated as hypothetical proteins or with unknown function, but there are also metabolic enzymes, translation machinery components (ribosomal proteins and eukaryotic initiation factors), and RNA binding proteins.

In 2014, Pawar et al. [36] reported a quite wide proteome of the *L. major* promastigote stage, in which 3613 proteins were identified. These authors followed a proteogenomic approach, as we did in this study, consisting of searching the mass spectra against a six-frame translated database generated from a complete genome sequence. An orthology-based comparison indicated that the *L. major* promastigote proteome and the *L. infantum* proteome of this study shared 1733 proteins (Figure 2). Moreover, considering the 1792 proteins identified in the *L. major* proteome, though not in our study, and the 615 proteins exclusively identified by us in the *L. infantum* proteome, the total number of identified proteins presumably expressed in the promastigote stage is 4140 (roughly half of the predicted proteins to be encoded in the *Leishmania* genome).

3.2. Representativeness of the Translational Machinery and RNA Binding Proteins in the *L. infantum* Experimental Proteome

Around 200 proteins from the *L. infantum* promastigote proteome were categorized as components of the translational machinery: 122 ribosomal proteins, 51 translation regulatory factors, and 24 tRNA synthases (Supplementary File, Table S2). As expected for a highly proliferative stage (the promastigotes were growing in the logarithmic phase when harvested for analysis), in which protein synthesis needs to be very active, all the ribosome components, tRNA-loading enzymes, and regulatory factors were abundant and easily detected by mass spectrometry. Nevertheless, in contrast, very few of the annotated mitoribosomal proteins [39] were identified in the *L. infantum* promastigote proteome. This observation may indicate that mitochondrial ribosomes are in relatively low amounts when promastigotes are grown in nutrient-rich culture media.

Proteins with RNA binding properties deserve special attention, since gene expression in *Leishmania* and related trypanosomatids is essentially being controlled at the post-transcriptional level [40,41]. In this scenario, RNA-binding proteins are key players in controlling RNA metabolism [42–44]. In the *L. infantum* promastigote proteome, a large number of known RNA binding proteins were detected (Supplementary File, Table S3). Apart from the mentioned ribosomal proteins and translation factors, 15 RNA helicases were detected, as well as many of the RNA-binding domain-containing proteins reported in a recent study aimed to the capture and identification of RNA-bound proteins in *L. donovani* [45]. The RNA-binding proteins of the Pumilio family (aka PUF proteins) are especially numerous (11 members) in *Leishmania* [46]. In this study, we identified 6 out of 11 PUF proteins that are being expressed in the promastigote stage of *L. infantum*; these are PUF 1, PUF 4, PUF 6, PUF 7, PUF 8, and PUF 10 (see Table S3 to see their gene IDs).

3.3. Metabolic Enzymes and Pathways

Going deeper, we performed an in-silico pathway reconstruction using the detected proteins in the *L. infantum* promastigote proteome. By using the KEGG database resource (<http://www.genome.jp/kegg/>) accessed via the DAVID package, a total of 578 (27.6%) of the detected proteins could be classified into pathways representing classical cellular processes. In particular, 236 proteins were identified as metabolic enzymes; 31% of these enzymes belong to glycolysis (Tables 1 and 2), the tricarboxylic acid (TCA) cycle, and the pentose phosphate cycle (Supplementary File, Table S4), which are three metabolic pathways playing essential maintenance functions in the cell [47]. Remarkably, the complete set of 29 enzymes that make up the TCA cycle were identified in the promastigote proteome (Figure 3).

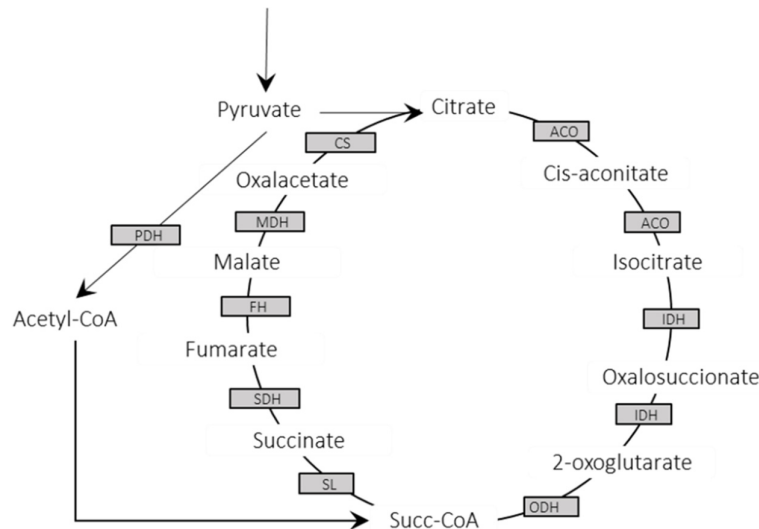


Figure 3. Detected enzymes in the *L. infantum* JPCM5 proteome composing the complete tricarboxylic acid (TCA) cycle. PDH: pyruvate dehydrogenase; ACO: aconitase; IDH: isocitrate dehydrogenase; ODH: 2-oxoglutarate dehydrogenase; SL: succinyl-CoA ligase; SDH: succinyl dehydrogenase; FH: fumarate hydratase; MDH: malate dehydrogenase; and CS: citrate synthase.

The glycosome is a trypanosomatid-specific, membrane-enclosed organelle that contains glycolytic enzymes, among others. Thus, glycolysis in *Leishmania* takes place in these organelles for the steps between glucose and 3-phosphoglycerate [48], as well as in the cytosol for those late steps leading to the formation of pyruvate [20]. The identified enzymes involved in these two stages are listed in Tables 1 and 2. Among them, there are 32 enzymes belonging to the glycosomal/cytosolic glycolysis (and gluconeogenesis) pathway until the formation of pyruvate by pyruvate kinases (IDs LINF_350005400 and LINF_350005300). Some enzymes involved in the mitochondrial electron transport respiratory chain were detected (Supplementary File, Table S5). Similar findings were found by Rosenzweig and collaborators in the *L. donovani* promastigote proteome [20]. Several proteins of the electron transport chain are encoded by the kinetoplast DNA maxicircles [49] such as cytochrome oxidase subunits and NADH dehydrogenase, but they were not searched in this study.

Table 1. List of glycosomal enzymes related to gluconeogenesis and glycolysis identified in *L. infantum* (according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database).

Gene ID	Unique Peptides	Description
LINF_040016700	12	Fructose-1-6-bisphosphatase
LINF_120010600	43	Glucose-6-phosphate isomerase
LINF_200006000	41	Phosphoglycerate kinase C-glycosomal
LINF_210007800	47	Hexokinase
LINF_210012000	13	Phosphoglucomutase
LINF_230009500	10	Aldose 1-epimerase-like protein
LINF_240013700	32	Triosephosphate isomerase
LINF_250017300	46	Aldehyde dehydrogenase—mitochondrial precursor
LINF_270024900	56	Glycosomal phosphoenolpyruvate carboxykinase
LINF_290032900	23	ATP-dependent phosphofructokinase
LINF_300035000	49	Glyceraldehyde 3-phosphate dehydrogenase—glycosomal
LINF_300039500	13	PAS-domain containing phosphoglycerate kinase
LINF_340040800	13	Aldose 1-epimerase-like protein

Table 2. List of cytosolic enzymes related to gluconeogenesis and glycolysis identified in the *L. infantum* proteome (according to the KEGG pathway database).

Gene ID	Unique Peptides	Description
LINF_140018000	55	Enolase
LINF_180019200	23	Pyruvate dehydrogenase e1 component α subunit
LINF_210011100	10	Dihydrolipoamide acetyltransferase
LINF_210012000	13	Phosphoglucomutase
LINF_230009000	43	NADP-dependent alcohol dehydrogenase
LINF_230014800	47	Acetyl-CoA synthetase
LINF_250023800	29	Pyruvate dehydrogenase e1 β subunit
LINF_290025700	4	Dihydrolipoamide dehydrogenase
LINF_310034500	2	Dihydrolipoamide dehydrogenase
LINF_320040600	37	Dihydrolipoamide dehydrogenase
LINF_350005300	43	Pyruvate kinase
LINF_350005400	43	Glyceraldehyde 3-phosphate dehydrogenase—cytosolic
LINF_360030600	19	Dihydrolipoamide acetyltransferase precursor

An active energy metabolism requires enzymes to be involved in redox homeostasis. Several of these enzymes have been identified in the *L. infantum* proteome, and they are likely abundant, as judged by the large number of unique peptides that were mapped to them. The detected proteins were trypanothione (LINF_150019000, with 31 unique peptides), peroxidothione (LINF_230005400, 31 peptides), cyclophilin (LINF_060006300, 20 peptides), iron superoxide dismutases (LINF_080007900 and LINF_320024000, with 14 and 18 peptides, respectively), and several elongation factors 1 β (LINF_340014200 and LINF_340014000 with 16 peptides each and LINF_360020500 with 19 peptides).

3.4. Components of the Proteostasis Network

The proteasome is a complex of multi-subunit proteases, associated with protein degradation, but in protozoan parasites, it has been also involved in cell differentiation and replication processes [50]. In fact, proteasomal inhibitors have been described as promising therapeutic targets for leishmaniasis and trypanosomiasis [51,52]. According to the KEGG database, the complete compendium of proteasomal proteins were identified in this study (Supplementary File, Table S6).

Protein degradation and protein folding cooperate to maintain protein homeostasis or proteostasis [53]. Multiple and drastic environmental changes (pH variation, sudden temperature up/down-shifts, and oxidative stress) occur along the *Leishmania* life cycle. Most often, these environmental insults promote protein unfolding and aggregation; to counteract these effects, cells possess specialized molecular chaperones (or heat shock proteins: HSPs) that serve as central integrators of protein homeostasis. Not surprisingly, *Leishmania* parasites possess a large number and variety of molecular chaperones [54]. In this study, we identified proteins belonging to the different HSP families: HSP100, HSP83/90, HSP70, HSP60, HSP40/DnaJ, and HSP20 (listed in Supplementary File, Table S7). Mitochondrion is a cellular organelle in which molecular chaperones are of particular relevance because they are involved in protein transport across membranes and protein refolding inside the mitochondria. Recently, the mitochondrial proteome was analyzed in *L. tropica* [22]. Taking advantage of that study, in Table 3, we list those HSPs identified in the *L. infantum* proteome that are potentially mitochondrial proteins.

Table 3. Identified molecular chaperones of *L. infantum* promastigotes, with putative mitochondrial location (according to Tasbihi et al. [22]).

Gene ID	Unique Peptides	Molecular Chaperone
LINF_260011400	9	Heat shock protein 10 (HSP10)
LINF_320040500	5	HSP40/JDP45
LINF_350035100	7	HSP40/JDP50
LINF_360027100	19	HSP60/cpn60.2
LINF_360027200	26	HSP60/cpn60.3
LINF_240010000	5	HSP40/JDP8
LINF_260017400	42	HSP70.4
LINF_280017800	51	Grp78/BiP
LINF_330033000	48	HSP75/TRAP-1
LINF_020012400	23	HSP78
LINF_330009000	22	HSP83/90

3.5. Glycosomal Proteins Represent a Substantial Fraction of the Experimentally Detected Proteins in the *L. infantum* Promastigote

As mentioned above, glycosomes are specialized peroxisomes that contain key enzymes involved on energy metabolism and purine salvage [55]. Moreover, as occurs in peroxisomes, glycosomal proteins are targeted for import to and location in glycosomes by the presence of the peroxisomal targeting signals (PTSs) PTS1 and PTS2. Two essential proteins for targeting newly synthesized proteins, with a PTS2 import signal, into the glycosome are peroxins (PEXs) 5 and 7 [56]. Remarkably, both proteins, PEX5 (LINF_350019100) and PEX7 (LINF_290012400), were identified in the experimental proteome of *L. infantum* promastigotes. Moreover, we made a direct comparison between the *L. infantum* proteome reported here and two studies focused on glycosomal proteomes in *Leishmania tarentolae* and *L. donovani* [48,57]. Colasante and collaborators [48] identified 464 proteins in a glycosomal membrane preparation from *L. tarentolae*, and they concluded that 258 would be glycosomal proteins, including 40 enzymes. Interestingly, the orthologs of 165 (64%) of these proteins were experimentally identified in the *L. infantum* promastigote proteome. In particular, complete enzymatic complements involved in glycosomal glycolysis and gluconeogenesis steps were identified in both studies. Jamdhade and coworkers [57] reported the proteome analysis of an enriched glycosome fraction from *L. donovani* promastigotes, identifying 1355 proteins. In our study, orthologs to 853 of those putative *L. donovani* glycosomal proteins were found; these are listed in Supplementary File, Table S8.

The purine salvage pathway, essential for trypanosomes, also takes place in glycosomes [58]. Notably, the 13 enzymes composing the purine salvage pathway (Figure 4) [58] were identified in this *L. infantum* experimental proteome, which was in agreement with the relevance of this metabolic route for parasite survival. In this regard, it is somewhat unexpected that only two enzymes (adenylosuccinate synthetase (ADSS) and inosine monophosphate dehydrogenase (IMPDH)) were identified in the *L. tarentolae* glycosomal proteome, and five of these enzymes were identified in the glycosomal fractions of *L. donovani*. Similarly, we identified most of the enzymes constituting the de novo pyrimidine biosynthesis pathway (Table 4) [59] in the *L. infantum* proteome, whereas Jamdhade et al. [57] only found one enzyme of this pathway in the *L. donovani* glycosomal proteome—the orotate phosphoribosyltransferase (LDBPK_160560).

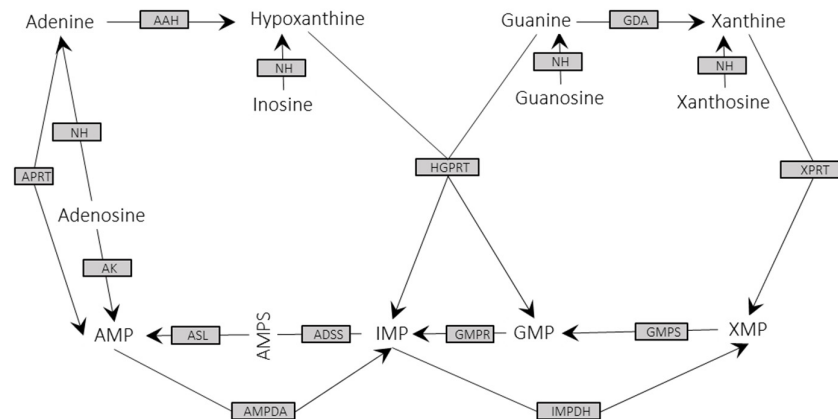


Figure 4. Enzymes from the purine salvage cycles identified in the *L. infantum* experimental proteome. All the enzymes (grey squares) that are required to complete the pathway were identified in this study. APRT: adenine phosphoribosyltransferase (LINF_130016900); NH: nucleoside hydrolase (LINF_180021400); AK: adenosine kinase (LINF_300014400); AAH: adenine aminohydrolase (LINF_350026800); ASL: adenylosuccinate lyase (LINF_040009600); AMPDA: AMP deaminase (LINF_130014700); ADSS: adenylosuccinate synthetase (LINF_130016900); GMPR: GMP reductase (LINF_170014800); GMPS: GMP synthase (LINF_220006100); HGPRT: hypoxanthine-guanine phosphoribosyltransferase (LINF_210014900); GDA: guanine deaminase (LINF_290014000); IMPDH: inosine monophosphate dehydrogenase (LINF_190022000); and XPRT: xanthine phosphoribosyltransferase (LINF_210015000). The cycle was depicted according to Boitz et al. [58].

Table 4. Enzymes involved in de novo pyrimidine biosynthesis identified in *L. infantum* promastigote proteome.

Gene ID	Unique Peptides	Description
LINF_060011200	7	Deoxyuridine triphosphatase
LINF_160010400	5	Dihydroorotate dehydrogenase (fumarate)
LINF_160010500	15	Aspartate carbamoyltransferase
LINF_160010700	28	Orotate phosphoribosyltransferase
LINF_160011200	6	Carbamoyl-phosphate synthase
LINF_180021400	17	Nonspecific nucleoside hydrolase
LINF_340016700	8	Uracil phosphoribosyltransferase

3.6. Exoproteome Components Identified in the *L. infantum* Experimental Proteome

Leishmania-secreted molecules and exosomes are particularly relevant for infection establishment because parasites and exosomes are co-egested during the insect blood meal [60]. A detailed analysis of the *L. infantum* secreted proteins (exoproteome) was carried out by Santarem et al. [35]. These authors found that the proteome profiles were distinct depending on the metabolic stage of the parasites (logarithmic or stationary phase promastigotes). The number of distinct proteins identified in that study was 297, and around 90% of them were also identified in the proteome reported here. In another outstanding study, Atayde et al. [61] analyzed the proteomic composition of *L. infantum* exosomes and extracellular vesicles that were directly isolated from the sand fly

midgut. Table 5 lists proteins commonly present in exosome preparations; all of them were identified in the *L. infantum* experimental proteome reported here.

Table 5. Common components of *Leishmania* exosomes identified in this *L. infantum* proteome study.

Gene ID	Description	Features [Ref.]
LINF_050017500; LINF_040007000	Surface antigens	Virulence factor [61]
LINF_090013900	Oligopeptidase b	Virulence factor [35]
LINF_100010100	GP63-leishmanolysin	Virulence factor [60,61]
LINF_120014700	Surface antigen protein 2	Virulence factor [61]
LINF_140018000	Enolase	Virulence factor [35,62]
LINF_150019000	Tryparedoxin peroxidase	Virulence factor [61]
LINF_170005900	Elongation factor 1- α	Exosome marker [61]
LINF_190020600	Cysteine peptidase A (CPA)	Virulence factor [61]
LINF_200018000	Calpain-like cysteine peptidase	Virulence factor [61]
LINF_280035000	HSP70	Exosome marker [61]
LINF_280036000		
LINF_280034700	Receptor for activated C kinase 1	Immunomodulator [35]
LINF_320036700	Nucleoside diphosphate kinase b	Immunomodulator [35]
LINF_330009000	HSP83/90	Exosome marker [61]
LINF_350027300		
LINF_350027500	Kinetoplastid membrane protein 11 (KMP11)	Immunomodulator [35]
LINF_360018400	Fructose-1-6-bisphosphate aldolase	Immunomodulator [60]

3.7. Other Relevant Proteins Identified in the *L. infantum* Promastigote Proteome

Proteins with a high molecular weight (HMW) represent a challenge for mass spectrometry-based assays, as they are usually underrepresented in protein extracts used for proteomic analysis. To overcome this limitation, Brotherton et al. [63] optimized extraction protocols to enrich HMW proteins and membrane proteins in *L. infantum* promastigotes and amastigotes. In our study, we confirmed the presence of tryptic and/or chymotryptic peptides from 35 HMW proteins with a molecular weight (MW) higher than 200 kDa (Supplementary File, Table S9). Among them, the identification of a calpain-like cysteine peptidase was remarkable, as it had an estimated MW of around 700 kDa (LINF_270010200) and was identified by 124 unique peptides, thus covering 24% of the amino acid sequence.

The flagellum is a characteristic organelle of *Leishmania* that confers motility to the parasite in the promastigote stage, during which this structure is particularly prominent. In a recent publication, an exhaustive structural and functional characterization of the *L. mexicana* promastigote flagellum was reported [64]. In that study, flagella preparations were analyzed by proteomics, and this allowed for the identification of 701 unique proteins for this organelle. Orthologues to around 400 flagellum-specific proteins were identified in the *L. infantum* proteome described here. More importantly, most of the proteins relevant for flagellum assembly and motility in *L. mexicana* promastigotes [64] were identified in the experimental proteome of *L. infantum* promastigotes (Supplementary File, Table S10).

3.8. Detection of Post-Translational Modifications

The PTMs of proteins influence their activity, structure, turnover, localization, and capacity to interact with other proteins. In *Leishmania*, PTMs, together with mRNA stability and translation processes, are the essential regulators of gene expression. In this study, based on MS/MS spectra, a significant number of phosphorylated, methylated, acetylated, glycosylated, and/or formylated proteins were identified in the *L. infantum* proteome. Thus, even though specific enrichments of modified proteins were not performed, we identified modified peptides that accounted for 10 phosphorylated, 144 methylated, 192 acetylated, 28 formylated, and 3 glycosylated proteins.

The phosphorylation of serine (S), threonine (T), and tyrosine (Y) amino acids implies an increase of 79.97 Da in their molecular weights (unimod.org). The phosphorylated proteins identified in this study and the modified residues are listed in Table 6. Apart from two unknown phosphoproteins (LINF_040005600 and LINF_220013200), the ribosomal protein S10, α tubulins, an rRNA biogenesis protein-like protein, the 3-ketoacyl CoA-thiolase, the flagellar protein KHARON1, the glycogen synthase kinase 3 (GSK-3), and the prototypical HSP70 might be regulated by phosphorylation (Table 6). In fact, the phosphorylation of HSP70 has been reported to occur during the stress response in both promastigotes and amastigotes of *L. donovani* [65].

Table 6. Phosphoproteins in the *L. infantum* proteome.

Gene ID	Description	Position
LINF_040005600	Hypothetical protein-conserved	T187
LINF_130007700; LINF_130007800; LINF_130008000; LINF_130008200; LINF_130008300; LINF_130008400; LINF_130008600; LINF_130008700	α tubulin	T334; Y357
LINF_180007700	Glycogen synthase kinase 3 (GSK-3)	Y186
LINF_190017000	Hypothetical protein-conserved	S120
LINF_200011800	rRNA biogenesis protein-like protein	Y571
LINF_220013200	Hypothetical protein-conserved	S45
LINF_230014600	3-ketoacyl-CoA thiolase	S229
LINF_280035400	HSP70	T159; T164
LINF_360015600; LINF_360015700	40S ribosomal protein S10	S157
LINF_360068400	Flagellum targeting protein KHARON1	S158

Most of the observed phosphorylation events occurred on S and T, but in some proteins, phosphorylation on Y residues was also detected. Rosenzweig et al. [66] identified 16 phosphorylated proteins in *L. donovani* in either promastigotes or amastigotes; in that study, all of phosphorylations occurred at S or/and T residues. No coincidences exist between the phosphoproteins identified by these authors and those identified in this study (Table 6); however, this was not unexpected because phosphorylation is a dynamic modification and the numbers of phosphorylated proteins identified in both studies were low. Though the occurrence of phosphorylation events has been proven to be much lower in tyrosine residues [67], it is remarkable that phosphorylated tyrosines were found in α tubulins, the rRNA biogenesis protein, and in GSK-3 (Table 6). Some of these phosphorylated proteins have also been identified in previous studies. Thus, for instance, the 3-ketoacyl-CoA thiolase was found to be phosphorylated (at serine 229) in *L. donovani* promastigotes [68]. Kinases and phosphatases are enzymes implicated in the regulation of phosphorylation/dephosphorylation processes, and, accordingly, several serine, threonine, and tyrosine kinases and phosphatases have been identified in the *L. infantum* proteome (Supplementary File, Table S11).

Methylation (+14 Da) is a physiological PTM that occurs at the C- and N-terminal ends of proteins, and on the side chain nitrogen of arginine (R) and lysine (K); this modification is critical for regulating several cellular processes. Apart from those amino acids, methylations have been found to occur in other amino acids like aspartic acid (D), glutamic acid (E), histidine (H), glutamine (Q), and asparagine (N) [69]. In the *L. infantum* proteome, 139 proteins were predicted to be methylated, 76 of them showed methylation at K or R residues, 123 of the modified proteins showed D and/or E methylated residues, and a methylated-H was found in β -tubulin. All the methylated proteins detected in this study are listed

in Supplementary File, Table S12. In summary, our findings pointed out that methylation at D and E residues would be relatively frequent in *Leishmania*; as suggested by Sprung et al. [69], methylations at E and D residues would increase the hydrophobicity of the modified proteins. Some examples of highly methylated proteins identified in this study are α and β tubulins, heat shock proteins HSP70 and HSP83/90, and the elongation factor 2 (eEF2). Many orthologs to the methylated proteins detected in this study were also identified as methylated in *L. donovani* [66].

Acetylation (+42.02 Da) is a PTM considered as relevant as phosphorylation in metabolic and signaling pathways. K acetylation has been described as a reversible enzymatic reaction that regulates protein function, as it is particularly relevant in chromatin compaction by the acetylation of histones [70]. Interestingly, the accumulation of acetylated histones has been observed at the polycistronic transcription initiation sites in *L. major* and *Trypanosoma cruzi* [71,72]. However, in the *L. infantum* proteomics data, peptides bearing acetylated K belonging to histones were not identified, as was the case in the *L. donovani* proteome [66], thus suggesting a relatively low proportion of acetylated histones in the bulk of total cellular histones. Some examples of proteins detected as acetylated in this study are β tubulins (LINF_330015100, LINF_330015200, and LINF_210028500; modified at K297), guanylate kinase (LINF_330018400, K3), the subunit β of ATP synthase (LINF_250018000; K511) and a calpain-like cysteine peptidase (LINF_140014400; K74).

In addition, acetylation at the N-terminal ends of proteins may occur either co- or post-translationally, as it is a frequent modification in eukaryotic proteins even though their physiological consequences remain poorly understood [73]. This irreversible modification affects protein fate in the cell and is carried out by N-terminal acetyltransferases (Nat). In the *L. infantum* promastigote proteome, we identified three of these enzymes: Nat-1, Nat-B, and Nat-C (Supplementary File, Table S11). On the other hand, among the 144 N-terminally acetylated proteins identified in this study (Supplementary File, Table S13), 40 proteins showed acetylation at their initial methionine (iM), and 104 were acetylated at the second amino acid, suggesting a cleavage of the iM during protein maturation [74]. In the cases in which acetylation takes place at the iM, we detected a bias in the amino acids located behind the iM. Thus, in 40% of the acetylated proteins, the second amino acid was the polar non-charged N or Q residues (Figure 5A also shows the other more frequent amino acids located at the second position). An acetylation reaction after iM removal was mainly found on S (55% of the cases) and alanine (A) (in 31%) residues (Figure 5B). These two amino acids, as well as threonine (found in 7.7% of the detected acetylated residues), have small side chains, a feature previously noted to favor a more efficient iM cleavage in the course of protein maturation [66,75].

On the other hand, the analysis of acetylated peptides allowed us to correct the initiator AUG codon (and, therefore, the predicted amino acid sequence) of four previously misannotated genes (whose new coordinates are indicated in Supplementary File, Table S13). One example is illustrated in Figure 5C; the coding sequence of the LINF_010008200 gene (coding for a poly (A) export protein) should be extended upstream 36 triplets based on the existence of an acetylated peptide encoded in that region.

Glycosylation also plays a relevant role in protein maturation, as well as in signal transduction mechanisms [76]. In this *L. infantum* proteomic study, we detected hexosylations or N-acetylhexosamine addition in three proteins; these modifications were consistent with N-linked glycosylation events at N residues. These PTM-modified proteins are cysteine peptidase A (CPA; LINF_190020600) modified at N345, 3-hydroxy-3 methylglutaryl CoA synthase (at N340; LINF_240027300), and PUF6 (at N483; LINF_330019100). The characterization of glycoproteins and the nature of their modifications remain challenging due to the complexity and variety of glycan moieties. In this study, two single modifications (hexosylation and N-acetylhexosamine addition) were searched, and this explained the extremely low number of detected glycosylated proteins. Rosenzweig et al. [66] found 13 glycosylated proteins—only one in asparagine and the rest in serine and threonine residues.

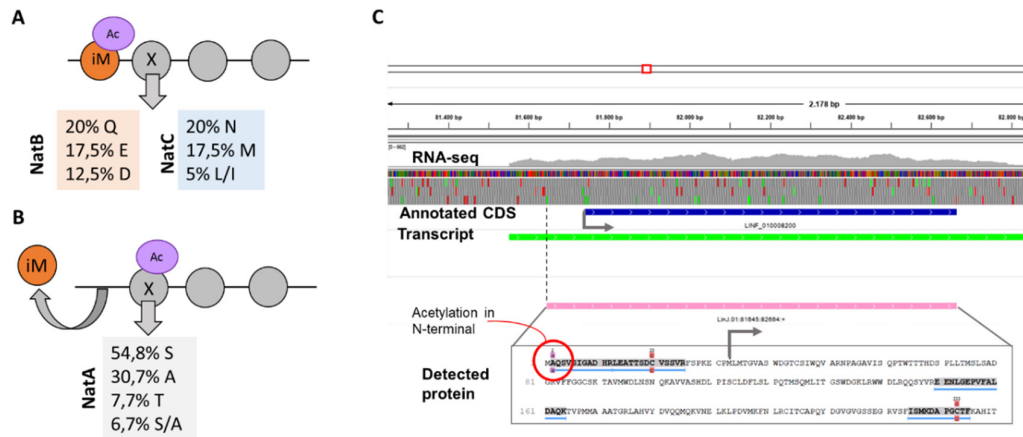


Figure 5. Features of the N-terminal acetylated proteins identified in this study. **(A)** Frequencies (percentages) of amino acids found next to the acetylated initial methionine (iM) and the putative enzymes responsible for the acetylation. **(B)** Percentages of amino acids found to be acetylated after cleavage of the iM and the putative enzyme involved in the reaction. **(C)** An example illustrating the usefulness of proteomics data for improving gene annotations. An acetylated peptide (red circle on the grey shaded sequence) was mapped to sequences located upstream of the currently annotated coding sequence for gene LINF_010008200 (blue box). The corrected gene (pink box) fit well into the transcript (green box). The image in **(C)** was generated using the Integrative Genome Viewer (IGV). CDS: coding sequence.

Formylation has not been previously described in trypanosomatids, but N-terminal methionine formylation in eukaryotes has been linked to cellular stress and protein degradation processes [77]. In particular, formyl-lysine residues has been found in histones and other nuclear proteins [78]. In our study, eight proteins were detected as formylated, mainly at leucine (in β -tubulin and typaredoxin), glycine (in a hypothetical protein; LINP_260024300), serine (in an RNA-guanylyltransferase) and lysine (in HSP83/90, a paraflagellar rod protein and a transaldolase) residues (see Supplementary File, Table S14). Future research on protein formylations in *Leishmania* and other organisms will provide insight into the physiological significance of this kind of PTM.

3.9. Proteogenomics

After assembling a genome, dedicated programs conduct the automatic annotation of ORFs. However, this annotation is not definitive, and a continuous effort of curation is needed. Transcriptomic analysis enables the obtainment of complete gene model annotation, including untranslated regions that are key to understand post-transcriptional regulation mechanisms. In addition, a proteogenomic analysis, such as that reported here, represents a powerful and useful approach for the identification of non-annotated genes, the correction of misannotations, or the validation of gene annotations [23]. For this purpose, in this work, the experimentally obtained peptide spectra were searched against all the polypeptides longer than 20 amino acids predicted from the ORFs found in the six possible translation frames of the recently re-sequenced genome for *L. infantum* JPCM5 strain [13]. The majority of the identified peptides fit well in current gene annotation (available at TriTryb.org and <http://leish-esp.cbm.uam.es/>). Nevertheless, some peptides mapped to non-annotated coding-regions in the *L. infantum* (JPCM5) genome, uncovering eight novel protein-coding genes (Supplementary File, Table S15). Figure 6 shows an example of a novel hypothetical protein found in chromosome 27, together with the MS spectra of the two peptides that allowed for its identification.

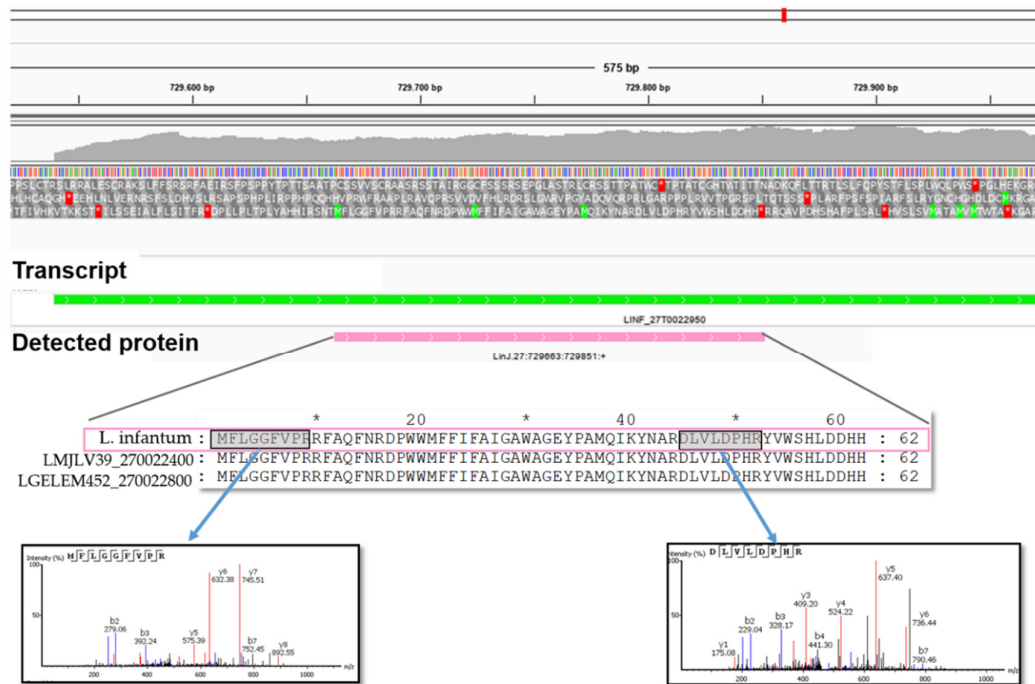


Figure 6. Identification of a novel protein based on the mapping of two experimentally detected-peptides in a region of *L. infantum* JPCM5 chromosome 27, which currently lacks an annotated ORF. The new CDS, named LINF_ 270,022,950 (pink box), fit well within a predicted transcript (LINF_27T0022950; green box). Interestingly, the predicted amino acid sequence was well-conserved when compared with proteins annotated in the genomic assemblies of *L. major* LV39 (ID: LmjVL39_270022400) and *Leishmania gerbilli* LEM452 (ID: LGELEM452_270022800).

As mentioned above (and illustrated in Figure 5), some of the detected peptides were mapped to regions located upstream of annotated coding sequences (CDS). This led to the addition of N-terminal extensions to 34 annotated proteins and the establishment of new translation start codons for their corresponding genes (Supplementary File, Table S16). All the detected peptides were confirmed to be unique, and the accuracy of their MS/MS spectra was manually revised.

4. Conclusions

In the last years, the characterization of trypanosomatid proteomes has become an active area of research. Here, we reported a proteomic analysis of *L. infantum*'s (JPCM5 strain) promastigote stage, and it was the first whole proteomic study in this species after the re-sequencing and de novo assembly of its genome in 2017 by González-De la Fuente et al. [13]. In addition, the search of the MS/MS spectra was performed against any possible ORFs larger than 20 triplets that existed in all-six frames from the *L. infantum* genome sequence. As a result, we identified 2352 proteins (Table S17), most of them corresponding to the predicted sequences in current gene annotations (TriTrypDB.org). Comparisons between the results of this study and previous proteomics data derived from promastigote stages in different *Leishmania* species showed a significant level of similarity regarding the type of detected proteins. Nevertheless, this proteomic study showed experimental evidence on the expression in this parasite stage of 123 proteins that were not detected in previous studies; these proteins are listed in Supplementary File, Table S18. In addition, this study allowed for the identification of several PTMs in proteins, such as phosphorylation, methylation, acetylation, glycosylation, and formylation. Finally, this study also allowed for the identification of eight new protein-coding genes and the extension of the ORFs for 34. In conclusion, whole proteomics and genomic studies are inextricable, the results of the former depend on an accurate genome annotation, and a genome cannot be only annotated in an automatic manner. Thus, the proteomics data obtained in this study have allowed for the correction of

annotation mistakes, the discovery of new genes, and experimental evidence of the existence of a large number of proteins that had to date been annotated as hypothetical.

All this new information, at the level of individual genes, is already available at Wikidata.org (searchable by the ID gene) and is going to be incorporated in the TriTrypDB database and the Leish-ESP website (<http://leish-esp.cbm.uam.es/>).

Supplementary Materials: The following are available online at www.mdpi.com/2073-4425/11/9/1036/s1, Table S1: Proteins detected in *L. infantum* promastigotes whose orthologues in *L. donovani* were found to be only expressed in the amastigote stage, Table S2: Translational machinery components (ribosomal proteins, translation factors and tRNA synthetases), Table S3: Proteins annotated as RNA binding proteins, Table S4: Identified proteins belonging to the pentose phosphate pathway, Table S5: Proteins of the mitochondrial electron transport chain, Table S6: Identified proteins belonging to the proteasome, Table S7: Identified proteins in the categories of chaperones and heat shock proteins, Table S8: Putative glycosomal proteins, Table S9: Identified proteins with predicted molecular weight higher than 200 kDa, Table S10: Identified proteins in *L. infantum* putatively located in the flagellum and/or involved in parasite motility, Table S11: Enzymes putatively involved on post-translational modifications, Table S12: Methylated proteins and amino acids found to be modified, Table S13: N-terminal acetylated proteins and the modified position, Table S14: Formylated proteins and the modified position, Table S15: Identification of novel ORFs based on peptide MS spectra, Table S16: N-terminal extended proteins based on the identification of peptides mapping upstream of the annotated ORFs, Table S17: Experimentally identified proteins in promastigotes of *L. infantum* JPCM5 strain, Table S18: Proteins identified for the first time in experimental proteomes of *Leishmania* promastigotes.

Author Contributions: Conceptualization, Á.S., E.M., A.R., A.M., B.A., and J.M.R.; methodology, Á.S., E.M., A.R., E.C., A.M., and S.G.-d.l.F.; formal analysis, Á.S., E.M., A.R., E.C., and S.G.-d.l.F.; data curation, A.S., E.M., S.G.-d.l.F., and J.M.R.; writing—original draft preparation, Á.S.; writing—review and editing, Á.S., A.M., B.A., and J.M.R.; funding acquisition, A.M., B.A., and J.M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by grants (to B.A. and J.M.R.) from Proyecto del Ministerio de Economía, Industria y Competitividad SAF2017-86965-R, and by the Network of Tropical Diseases Research RICET (RD16/0027/0008); both grants are co-funded with FEDER funds. A.S. was funded by a postdoctoral contract from the “Programa de Empleo Juvenil” of the Community of Madrid, Spain, within the European Youth Employment Initiative (YEI). A.M. was funded by project PRB3-ISCIII (supported by grant PT17/0019) of the PE I+D+i 2013–2016, funded by ISCIII and ERDF. The CBMSO receives institutional grants from the Fundación Ramón Areces and from the Fundación Banco de Santander.

Acknowledgments: We thank the Genomics and NGS Core Facility at the Centro de Biología Molecular Severo Ochoa (CBMSO, CSIC-UAM) for helping with the bioinformatics analysis. The CBMSO Proteomics Facility is a member of Proteored.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization (WHO). *Leishmaniasis: Disease, Epidemiology, Diagnosis, Detection and Surveillance, Vector Control, Access to Medicines and Information Resources*; WHO: Geneva, Switzerland, 2019.
2. Lukes, J.; Mauricio, I.L.; Schonian, G.; Dujardin, J.-C.; Soteriadou, K.; Dedet, J.-P.; Kuhls, K.; Tintaya, K.W.Q.; Jirku, M.; Chocholova, E.; et al. Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9375–9380.
3. Ivens, A.C.; Peacock, C.S.; Worthey, E.A.; Murphy, L.; Berriman, M.; Sisk, E.; Rajandream, M.; Adlem, E.; Anupama, A.; Apostolou, Z.; et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **2006**, *309*, 436–442.
4. Peacock, C.S.; Seeger, K.; Harris, D.; Murphy, L.; Ruiz, J.C.; Quail, M.A.; Peters, N.; Adlem, E.; Tivey, A.; Aslett, M.; et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat. Genet.* **2007**, *39*, 839–847.
5. Downing, T.; Imamura, H.; Decuypere, S.; Clark, T.G.; Coombs, G.H.; Cotton, J.A.; Hilley, J.D.; de Doncker, S.; Maes, I.; Mottram, J.C.; et al. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res.* **2011**, *21*, 2143–2156.

6. Gupta, A.K.; Srivastava, S.; Singh, A.; Singh, S. De novo whole-genome sequence and annotation of a Leishmania strain isolated from a case of post-kala-azar dermal leishmaniasis. *Genome Announc.* **2015**, *3*, 4–5.
7. Llanes, A.; Restrepo, C.M.; Vecchio, G. Del; Anguizola, F.J.; Lleonart, R. The genome of *Leishmania panamensis*: Insights into genomics of the L. (Viannia) subgenus. *Sci. Rep.* **2015**, *5*, 1–10.
8. Raymond, F.; Boisvert, S.; Roy, G.; Ritt, J.F.; Légaré, D.; Isnard, A.; Stanke, M.; Olivier, M.; Tremblay, M.J.; Papadopolou, B.; et al. Genome sequencing of the lizard parasite *Leishmania tarentolae* reveals loss of genes associated to the intracellular stage of human pathogenic species. *Nucleic Acids Res.* **2012**, *40*, 1131–1147.
9. Real, F.; Vidal, R.O.; Carazzolle, M.F.; Mondego, J.M.C.; Costa, G.G.L.; Herai, R.H.; Würtele, M.; De Carvalho, L.M.; E.; Ferreira, R.C.; Mortara, R.A.; et al. The genome sequence of leishmania (*Leishmania*) amazonensis: Functional annotation and extended analysis of gene models. *DNA Res.* **2013**, *20*, 567–581.
10. Forsdyke, D.R. *Evolutionary Bioinformatics*; Springer: New York, NY, USA, 2006; pp. 1–424.
11. Imamura, H.; Downing, T.; Van den Broeck, F.; Sanders, M.J.; Rijal, S.; Sundar, S.; Mannaert, A.; Vanaerschot, M.; Berg, M.; De Muylder, G.; et al. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. *Elife* **2016**, *5*, e12613.
12. Coughlan, S.; Mulhair, P.; Sanders, M.; Schonian, G.; Cotton, J.A.; Downing, T. The genome of *Leishmania adleri* from a mammalian host highlights chromosome fission in *Sauroleishmania*. *Sci. Rep.* **2017**, *7*, 1–13.
13. González-De La Fuente, S.; Peiró-Pastor, R.; Rastrojo, A.; Moreno, J.; Carrasco-Ramiro, F.; Requena, J.M.; Aguado, B. Resequencing of the *Leishmania infantum* (strain JPCM5) genome and de novo assembly into 36 contigs. *Sci. Rep.* **2017**, *7*, 1–10.
14. Alonso, G.; Rastrojo, A.; López-Pérez, S.; Requena, J.M.; Aguado, B. Resequencing and assembly of seven complex loci to improve the *Leishmania major* (Friedlin strain) reference genome. *Parasites Vectors* **2016**, *9*, 1–13.
15. González-De La Fuente, S.; Camacho, E.; Peiró-Pastor, R.; Rastrojo, A.; Carrasco-Ramiro, F.; Aguado, B.; Requena, J.M. Complete and de novo assembly of the *Leishmania braziliensis* (M2904) genome. *Mem. Inst. Oswaldo Cruz* **2019**, *114*, 1–6.
16. Requena, J.M.; Alcolea, P.J.; Alonso, A.; Larraga, V. Omics approaches for understanding gene expression in *Leishmania*: Clues for tackling leishmaniasis. In *Protozoan Parasitism: From Omics to Prevention and Control*; Pablos-Torró, L.M., Lorenzo-Morales, J., Eds.; Caister Academic Press: Norfolk, UK, 2018; pp. 77–112.
17. Capelli-Peixoto, J.; Mule, S.N.; Tano, F.T.; Palmisano, G.; Stolf, B.S. Proteomics and Leishmaniasis: Potential clinical applications. *Proteom. Clin. Appl.* **2019**, *13*, 1800136.
18. Nugent, P.G.; Karsani, S.A.; Wait, R.; Tempero, J.; Smith, D.F. Proteomic analysis of *Leishmania mexicana* differentiation. *Mol. Biochem. Parasitol.* **2004**, *136*, 51–62.
19. Leifso, K.; Cohen-Freue, G.; Dogra, N.; Murray, A.; McMaster, W.R. Genomic and proteomic expression analysis of *Leishmania* promastigote and amastigote life stages: The *Leishmania* genome is constitutively expressed. *Mol. Biochem. Parasitol.* **2007**, *152*, 35–46.
20. Rosenzweig, D.; Smith, D.; Oppendoes, F.; Stern, S.; Olafson, R.W.; Zilberstein, D. Retooling *Leishmania* metabolism: From sand fly gut to human macrophage. *FASEB J.* **2008**, *22*, 590–602.
21. Jardim, A.; Hardie, D.B.; Boitz, J.; Borchers, C.H. Proteomic profiling of *Leishmania donovani* promastigote subcellular organelles. *J. Proteome Res.* **2018**, *17*, 1194–1215.
22. Tasbihi, M.; Shekari, F.; Hajjaran, H.; Masoori, L.; Hadighi, R. Mitochondrial proteome profiling of *Leishmania tropica*. *Microb. Pathog.* **2019**, *133*, 103542.
23. Armengaud, J. Proteogenomics and systems biology: Quest for the ultimate missing parts. *Expert Rev. Proteom.* **2010**, *7*, 65–77.
24. Moreno, M.-L.; Escobar, J.; Izquierdo-Alvarez, A.; Gil, A.; Perez, S.; Pereda, J.; Zapico, I.; Vento, M.; Sabater, L.; Marina, A.; et al. Disulfide stress: A novel type of oxidative stress in acute pancreatitis. *Free Radic. Biol. Med.* **2014**, *70*, 265–277.
25. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **1996**, *68*, 850–858.
26. Torres, L.L.; Cantero, A.; del Valle, M.; Marina, A.; Lopez-Gallego, F.; Guisan, J.M.; Berenguer, J.; Hidalgo, A. Engineering the substrate specificity of a thermophilic penicillin acylase from *Thermus thermophilus*. *Appl. Environ. Microbiol.* **2013**, *79*, 1555–1562.
27. Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G.A.; Ma, B. PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteom.* **2012**, *11*, M111.010587.

28. Yonghua, H.; Bin, M.; Zaizhong, Z. SPIDER: Software for protein identification from sequence tags with de novo sequencing error. In Proceedings of the IEEE Computational Systems Bioinformatics Conference, Stanford, CA, USA, 19 August 2004.
29. Han, X.; He, L.; Xin, L.; Shan, B.; Ma, B. PeaksPTM: Mass spectrometry-based identification of peptides with unspecified modifications. *J. Proteome Res.* **2011**, *10*, 2930–2936.
30. McNicoll, F.; Drummelsmith, J.; Müller, M.; Madore, É.; Boilard, N.; Ouellette, M.; Papadopolou, B. A combined proteomic and transcriptomic approach to the study of stage differentiation in *Leishmania infantum*. *Proteomics* **2006**, *6*, 3567–3581.
31. Foucher, A.L.; Papadopolou, B.; Ouellette, M. Prefractionation by digitonin extraction increases representation of the cytosolic and intracellular proteome of *Leishmania infantum*. *J. Proteome Res.* **2006**, *5*, 1741–1750.
32. Brotherton, M.-C.; Racine, G.; Foucher, A.L.; Drummelsmith, J.; Papadopolou, B.; Ouellette, M. Analysis of stage-specific expression of basic proteins in *Leishmania infantum*. *J. Proteome Res.* **2010**, *9*, 3842–3853.
33. Alcolea, P.J.; Alonso, A.; Larraga, V. Proteome profiling of *Leishmania infantum* promastigotes. *J. Eukaryot. Microbiol.* **2011**, *58*, 352–358.
34. Braga, M.S.; Neves, L.X.; Campos, J.M.; Roatt, B.M.; de Oliveira Aguiar Soares, R.D.; Braga, S.L.; de Melo Resende, D.; Reis, A.B.; Castro-Borges, W. Shotgun proteomics to unravel the complexity of the *Leishmania infantum* exoproteome and the relative abundance of its constituents. *Mol. Biochem. Parasitol.* **2014**, *195*, 43–53.
35. Santarém, N.; Racine, G.; Silvestre, R.; Cordeiro-da-Silva, A.; Ouellette, M. Exoproteome dynamics in *Leishmania infantum*. *J. Proteom.* **2013**, *84*, 106–118.
36. Pawar, H.; Renuse, S.; Khobragade, S.N.; Chavan, S.; Sathe, G.; Kumar, P.; Mahale, K.N.; Gore, K.; Kulkarni, A.; Dixit, T.; et al. Neglected tropical diseases and omics science: Proteogenomics analysis of the promastigote stage of leishmania major parasite. *Omi. A J. Integr. Biol.* **2014**, *18*, 499–512.
37. Nirujogi, R.S.; Pawar, H.; Renuse, S.; Kumar, P.; Chavan, S.; Sathe, G.; Sharma, J.; Khobragade, S.; Pande, J.; Modak, B.; et al. Moving from unsequenced to sequenced genome: Reanalysis of the proteome of *Leishmania donovani*. *J. Proteom.* **2014**, *97*, 48–61.
38. Pawar, H.; Sahasrabudhe, N.A.; Renuse, S.; Keerthikumar, S.; Sharma, J.; Kumar, G.S.S.; Venugopal, A.; Sekhar, N.R.; Kelkar, D.S.; Nemade, H.; et al. A proteogenomic approach to map the proteome of an unsequenced pathogen—*Leishmania donovani*. *Proteomics* **2012**, *12*, 832–844.
39. Ramrath, D.J.F.; Niemann, M.; Leibundgut, M.; Bieri, P.; Prange, C.; Horn, E.K.; Leitner, A.; Boehringer, D.; Schneider, A.; Ban, N. Evolutionary shift toward protein-based architecture in trypanosomal mitochondrial ribosomes. *Science* **2018**, *362*, eaau7735.
40. Clayton, C.; Shapira, M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol. Biochem. Parasitol.* **2007**, *156*, 93–101.
41. Requena, J.M. Lights and shadows on gene organization and regulation of gene expression in Leishmania. *Front. Biosci. Landmark Ed.* **2011**, *16*, 2069–2085.
42. Beckmann, B.M.; Castello, A.; Medenbach, J. The expanding universe of ribonucleoproteins: Of novel RNA-binding proteins and unconventional interactions. *Pflugers Arch. Eur. J. Physiol.* **2016**, *468*, 1029–1040.
43. Kramer, S.; Carrington, M. Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. *Trends Parasitol.* **2011**, *27*, 23–30.
44. De Gaudenzi, J.G.; Frasch, A.C.; Clayton, C. RNA-binding domain proteins in *Kinetoplastids*: A comparative analysis. *Eukaryot. Cell* **2014**, *4*, 2106–2114.
45. Nandan, D.; Thomas, S.A.; Nguyen, A.; Moon, K.M.; Foster, L.J.; Reiner, N.E. Comprehensive identification of mRNABinding proteins of *Leishmania donovani* by interactome capture. *PLoS ONE* **2017**, *12*, e0170068.
46. Folgueira, C.; Martínez-Bonet, M.; Requena, J.M. The *Leishmania infantum* PUF proteins are targets of the humoral response during visceral leishmaniasis. *BMC Res. Notes* **2010**, *3*, 13.
47. Subramanian, A.; Jhawar, J.; Sarkar, R.R. Dissecting *Leishmania infantum* energy metabolism—A systems perspective. *PLoS ONE* **2015**, *10*, e0137976.
48. Colasante, C.; Voncken, F.; Manful, T.; Ruppert, T.; Tielens, A.G.M.; van Hellemond, J.J.; Clayton, C. Proteins and lipids of glycosomal membranes from *Leishmania tarentolae* and *Trypanosoma brucei*. *F1000Research* **2013**, *2*, 1–15.
49. Camacho, E.; Rastrojo, A.; Sanchiz, Á.; González-de la Fuente, S.; Aguado, B.; Requena, J.M. Leishmania mitochondrial genomes: Maxicircle structure and heterogeneity of minicircles. *Genes (Basel)*. **2019**, *10*, 758–777.

50. Paugam, A.; Bulteau, A.L.; Dupouy-Camet, J.; Creuzet, C.; Friguet, B. Characterization and role of protozoan parasite proteasomes. *Trends Parasitol.* **2003**, *19*, 55–59.
51. Silva-Jardim, I.; Fátima Horta, M.; Ramalho-Pinto, F.J. The *Leishmania chagasi* proteasome: Role in promastigotes growth and amastigotes survival within murine macrophages. *Acta Trop.* **2004**, *91*, 121–130.
52. Khare, S.; Nagle, A.S.; Biggart, A.; Lai, Y.H.; Liang, F.; Davis, L.C.; Barnes, S.W.; Mathison, C.J.N.; Myburgh, E.; Gao, M.-Y.; et al. Proteasome inhibition for treatment of leishmaniasis, Chagas disease and sleeping sickness. *Nature* **2016**, *537*, 229–233.
53. Balchin, D.; Hayer-Hartl, M.; Hartl, F.U. In vivo aspects of protein folding and quality control. *Science* **2016**, *353*, aac4354.
54. Requena, J.M.; Montalvo, A.M.; Fraga, J. Molecular chaperones of Leishmania: Central players in many stress-related and -unrelated physiological processes. *Biomed Res. Int.* **2015**, *2015*, 1–21.
55. Bauer, S.; Morris, M.T. Glycosome biogenesis in trypanosomes and the de novo dilemma. *PLoS Negl. Trop. Dis.* **2017**, *11*, e0005333.
56. Pilar, A.V.C.; Strasser, R.; McLean, J.; Quinn, E.; Cyr, N.; Hojjat, H.; Kottarampatel, A.H.; Jardim, A. Analysis of the Leishmania peroxin 7 interactions with peroxin 5, peroxin 14 and PTS2 ligands. *Biochem. J.* **2014**, *460*, 273–282.
57. Jamdhade, M.D.; Pawar, H.; Chavan, S.; Sathe, G.; Umasankar, P.K.; Mahale, K.N.; Dixit, T.; Madugundu, A.K.; Prasad, T.S.K.; Gowda, H.; et al. Comprehensive Proteomics analysis of glycosomes from *Leishmania donovani*. *OMICS A J. Integr. Biol.* **2015**, *19*, 157–170.
58. Boitz, J.M.; Ullman, B.; Jardim, A.; Carter, N.S. Purine salvage in Leishmania: Complex or simple by design? *Trends Parasitol.* **2012**, *28*, 345–352.
59. Tiwari, K.; Dubey, V.K. Fresh insights into the pyrimidine metabolism in the trypanosomatids. *Parasites Vectors* **2018**, *11*, 1–15.
60. Pérez-Cabezas, B.; Santarém, N.; Cecílio, P.; Silva, C.; Silvestre, R.; AM Catita, J.; Cordeiro da Silva, A. More than just exosomes: Distinct *Leishmania infantum* extracellular products potentiate the establishment of infection. *J. Extracell. Vesicles* **2019**, *8*, 1541708.
61. Atayde, V.D.; Aslan, H.; Townsend, S.; Hassani, K.; Kamhawi, S.; Olivier, M. Exosome secretion by the parasitic protozoan Leishmania within the sand fly midgut. *Cell Rep.* **2015**, *13*, 957–967.
62. Avilán, L.; Gualdrón-López, M.; Quiñones, W.; González-González, L.; Hannaert, V.; Michels, P.A.M.; Concepción, J.-L. Enolase: A key player in the metabolism and a probable virulence factor of *Trypanosomatid* parasites—Perspectives for its use as a therapeutic target. *Enzyme Res.* **2011**, *2011*, 932549.
63. Brotherton, M.C.; Racine, G.; Ouameur, A.A.; Leprohon, P.; Papadopolou, B.; Ouellette, M. Analysis of membrane-enriched and high molecular weight proteins in *Leishmania infantum* promastigotes and axenic amastigotes. *J. Proteome Res.* **2012**, *11*, 3974–3985.
64. Beneke, T.; Demay, F.; Hookway, E.; Ashman, N.; Jeffery, H.; Smith, J.; Valli, J.; Becvar, T.; Myskova, J.; Lestonova, T.; et al. Genetic dissection of a *Leishmania flagellar* proteome demonstrates requirement for directional motility in sand fly infections. *PLoS Pathog.* **2019**, *15*, e1007828.
65. Morales, M.A.; Watanabe, R.; Dacher, M.; Chafey, P.; Osorio, Y.; Fortéa, J.; Scott, D.A.; Beverley, S.M.; Ommen, G.; Clos, J.; et al. Phosphoproteome dynamics reveal heat-shock protein complexes specific to the *Leishmania donovani* infectious stage. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8381–8386.
66. Rosenzweig, D.; Smith, D.; Myler, P.J.; Olafson, R.W.; Zilberstein, D. Post-translational modification of cellular proteins during *Leishmania donovani* differentiation. *Proteomics* **2008**, *8*, 1843–1850.
67. Santos, A.L.S.; Branquinho, M.H.; D’Avila-Levy, C.M.; Kneipp, L.; Sodré, C.L. *Proteins and Proteomics of Leishmania and Trypanosoma*; Springer: New York, NY, USA, 2014; ISBN 978-94-007-7304-2.
68. Tsigankov, P.; Gherardini, P.F.; Helmer-Citterich, M.; Späth, G.F.; Zilberstein, D. Phosphoproteomic analysis of differentiating Leishmania parasites reveals a unique stage-specific phosphorylation motif. *J. Proteome Res.* **2013**, *12*, 3405–3412.
69. Sprung, R.; Chen, Y.; Zhang, K.; Cheng, D.; Zhang, T.; Peng, J.; Zhao, Y. Identification and validation of eukaryotic aspartate and glutamate methylation in proteins. *J. Proteome Res.* **2008**, *7*, 1001–1006.
70. Alonso, V.L.; Serra, E.C. Lysine acetylation: Elucidating the components of an emerging global signaling pathway in trypanosomes. *J. Biomed. Biotechnol.* **2012**, *2012*.
71. Thomas, S.; Green, A.; Sturm, N.R.; Campbell, D.A.; Myler, P.J. Histone acetylations mark origins of polycistronic transcription in Leishmania major. *BMC Genom.* **2009**, *10*, 152.

72. Respuela, P.; Ferella, M.; Rada-Iglesias, A.; Aslund, L. Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *Trypanosoma cruzi*. *J. Biol. Chem.* **2008**, *283*, 15884–15892.
73. Cuervo, P.; Domont, G.B.; De Jesus, J.B. Proteomics of trypanosomatids of human medical importance. *J. Proteom.* **2010**, *73*, 845–867.
74. Linster, E.; Wirtz, M. N-terminal acetylation: An essential protein modification emerges as an important regulator of stress responses. *J. Exp. Bot.* **2018**, *69*, 4555–4568.
75. Gupta, N.; Tanner, S.; Jaitly, N.; Adkins, J.N.; Lipton, M.; Edwards, R.; Romine, M.; Osterman, A.; Bafna, V.; Smith, R.D.; et al. Whole proteome analysis of post-translational modifications: Applications of mass-spectrometry for proteogenomic annotation. *Genome Res.* **2007**, *17*, 1362–1377.
76. Manzano-Román, R.; Fuentes, M. Relevance and proteomics challenge of functional posttranslational modifications in *Kinetoplastid* parasites. *J. Proteom.* **2020**, *220*, 103762.
77. Eldeeb, M.A.; Fahlman, R.P.; Esmaili, M.; Fon, E.A. Formylation of eukaryotic cytoplasmic proteins: Linking stress to degradation. *Trends Biochem. Sci.* **2019**, *44*, 181–183.
78. Wiśniewski, J.R.; Zougman, A.; Mann, M. N ϵ -Formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. *Nucleic Acids Res.* **2008**, *36*, 570–577.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).