Peer-reviewed version available at *Genes* **2020**, *11*, 1036; doi:10.3390/genes11091036

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

* Correspondence: baguado@cbm.csic.es (B.A.); jmrequena@cbm.csic.es (J.M.R)

Abstract: *Leishmania infantum* is causative of visceral leishmaniasis (kala-azar), the most severe form of leishmaniasis, lethal if untreated. Few years ago, re-sequencing and *de novo* assembling of the *L. infantum* (strain JPCM5) genome was accomplished, and now we aimed to describe and characterize the experimental proteome of this species. In this work, we have performed a proteomic analysis from axenic cultured promastigotes and carried out a detailed comparison with other *Leishmania* experimental proteomes published to date. We identified 2,352 proteins based on the search of mass spectrometry data against a database built from the six-frame translated genome sequence of *L. infantum*. We have detected many proteins belonging to organelles such as glycosomes, mitochondria or flagellum, as well as many metabolic enzymes, and a large number of putative RNA binding proteins and molecular chaperones. Moreover, we listed the proteins presenting post-translational modifications, such as phosphorylations, acetylations and methylations, among others. On the other hand, the identification of peptides mapping to genomic regions previously annotated as non-coding has allowed to correct annotations, leading to 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 *L. infantum* reference genome annotation.

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 responsible for a complex of diseases named leishmaniasis, second most common cause of mortality among tropical infectious diseases, after malaria [1]. Some species of *Leishmania* are human pathogens, causative of clinical manifestations as cutaneous (CL), mucosal (MCL) or visceral (VL) leishmaniasis. Old World species *Leishmania infantum* and *Leishmania donovani* cause VL or kala azar, 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 developing extracellularly in the gut of their sand-fly vector. 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, *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 drafting the genomes for many other *Leishmania* species [5–12] and improving the assemblies of the first sequenced [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 mass spectrometry (MS)-based experimental approaches have become the method of choice for the analysis of complex protein mixtures, such as cells, tissues or even whole organisms. In particular, a variety of proteomic technologies are being used to study diverse aspects of *Leishmania* biology, as parasite development, virulence and drug resistance [17]. Thus, different proteomic approaches have been used to determine differential patterns of protein expression between the promastigote and amastigote stages in *L. mexicana* [18], *L. infantum* [19], and *L. donovani* [20], among others. Other studies were 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 activation of specific pathways in a given growing condition, improving the knowledge on protein interactions in complex networks.

Here, we present a wide and detailed proteome of the *L. infantum* strain JPCM5, 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 the identification of PTM (phosphorylation, acetylation, methylation, formylation and glycosylation) at specific protein sites, which 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 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 strain JPCM5 parasites were grown at 26^{0}C in RPMI 1640 media supplemented with 15% of heat inactivated foetal calf serum (Biowest SAS, Nuaillé, France). Promastigote cultures was initiated at 1 x 10^6 parasites/ml and harvested at mid-logarithmic phase (1-2 x 10^7 parasites/ml). Around 1-2 x 10^8 parasites were collected and washed twice with PBS; finally, parasites were suspended by pipetting in 300 µl of RIPA buffer (Thermo Scientific Pierce) in the presence of EDTA-free Easy Pack Protease inhibitor (*Roche*). After 6 cycles (30 s pulse/30 s pause) of sonication in bath at 4^{0}C, samples were incubated for 90 min at 4^{0}C and, afterwards, protein lysates were centrifuged at 14000g 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 in-gel digestion of proteins, samples were mixed with an equal volume of 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 x 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 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) 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, 10mM $CaCl_2$, pH 8 with 60 ng/µl trypsin or chymotrypsin at 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% TFA. Whole supernatants were dried down and then desalted onto OMIX Pipette tips C18 (Agilent Technologies) until the MS analysis.

Additionally, in-solution digestion was performed as described elsewhere [26]. After denaturation of proteins with 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 ^{0}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) or chymotrypsin (Roche) 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 Pipette tips C18 (Agilent Technologies) until 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 Scientific). The peptides were concentrated (on-line) by reverse phase chromatography using a 0.1mm × 20 mm C18 RP precolumn (Thermo Scientific), and then separated using a 0.075mm x 250 mm C18 RP column (Thermo 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 45min, 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, 80% acetonitrile in water). ESI ionization was done using a Nano-bore emitters 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), normalized collision energy of 35%, and dynamic exclusion applied during 60 seconds periods.

2.4. Data Analysis

Peptide identification from raw data was carried out using PEAKS Studio X search engine (Bioinformatics Solutions Inc.). A custom Python script was used to create a database comprising all possible open reading frames (ORF) coding for protein sequences ≥ 20 amino acids existing in any of the six-frames in the *L. infantum* strain JPCM5 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, and tolerances of 20 ppm for precursor ions and 0.6 Da for MS/MS fragment ions and the searches were performed allowing optional Met oxidation and Cys carbamidomethylation. False discovery rates (FDR) for peptide spectrum matches (PSM) was limited to 0.01 or lower. Those proteins having been 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 either miss-annotations or novel proteins. Also, Python scripts were used to ascribe post-translational modifications to particular protein entries.

Functional categories and enzymatic pathways using DAVID program (Functional Annotation Tool, DAVID Bioinformatics Resources 6.8) and 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 re-sequenced genome, recently published [13], was used. However, we did not restrict the search of peptide spectra on currently annotated protein-coding genes, but 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 Fig. 1. From the MS/MS data, only those associated with peptides longer than 7 amino acids were considered for protein identification. Among the identified proteins, 2,344 proteins matched with previously annotated proteins [13]. However, eight novel proteins were uncovered, legitimating the search strategy. Moreover, some ORFs had to be extended to accommodate MS identified peptides (see below for further details about these findings). Most of the proteins, 70.5% (1,659 out of 2,352), 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, 3,482 out of 8,590 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.

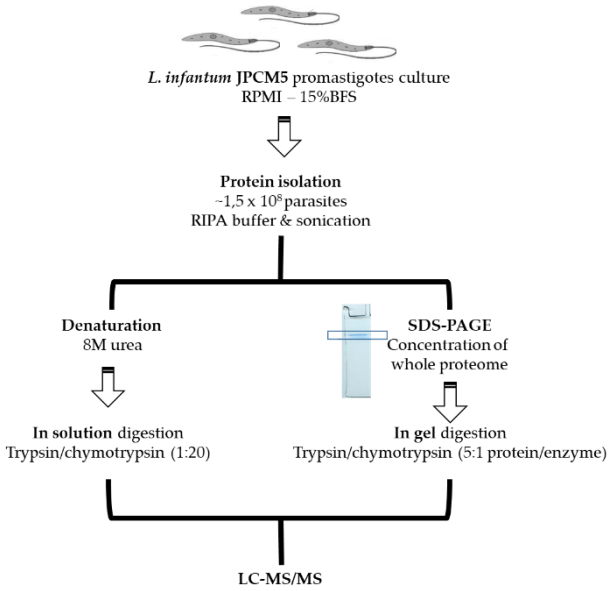

Figure 1. Workflow for protein extraction and proteomic analyses of L. infantum promastigotes. The experimental MS data were searched against UniProt protein database and a database consisting of all possible polypeptides encoded in the six-frames of the L. infantum genome (based on v2/2018; www.leish-esp.cbm.uam.es;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

The first comprehensive study aimed to characterize the L. infantum proteome, was carried out by the Papadopoulou's group [30]. Using two-dimensional (2-D) gel electrophoresis, these authors visualized 2261 protein spots in promastigote samples and 2273 spots in the 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 to Leifso and co-workers the identification of 91 L. infantum proteins [19]. An enrichment for basic proteins using the technique of free-flow electrophoresis prior to separation by 2-D 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 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 is providing 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. Rosenzweig and collaborators, in 2008, reported the identification of 1,713 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 1,218 proteins were common (orthologs) in both studies (Figure 2). We failed to identify 207 proteins of those reported in L. donovani, whereas we found 1,130 proteins that are absent from the L. donovani proteome reported by Rosenzweig et al [20]

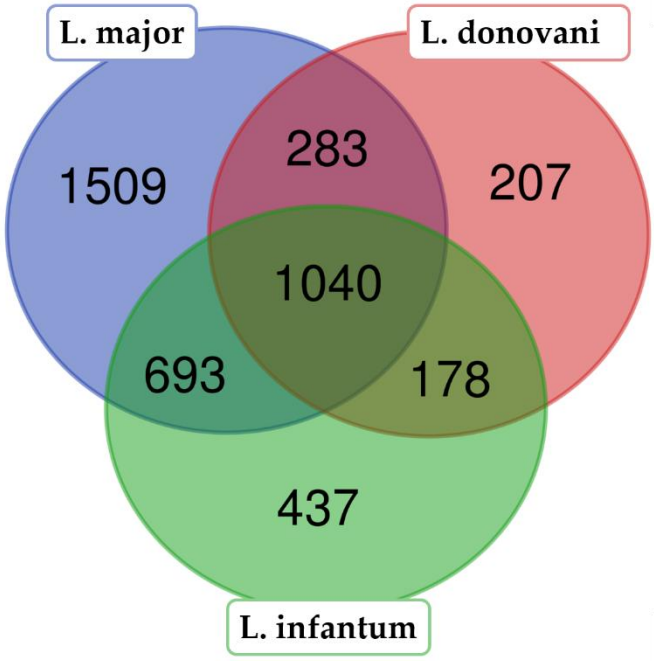

Figure 2. Comparison (Venn diagram) between the identified proteins in this work (*L. infantum,* in green*)* and those identified in two previous studies [20,36] performing proteomic analysis in *L. donovani* (in red) and *L. major* (in blue). Venn diagram was created by the tool available at bioinformatics.psb.ugent.be/webtools/.

More recently, Pandey and coworkers reported the identification of 3,386 different proteins in L. donovani promastigote and amastigote stages [37, 38]. 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, indicating that these proteins are being expressed in the promastigotes stage too, at least in L. infantum (see Supplementary file, Table S1). Most of them are 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 3,613 proteins were identified. These authors followed a proteogenomic approach, as we did in this study, consisting in searching the mass spectra against a six-frame translated database generated from the complete genome sequence. An orthology-based comparison indicated that the L. major promastigote proteome and the L. infantum proteome of this study share 1,733 proteins (Figure 2). Moreover, considering the 1,792 proteins identified in the L. major proteome, but no 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 4,140 (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 having RNA binding properties deserve special attention, since gene expression in Leishmania, and related trypanosomatids, is being controlled essentially 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]. RNA-binding proteins of the Pumilio family (aka PUF proteins) are especially numerous (11 members) in Leishmania [46]. In this study, we have identified 6 out of 11 PUF proteins, which 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 for retrieving their gene IDs).

3.3. Metabolic enzymes and pathways

Going deeper, we have performed an in-silico pathway reconstruction using the detected proteins in the L. infantum promastigote proteome. By the KEGG database resource (http://www.genome.jp/kegg/) accessed via 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), tricarboxylic acid (TCA) cycle and pentose phosphate cycle (Supplementary file, Table S4), 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).

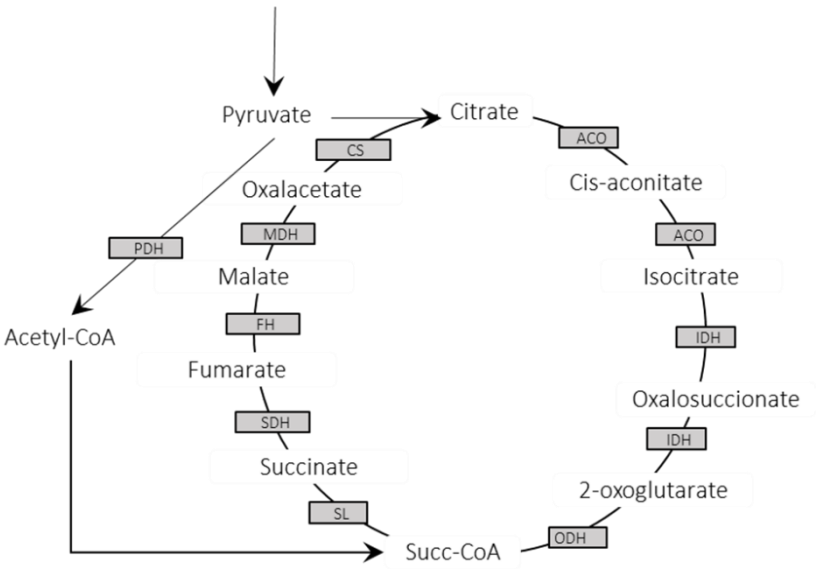

Figure 3. Detected enzymes in *L. infantum* JPCM5 composing the complete TCA cycle (TCA). PDH: pyruvate dehydrogenase; ACO: aconitase; IDH: isocitrate dehydrogenase; ODH: 2-oxoglutarate dehydrogenase; SL: succinyl-CoA ligase; SDH: succinyl dehydrogenase; FH: fumarate hydratase; MDH: malate dehydrogenase; 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], and 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 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 maxicircle kDNA [49], 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 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 alpha 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 beta subunit
LINF_290025700	4	Dihydrolipoamide dehydrogenase
LINF_310034500	2	Dihydrolipoamide dehydrogenase
LINF_320040600	37	Dihydrolipoamide dehydrogenase
LINF_350005300LINF_350005400	43	Pyruvate kinase
LINF_360030600	43	Glyceraldehyde 3-phosphate dehydrogenase, cytosolic
LINF_360034400	19	Dihydrolipoamide acetyltransferase precursor

An active energy metabolism requires enzymes involved in redox homeostasis. Several of these enzymes have been identified in the L. infantum proteome, being likely abundant, as judged by the large number of unique peptides that mapped to them. The detected proteins were typaredoxin (LINF_150019000, with 31 unique peptides), peroxidoxin (LINF_230005400, 31 peptides), cyclophilin (LINF_060006300, 20 peptides), iron superoxide dismutases (LINF_080007900 and LINF_320024000, with 14 and 18 peptides, respectively) and elongation factors 1β (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 KEGG database, the complete compendium of proteasomal proteins would have been 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 have 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 as they are involved in the protein transport across membranes and protein refolding inside the mitochondria. The mitochondrial proteome has been recently analyzed in L. tropica [22]. Taking advantage of that study, we are listing in Table 3 those HSPs identified in the L. infantum proteome that are candidate to be 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	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 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) have been 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 1,355 proteins. In our

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 August 2020 doi:10.20944/preprints202008.0281.v1

Peer-reviewed version available at Genes 2020, 11, 1036; doi:10.3390/genes11091036

study, orthologs to 853 of those putative L. donovani glycosomal proteins were found; these are listed in the Supplementary file (Table S8).

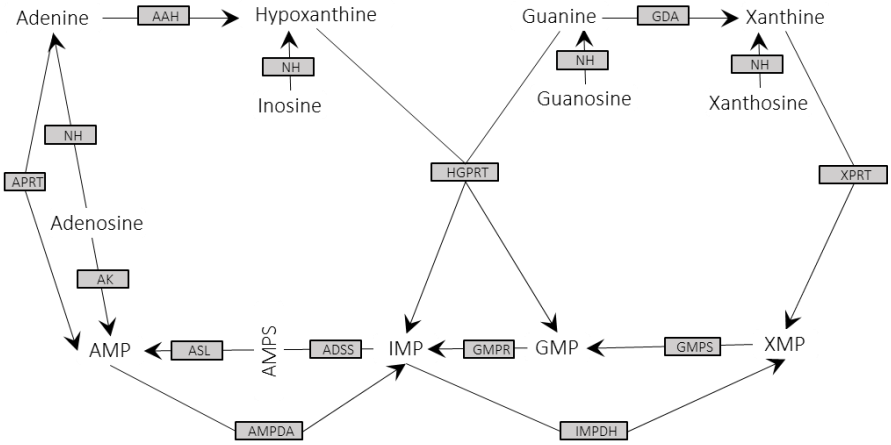

Figure 4. Enzymes from the purine salvage cycles identified in the L. infantum experimental proteome. All the enzymes (grey squares) that are required to complete the pathway were identified in this study. APRT: adenine phosphoribosyltransferase (LINF_130016900); NH: nucleoside hydrolase (LINF_180021400); AK: adenosine kinase (LINF_300014400); AAH: adenine aminohydrolase (LINF 350026800); ASL: adenylsuccinate 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); XPRT: xanthine phosphoribosyltransferase (LINF 210015000). The cycle was depicted according to Boitz et al [58].

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

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 have been found particularly relevant for infection establishment, being parasites and exosomes 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 have been 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, 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-alpha	Exosome marker [61]
LINF_190020600	Cysteine peptidase A (CPA)	Virulence factor [61]
LINF_200018000	Calpain-like cysteine peptidase	Virulence factor [61]
LINF_280035000 - LINF_280036000	HSP70	Exosome marker [61]
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 high molecular weight (HMW) represent a challenge for mass spectrometry-based assays, being 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, it is remarkable the identification of a calpain-like cysteine peptidase, with an estimated MW around 700 kDa (LINF_270010200), that was identified by 124 unique peptides, covering 24% of the amino acid sequence.

Flagellum is a characteristic organelle of Leishmania that confers motility to the parasite in the promastigote stage, in 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 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

PTMs of proteins influence their activity, structure, turnover, localization and their 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 accounting for 10 phosphorylated, 144 methylated, 192 acetylated, 28 formylated and 3 glycosylated proteins.

Phosphorylation of a 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, alpha 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 was 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	Alpha 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 serine (S) and threonine (T), but in some proteins, phosphorylation on tyrosine (Y) residues was also detected. Rosenzweig et al [66], in L. donovani, identified 16 phosphorylated proteins, 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 is not unexpected as phosphorylation is a dynamic modification and the numbers of phosphorylated proteins identified in both studies are low. Although abundance of phosphorylation events is probed to be much lower in tyrosine residues [67], it is remarkable that phosphorylated tyrosines were found in alpha tubulins, the rRNA biogenesis protein and in GSK-3 (Table 6). Some of these phosphorylated proteins were also 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 the enzymes implicated on regulation of phosphorylation/dephosphorilation 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 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 in beta-tubulin a methylated-H was found. 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. Examples of highly methylated proteins identified in this study are alpha and beta 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. Lysine (K) acetylation has been described as a reversible enzymatic reaction that regulates protein function, being particularly relevant in chromatin compaction by acetylation of histones [70]. Interestingly, accumulation of acetylated histones has been observed at the polycistronic transcription initiation sites in L. major and T. cruzi [71,72]. However, in the L. infantum proteomics data, peptides bearing acetylated K belonging to histones were not found, as occurred in the L. donovani proteome [66], suggesting a relatively low proportion of acetylated histones in the bulk of total cellular histones. Examples of proteins detected as acetylated in this study are: beta tubulins (LINF_330015100, LINF_330015200, LINF_210028500; modified at K297), guanylate kinase (LINF_330018400, K3), subunit beta of ATP synthase (LINF_250018000; K511) and a calpain-like cysteine peptidase (LINF_140014400; K74).

In addition, acetylation at the N-terminal end of proteins may occur either co- or post-translationally, being 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 it is carried out by N-terminal acetyltransferases (Nat). In the L. infantum promastigote proteome, we have 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 asparagine (N) or glutamine (Q) residues (Figure 5A also shows the other more frequent amino acids located at the second position). Acetylation reaction after iM removal has been mainly found on serine (S) (55% of the cases) and alanine (A) (in 31%) residues (Figure 5B). These two amino acids, and also 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 miss-annotated genes (whose new coordinates are indicated in Supplementary file, Table S13). One example is illustrated in Figure 5C; the coding sequence of gene LINF_010008200 (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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 August 2020 doi:10.20944/preprints202008.0281.v1

Peer-reviewed version available at Genes 2020, 11, 1036; doi:10.3390/genes11091036

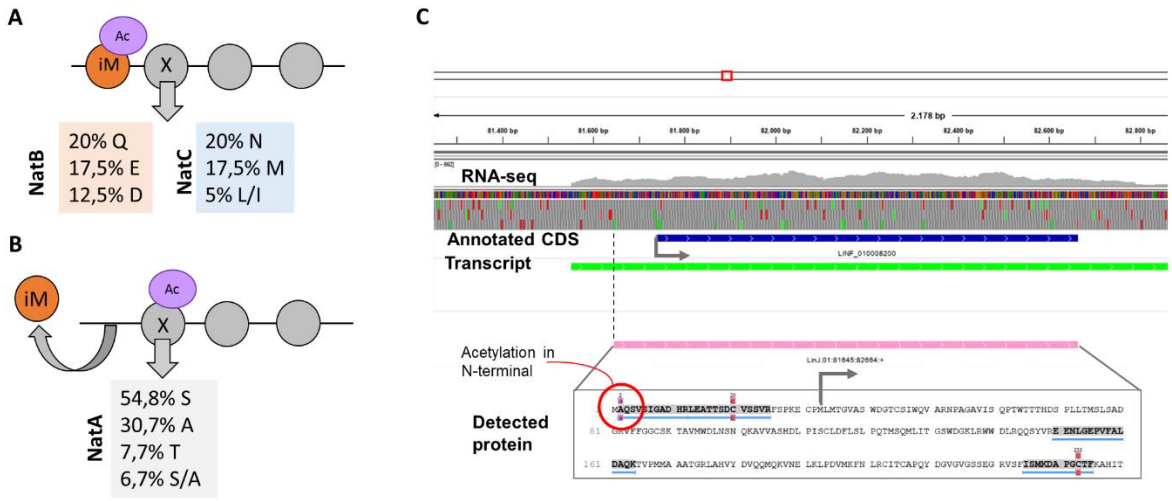

Figure 5. Features of the N-terminal acetylated proteins identified in this study. A) Frequencies (percentages) of amino acids found next to the acetylated initial methionine (iM), and the putative enzymes responsible for the acetylation. B) Percentages of amino acids found to be acetylated after cleavage of the iM, and the putative enzyme involved in the reaction. C) 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). Image in C was generated using the Integrative Genome Viewer (IGV).

Glycosylation also plays a relevant role in protein maturation, as well as in signal transduction mechanisms [76]. In this L. infantum proteomic study, we have detected hexosylations or N-acetylhexosamine addition in three proteins; these modifications are consistent with N-linked glycosylation events at asparagine (N) residues. These PTM-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). Rosenzweig et al [66] found 13 glycosylated proteins, only one in asparagine and the rest in serine and threonine residues.

Formylation was not 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, a total of eight proteins were detected as formylated, mainly at leucine (in β-tubulin and typaredoxin), glycine (in a hypothetical protein; LINF_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 researches on protein formylations in Leishmania, and other organisms will provide insight into the physiological significance of this kind of PTM.

3.9. Proteogenomics

After a genome assembly, dedicated programs conduct automatic annotation of ORFs. However, this annotation is not definitive and a continuous effort of curation is needed. Transcriptomic analysis enables to obtain complete gene model annotation, including untranslated regions that are key to understand post-transcriptional regulation mechanisms. In addition, a proteogenomic analysis, as that reported here, represent a powerful and useful approach for the identification of non-annotated genes, correcting miss-annotations or validating 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 strain JPCM5 [13]. The majority of the identified peptides fitted well in current gene annotation (available at

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 August 2020 doi:10.20944/preprints202008.0281.v1

Peer-reviewed version available at Genes 2020, 11, 1036; doi:10.3390/genes11091036

TriTryDB.org and http://leish-esp.cbm.uam.es/). Nevertheless, some peptides mapped to non-annotated coding-regions in the L. infantum (JPCM5) genome, resulting in 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 its identification.

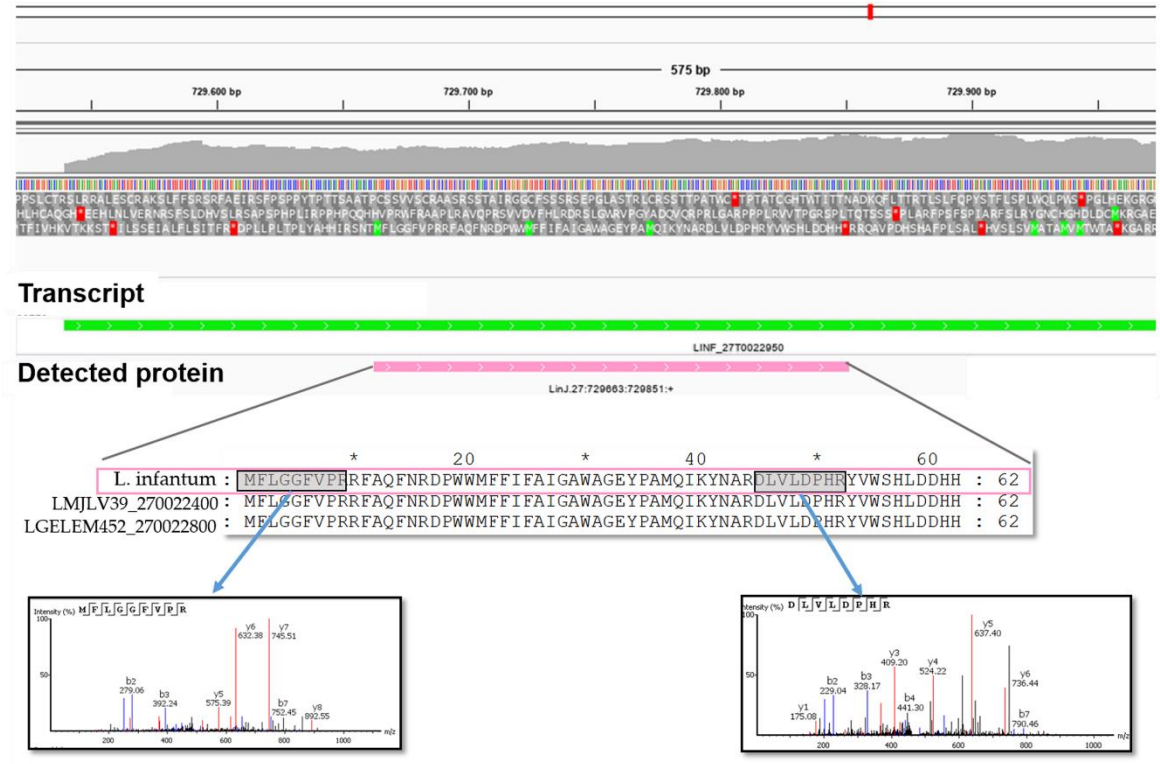

Figure 6. Identification of a novel protein based on the mapping of two experimentally detected-peptides in a region of L. infantum JPCM5 chromosome 27, currently lacking of annotated ORF. The new CDS, named LINF_ 270022950 (pink box), fit well within a predicted transcript (LINF_27T0022950; green box). Interestingly, the predicted amino acid sequence is 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 CDS. This led to adding N-terminal extensions to 34 annotated proteins and establishing 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 is being an active area of research. Here, we reported a proteomic analysis of L. infantum (strain JPCM5) promastigote stage, being 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 ORF, larger than 20 triplets, existing in all-six frames from the L. infantum genome sequence. As a result, we identified 2,352 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 have shown a significant level of similarity regarding the type of proteins detected. Nevertheless, this proteomic study has shown 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 has allowed the identification of several PTMs in proteins, such as phosphorylation, methylation, acetylation, glycosylation and formylation. Finally, this study also served for the identification of eight new protein-coding genes and the extension of the ORFs for 34 ones. 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 automatics manner. Thus, the proteomics data obtained in this study have allowed to correct annotation mistakes, to discover new genes and to show experimental evidence of the existence of a large number of proteins, annotated to date 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 web site (http://leish-esp.cbm.uam.es/).

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1. Supplementary file (Excel file containing 18 tables).

Author Contributions: Conceptualization, A.S., E.M., A.R., A.M., B.A. and J.M.R.; methodology, A.S., E.M., A.R., E.C., A.M., S.G.F.; formal analysis, A.S., E.M., A.R., E.C., S.G.F.; data curation, A.S., E.M., S.G.F., J.M.R.; writing—original draft preparation, A.S.; writing—review and editing, A.S., A.M., B.A., J.M.R.; funding acquisition, A.M., B.A., 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.; 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. U. S. A.* **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 (80-.).* **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.; Papadopoulou, 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. *Evol. Bioinforma.* **2006**, 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 and 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 Parasit. From Omi. to Prev. Control*; Pablos-Torró, L.M., Lorenzo-Morales, J., Eds.; 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. *PROTEOMICS – 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.; Opperdoes, 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. Proteomics* **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. Proteomics* **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; 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.; Papadopoulou, 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.; Papadopoulou, 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.; Papadopoulou, 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. Proteomics* **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. Proteomics* **2014**, *97*, 48–61.
38. Pawar, H.; Sahasrabuddhe, 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*.
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*, 2069–85.
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*, 1–16.
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*.
47. Subramanian, A.; Jhawar, J.; Sarkar, R.R. Dissecting leishmania infantum energy metabolism - A systems perspective. *PLoS One* **2015**, *10*, 1–34.
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*, 1–13.
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. *Omi. 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 and Vectors* **2018**, *11*, 1–15.
60. Pérez-Cabezas, B.; Santarém, N.; Cecílio, P.; Silva, C.; Silvestre, R.; A. M. 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*.
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.; Papadopoulou, 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.; Lestinova, 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.; Hem, S.; et al. Phosphoproteome dynamics reveal heat-shock protein complexes specific to the Leishmania donovani infectious stage. *Proc. Natl. Acad. Sci. U. S. A.* **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.; Branquinha, M.H.; D'Avila-Levy, C.M.; Kneipp, L..; Sodré, C.L. *Proteins and Proteomics of Leishmania and Trypanosoma*; Springer, 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 Genomics* **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. Proteomics* **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. Proteomics* **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. Nε-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.