Expanded Proteomic Survey of the Human Parasite *Leishmania major* Focusing on Changes in Null Mutants of the Golgi GDP-Mannose/Fucose/Arabinopyranose Transporter *LPG2* and of the Mitochondrial Fucosyltransferase *FUT1*

Gloria Polanco, a Nichollas E. Scott, b Lon F. Lye, a Stephen M. Beverley a

aDepartment of Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri, USA
bDepartment of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia

**ABSTRACT** The trypanosomatid protozoan parasite *Leishmania* has a significant impact on human health globally. Understanding the pathways associated with virulence within this significant pathogen is critical for identifying novel vaccination and chemotherapy targets. Within this study we leverage an ultradeep proteomic approach to improve our understanding of two virulence-associated genes in *Leishmania*, encoding the Golgi mannos/arabinopyranose/fucose nucleotide-sugar transporter (*LPG2*) and the mitochondrial fucosyltransferase (*FUT1*). Using deep peptide fractionation followed by complementary fragmentation approaches with higher-energy collisional dissociation (HCD) and electron transfer dissociation (ETD) allowed the identification of over 6,500 proteins, nearly doubling the experimentally known *Leishmania major* proteome. This deep proteomic analysis revealed significant quantitative differences in both Δ*lpg2* and Δ*fut1* mutants with *FUT1*-dependent changes linked to marked alterations within mitochondrion-associated proteins, while *LPG2*-dependent changes impacted many pathways, including the secretory pathway. While the FUT1 enzyme has been shown to fucosylate peptides in *vitro*, no evidence for protein fucosylation was identified within our ultradeep analysis, nor did we observe fucosylated glycans within *Leishmania* glycopeptides isolated using hydrophilic interaction liquid chromatography (HILIC) enrichment. This work provides a critical resource for the community on the observable *Leishmania* proteome as well as highlighting phenotypic changes associated with *LPG2* or *FUT1*, ablation of which may guide the development of future therapeutics.

**IMPORTANCE** *Leishmania* is a widespread trypanosomatid protozoan parasite of humans, with ~12 million cases currently, ranging from mild to fatal, and hundreds of millions asymptotically infected. This work advances knowledge of the experimental proteome by nearly 2-fold, to more than 6,500 proteins and thus provides a great resource to investigators seeking to decode how this parasite is transmitted and causes disease and to identify new targets for therapeutic intervention. The ultradeep proteomics approach identified potential proteins underlying the “persistence-without-pathology” phenotype of mutants with deletion of the Golgi nucleotide transporter *LPG2*, showing many alterations and several candidates. Studies of a rare mutant with deletion of the mitochondrial fucosyltransferase *FUT1* revealed changes underlying its strong mitochondrial dysfunction but did not reveal examples of fucosylation of either peptides or N-glycans. This suggests that this vital protein’s elusive target(s) may be more complex than the methods used could detect or that this target may not be a protein but perhaps another glycoconjugate or glycolipid.

**KEYWORDS** N-linked glycans, N-linked glycoconjugates, trypanosomatid protozoan parasite, fucose, glycoproteome, kinetoplast, mitochondria, ultradeep proteomics
Leishmaniasis is a devastating parasitic disease caused by species of the trypanosomatid protozoan parasite genus *Leishmania*. Currently, there are over 12 million cases, with 1.7 billion people at risk of infection worldwide, and estimates of asymptomatic infections ranging as high as >300 million (1–4). Clinical manifestations include self-healing, localized or diffuse cutaneous lesions (cutaneous leishmaniasis), destruction of the nasopharyngeal mucosa (mucocutaneous leishmaniasis), and enlargement of the spleen or liver, which can lead to death (visceral leishmaniasis) (3).

*Leishmania* parasites are transmitted to mammalian hosts via the bite of phlebotomine sand flies and undergo a series of transformations throughout their life cycle, predominantly as promastigotes (insect stage) and amastigotes (intracellular in mammalian host) (3, 5). Powerful genetic tools have uncovered a number of loci that are essential to the completion of the parasite’s life cycle and/or potentially suitable as targets for chemotherapy (6–10). Here, we focus on two mutants relevant to such efforts, whose properties are depicted in Fig. 1.

*Leishmania* parasites are completely coated in a dense glycocalyx largely containing glycoconjugate lipophosphoglycan (LPG), composed of (i) a 1-O-alkyl-2-lyso-phosohotidylinositol lipid anchor, (ii) a heptasaccharide core, (iii) a phosphoglycan polymer consisting of 15 to 30 Galβ1,4Manα1-PO₄ repeating units (phosphoglycan [PG] repeats), often bearing side chain sugars such as galactose in *Leishmania major*, and (iv) a small oligosaccharide cap (11–13). Biochemical and genetic studies of *L. major Δlpg1*– (this nomenclature indicates homozygous deletion and precise replacement of the LPG1 ORF) mutants specifically lacking LPG have implicated it in many key steps of the parasite infectious cycle: in the sand fly vector, as well as in the establishment of infection within the mammalian host following the sand fly bite (14–18). However, the amastigote stage lacks significant levels of LPG, and the few Δlpg1– parasites able to survive the initial host response are thereafter highly virulent (17). In contrast, deletion of LPG2, encoding a Golgi nucleotide sugar transporter (NST; described below), resulted in loss of LPG along with related glycoconjugates, including proteophosphoglycans, which are normally expressed in amastigotes (19). While many Δlpg2– phenotypes in both sand flies and establishment of mammalian infection mirrored those of Δlpg1– mutants, infections of susceptible mice with Δlpg2– mutants showed no overt disease and long-term persistence of a small number of parasites typical of long-term asymptomatic infections (19). This “persistence-without-pathology” phenotype induced protective immunity in a manner similar to that seen in natural healed infections (20).

Although the attenuated Δlpg2– phenotype had been attributed solely to the loss of LPG and related glycoconjugates (19), this view was challenged by studies of an analogous LPG-deficient mutant obtained by genetic deletion of the partially redundant *LPG5A* and *LPG5B* genes, encoding the Golgi UDP-galactose transporters (21). These parasites resemble the Δlpg2– mutant in lacking LPG and related PG-bearing glycoconjugates but, unlike Δlpg2– parasites, retained pathology following inoculation into susceptible mice, similar to that seen with Δlpg1– mutants lacking only LPG (22). This suggested the possibility of LPG2-dependent “off-LPG/phosphoglycan” effects, identification of which could shed new insight on both the persistence-without-pathology phenotype of the Δlpg2– strain, its ability to serve as a live vaccine line, and its tendency to revert toward virulence via second-site events (20, 23, 24).

The second mutant studied arises from finding that LPG2 was the first example of a multispecific NST, able to transport GDP-L-fucose as well as GDP-α-arapyranose (25). While the roles of GDP-Man in phosphoglycan repeat synthesis and α-Arap as a “capping” sugar able to block LPG interaction with the sand fly epithelium had been thoroughly studied (26, 27), GDP-Fuc transport was enigmatic. Although *L. major* expresses low levels of GDP-fucose mediated by the closely related bifunctional salvage enzymes AFKP80 and FKP40 (28, 29), convincing evidence for fucoconjugates has been hard to find. However, deletion of the two salvage enzymes could not be achieved unless...
“metabolic complementation” of GDP-fucose was engineered through expression of the trypanosome de novo GDP-fucose-synthetic pathway (29). Reasoning that this implied the existence of an essential fucosyltransferase (FUT), we identified 5 candidate fucosyltransferases, four of which appeared to be targeted as expected to the parasite secretory pathway. Null-mutant Δfut1 parasites exhibit severe defects, including significantly decreased growth rate, abnormal mitochondrial morphology, decreased mitochondrial membrane potential (Ψm), and loss of kDNA compactness or the kDNA network altogether. Arap, arabinopyranose (green star); Man, mannose (green circle); Fuc, fucose (red triangle); LPG, lipophosphoglycan. This figure was created with BioRender.com.

The fifth candidate, FUT1, however, was important for parasite survival and potentially the key player in the GDP-fucose requirement. Unlike the other candidates, FUT1 was found throughout all trypanosomatid species but not in other organisms (30, 31). Unexpectedly, FUT1 was found to target the parasite mitochondrion, as does its homolog TbFUT1 from Trypanosoma brucei, and both the mitochondrial localization and fucosyltransferase activity were essential (30, 31). While the gene was impossible to knock out by conventional approaches, through plasmid shuffling and examination of more than 1,000 events, a single, rare Leishmania Δfut1 deletion segregant (Δfut1-) mutant was obtained (30). This mutant line displayed severe growth and mitochondrial defects, which were rescued by L. major FUT1 (LmjFUT1) or TbFUT1 re-expression (30). While FUT1 proteins described in other organisms are able to fucosylate a variety of glycan substrates (32, 33), recombinant L. major FUT1 was unexpectedly able to fucosylate both glycan and peptide substrates in vitro (30). Thus far, neither the native acceptor in Leishmania nor that in trypanosomes has been identified, despite considerable effort in both species (30, 31).
predicted proteome, expanding the previously known experimental *L. major* proteome by nearly 2-fold. Differential proteomic analysis showed numerous differences in the Δ*pg*2− and Δ*fut*1+ parasites, including some impacted in both. Despite the depth of this analysis, we failed to identify any fucose-bearing proteins across these proteomes. Furthermore, while glycopeptide enrichment analysis of WT samples also identified numerous instances of N-linked glycopeptides, no O-fucosylation or fucosylated N-linked glycans were observed. Taken together, this work highlights the challenges of identifying FUT1 targets and indicates that alternative approaches will be required in future studies.

RESULTS

Acquisition of a high coverage of the *L. major* proteome. We examined the WT Fn strain of *L. major*, whose genome is one of the best-characterized references for *Leishmania* spp., as well as *L. major* Fn Δ*fut*1+ (30) and a newly created Fn Δ*pg*2− mutant. CRISPR/Cas9 mutagenesis was used to readily generate homozygous open reading frame (ORF) replacements (see Fig. S1 in the supplemental material), and one Fn Δ*pg*2− clonal line (c14.2) was chosen for further study. Preliminary studies showed that it lacks LPG by agglutination tests and does not exhibit lesion pathology when inoculated into susceptible mice but persists at low levels, similar to the previously described *L major* LV39 clone 5 Δ*pg*2− mutant (19). Previous studies showed that restoration of LPG2 expression restores all WT phenotypes tested in both the Fn and LV39 clone 5 backgrounds (19, 34).

For each line, four replicate cultures were initiated and harvested in logarithmic growth phase. Parasite lysates were digested with trypsin and separated into 12 concatenated fractions by basic reverse-phase C18 chromatography (35) and then individually separated and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). To ensure the ability to localize any potential glycosylation events, precursors were subjected to both higher-energy collisional dissociation (HCD) and electron transfer dissociation (ETD) fragmentation with spectra searched against a database consisting of the predicted *L. major* proteome. Protein matches meeting a 0.01 false discovery rate (FDR) cutoff, excluding contaminants, reverse decoys, and those identified only by site, were retained for an initial analysis of the coverage (Table S1).

The numbers of proteins detected in WT, Δ*pg*2−, and Δ*fut*1+ strains were similar (6,208, 6,102, and 5,934, respectively), totaling 6,744 (Table S1) and representing nearly 80% of the predicted *L. major* proteome (8,307 and 8,038 in TriTrypDB and UniProt, respectively). In comparisons with two previous proteomic studies of *L. major* (36, 37), 3,484 proteins identified had no prior MS-based evidence, while 3,260 had prior MS-based evidence, and 308 detected previously did not appear in our data set (Fig. 2A). Thus, this work increased experimental coverage of the *Leishmania* proteome by nearly 2-fold.

No predicted mitochondrion-encoded proteins were identified, despite the inclusion of many of them in our predicted proteome database (described in Materials and Methods). Often these are omitted from proteomic computational analysis completely, and due to their hydrophobicity, they are often not detected by standard or general approaches similar to those used here (38–40).

Protein glycosylation and absence of detectable fucosylation. We showed previously that recombinant FUT1 is able to modify both glycan and peptide substrates in vitro (30), and the expanded proteome provided an opportunity to search for evidence of in vivo fucosylation. To allow the localization of fucosylation events, ETD MS fragmentation was used, as O-fucosylated peptides are poorly localized with collision-based approaches (41). We searched first for peptide-O-fucose by considering deoxyhexose (dHex) modifications of serine or threonine residues, yet despite our proteome depth and the ability of our group to identify glycosylation events from deep proteome data sets (42, 43), no high-confidence O-fucosylation events were observed. Moreover, we were specifically unable to detect HSP70 or HSP60 peptide O-fucose, inferred previously in *Leishmania donovani* (44), peptides of which were fucosylated by recombinant FUT1
in vitro (30). Open searching (45) was also undertaken using MSFragger (46), which also failed to reveal any putative fucose-containing glycopeptides from our deep proteome analysis. To further explore the diversity of glycans which may exist within *L. major*, we surveyed glycopeptides derived from *L. major* using zwitterionic-hydrophilic interaction liquid chromatography (ZIC-HILIC) (47) coupled to LC-MS analysis. Using open searching with MSFragger (46), multiple unique glycoforms were observed, the majority of which corresponded to N-linked glycoforms. While short O-linked glycopeptides can be enriched with this approach (48), our observation of predominately N-linked glycopeptide supports the idea that these escaped detection in *L. major* using ZIC-HILIC-based enrichment. We identified 65 glycopeptides bearing a variety of high-mannose N-linked glycans, as reported previously for *Leishmania* (49–52). The glycopeptides mapped to 49 different proteins, with 14 showing multiple sites and/or glyco-heterogeneity; 24 were predicted as hypothetical proteins of unknown function. The annotated glycosylated proteins are summarized in Table 1 and include many described previously, such as gp63/leishmanialysin, nucleotidases, and phosphatases (50), or ones likely to be so due to the presence of targeting motifs for the secretory pathway such as *FUT2* (30). While these findings highlight the idea that multiple N-linked glycopeptides could be recovered from *L. major*, no fucosylated N-linked glycans were observed (Table S2).

**WT and mutant cell lines display qualitatively similar proteomes.** To compare the impacts of *LPG2* and *FUT1* loss, we first created a high-confidence data set by parsing the total proteome (Table S1) for proteins detected in two or more biological replicates, in at least one cell line. This retained 5,841 proteins (Table S3), of which 95% (5,576) occurred in all three lines (Fig. 2B), evidence of high congruency. Of the remainder, 16 to 21 were unique to one line, and 62 to 97 were shared by two lines (Fig. 2B). As expected, the *LPG2* protein was detected in all WT and Δ*fut1* lysates but not in
Mitochondrial proteome and gene ontology. As FUT1 is targeted to the mitochondrion, where its loss results in numerous mitochondrial abnormalities (30), we queried the high-confidence proteomic data set for nuclearly encoded mitochondrial proteins present in the trypanosomatid predicted mitochondrial protein database MiNT (54). Of 1,559 mitochondrial L. major proteins in the MiNT database (approximately 19% of the total proteome), 1,268 (81%) were identified (Fig. 3A; Table S3), with the WT having a slightly higher number of identified proteins (1,241), followed by the Δlpg2− (1,239) and Δfut1−s (1,226) mutants (Fig. 3B). Quantitative changes were apparent (discussed below), and no mitochondrially (maxicircle) encoded proteins were identified, as discussed above.

A gene ontology (GO) analysis was performed using the PANTHER classification system (Fig. S2; Table S4) (55). As seen previously (36), much of the L. major proteome consists of hypothetical proteins lacking annotation (over 60%) (Fig. S2A), and GO analysis of annotated proteins showed considerable similarity among all three lines. For biological processes (Fig. S2B), the categories with the highest number of genes were cellular process (GO:0009987), metabolic process (GO:0008152), localization (GO:0051179), and biological regulation (GO:0065007). For cellular components (Fig. S2C), the most commonly assigned categories were cellular anatomical entity (GO:0110165), intracellular (GO:0005622) and protein-containing complex (GO:0032991). For molecular functions (Fig. S2D), the categories were catalytic activity (GO:0003824) and binding (GO:0005488).

Knockouts of FUT1 and LPG2 have significant differential impacts on L. major protein abundance. Quantitative differences among WT, Δfut1−, and Δlpg2− parasites were assessed by analysis of variance (ANOVA) with an FDR of 0.05 applied to the protein abundance values (label-free quantitation [LFQ]) (Table S3), which were then normalized by Z-scoring to construct a heat map (Fig. 4). Clustering showed clear grouping of replicates within each line, as well as blocks encompassing 465 proteins.
differing significantly (Fig. 4). Of these, 107 increased and 142 decreased in Δfut1 mutant parasites only, 173 decreased and 13 increased in Δlpg2 mutant parasites only, and 30 decreased in Δfut1 mutant parasites but increased in 2 of 4 Δlpg2 biological replicates (Tables S3, S5, and S6; Fig. 4). Again, the majority of differentially expressed proteins were hypothetical proteins (254/465 total) (Table S3).

**FIG 3** Representation of the predicted *L. major* mitochondrial proteome. Proteins from WT or mutant parasite lines (Table S3) were compared to proteins in MiNT (predicted nuclearly encoded mitochondrial protein), supplemented with maxicircle (mitochondrion)-encoded proteins as described in Materials and Methods. (A) A total of 1,268 proteins, representing 81% of the mitochondrial proteome, were identified in this study. (B) Distribution of mitochondrial proteins by cell line.

**FIG 4** Patterns of variation of significantly differentially expressed proteins among WT and *FUT1* and *LPG2* knockout *L. major* lines. ANOVA using an FDR of 0.05 and an S0 of 1 was performed on LFQ values for the high-confidence proteome (Table S3), to identify proteins that are significantly different between WT, Δfut1, and Δlpg2 clones. From these data, a heat map was generated, clustering the parasite lines, each with four replicas (top dendrogram), and the differentially expressed genes (left dendrogram). The cluster properties are summarized on the right, and the specific proteins included in each cluster can be found in Table S5 and S6.
TABLE 2 Mitochondrial proteins significantly changed in Δfut1⁺ parasites

| Cluster pattern       | Protein name                                                                 | TriTrypDB ID     | Difference in log2(LFQ intensity) | Corrected ANOVA P value |
|-----------------------|------------------------------------------------------------------------------|------------------|----------------------------------|-------------------------|
| Down in Δfut1⁺        | 3-HMG-CoA reductase (EC 1.1.1.34)                                             | LmjF.30.3190     | 0.87 vs WT                        | 0.01                    |
|                       | Aldehyde dehydrogenase, mitochondrial (EC 1.2.1.3)                           | LmjF.25.1120     | 1.02 vs WT                        | 0.01                    |
|                       | Chaperonin HSP60, mitochondrial                                              | LmjF.36.2030     | −0.81 vs WT                       | 0.04                    |
|                       | Kinetoplast-associated protein-like protein                                   | LmjF.27.0240     | −1.61 vs Δfut1⁺                  |                         |
|                       | Metalloleptidase, clan ME, family M16 (EC 1.10.2.2)                           | LmjF.35.1380     | −1.06 vs WT                       | 0.02                    |
|                       | Proline dehydrogenase (EC 1.5.5.2)                                           | LmjF.26.1610     | −1.40 vs Δfut1⁺                  |                         |
|                       | Putative asparaginyl-tRNA synthetase (EC 6.1.1.22)                            | LmjF.34.2340     | −1.48 vs Δfut1⁺                  |                         |
|                       | Putative carbamoyl-phosphate synthase (EC 6.3.4.16)                          | LmjF.16.0590     | −1.25 vs Δfut1⁺                  |                         |
|                       | Putative cytochrome c oxidase VIII (EC 1.9.3.1)                               | LmjF.31.1570     | −2.79 vs WT                       |                         |
|                       | Putative cytochrome c oxidase subunit 10 (EC 1.9.3.1)                        | LmjF.23.0370     | −1.15 vs WT                       |                         |
|                       | Putative DNA-directed RNA polymerase I largest subunit (EC 2.7.7.6)          | LmjF.32.0650     | −1.42 vs WT                       |                         |
|                       | Putative heat shock protein                                                  | LmjF.18.1370     | −1.38 vs WT                       |                         |
|                       | Putative NADH dehydrogenase (EC 1.6.99.3)                                    | LmjF.36.5380     | −1.67 vs WT                       |                         |
|                       | Putative translation initiation factor                                        | LmjF.17.1290     | −1.36 vs WT                       |                         |
| Up in Δfut1⁺          | Inositol phosphosphingolipid phospholipase C like                             | LmjF.08.0200     | 2.00 vs WT                        | 0.01                    |
|                       | Putative 3,2-trans-enoyl-CoA isomerase mitochondrial (EC 1.1.1.35) (EC 5.3.3.8) | LmjF.31.2330     | 2.93 vs Δfut1⁺                   |                         |

*Annotated mitochondrial proteins differing significantly in abundance, as determined by ANOVA, between the Δfut1⁺ and WT or Δlpq2⁻ parasites (Fig. 4; Tables S3 and S5). Overall, there were 279 proteins in clusters where proteins decreased in Δfut1⁺ parasites (142 total; 26 mitochondrial, of which 13 are unannotated), increased in Δfut1⁺ parasites (107 total; 17 mitochondrial, of which 15 are unannotated), or decreased in Δfut1⁺ and varied in Δlpq2⁻ parasites (30 total; 13 mitochondrial, of which 12 are unannotated). CoA, coenzyme A.

SHERP, a stationary-phase/metacyclic parasite marker (56), was not seen in any Δlpq2⁻ replicates but was unexpectedly detected in log-phase WT and Δfut1⁺ parasites at similar levels (Tables S1 and S3). These results were corroborated by Western blot analysis with anti-SHERP antisera (Fig. S3).

Significantly different proteins accompanying the severe mitochondrial dysfunction of Δfut1⁺. The Δfut1⁺ mutant shows considerably slower growth and numerous mitochondrial abnormalities, including lowered mitochondrial membrane potential and complete loss or loss in compactness of the kinetoplast DNA (kDNA) network (30). From the cluster of 142 significantly less abundant proteins in Δfut1⁺ parasites (Fig. 4; Table S5), 26 (18%) were predicted as mitochondrial, of which 13 were annotated (Table 2). Of 107 proteins that were significantly more abundant in Δfut1⁺ parasites, 15 were mitochondrial, of which 2 were annotated (Table 2). The differentially expressed mitochondrial proteins are depicted in Fig. 5, many of which are positioned to contribute to Δfut1⁺ mitochondrial dysfunction (see Discussion).

Protein changes seen in the secretory pathway mutant Δlpq2⁻. Overall, 174 proteins were significantly decreased in Δlpq2⁻ parasites, of which 89 are of unknown function, while 13 increased in Δlpq2⁻ parasites, of which 6 are of unknown function (Fig. 4; Tables S3 and S6). The predominance of proteins downregulated in Δlpq2⁻ relative to Δfut1⁺ organisms could be related to the relative health of the Δlpq2⁻ mutant (this work and reference 19), as many of the changes seen in Δfut1⁺ parasites involved stress responses arising from mitochondrial dysfunction. Notably, the proteins altered in Δlpq2⁻ parasites mapped to a variety of cellular compartments and metabolic pathways.
(Tables S4 and S6), suggesting that despite the normal growth of this mutant in culture, the impact of loss of \textit{LPG2} was nonetheless profound. The broad impact of \textit{LPG2}-dependent effects thus hinders efforts to pinpoint those most important to the biological alterations of most interest, such as persistence without pathology.

\textit{LPG2} loss affected several proteins involved in the glycosylation pathways. Two examples are decreases of phosphomannomutase (PMM) and phosphoacetylglucosamine mutase (PAGM), which catalyze the reversible transfer of phosphate between C-6 and C-1 hydroxyl groups of mannose and N-acetylglucosamine (57). PMM in particular is necessary for establishing infection in macrophages (27).

Potential interactions between \textit{LPG2}- and \textit{FUT1}-dependent pathways. While the proteins differentially altered in \textit{Δlpg2} and \textit{Δfut1} parasites were mostly quite distinct (Fig. 4; Tables S5 and S6), in 2 of 4 replicate \textit{Δlpg2} lines a small group of 30 proteins showed inverse regulation to changes observed in \textit{Δfut1} parasites (Fig. 4; Table S5). Of these, 12 were unannotated predicted mitochondrial proteins. As the replicates were grown at the same time from the same inoculation under the same conditions, we cannot account for the variability. Changes in \textit{Δfut1} parasites were also observed in proteins outside the mitochondrion (Table S5). Interestingly, these included increase of other glycosyltransferases, including α-1,2-mannosyltransferase, phosphoglycan β1,2-arabinosyltransferase, and phosphoglycan β1,3-galactosyltransferase 3 (Table S5). We

\textbf{FIG 5} Schematic overview of a \textit{Leishmania} mitochondrion showing proteins with significantly altered expression in \textit{Δfut1} relative to WT \textit{L. major}. The specific proteins are listed in Table 1. Only mitochondrial proteins showing differential expression (Table 2) are shown. In the panel depicting the electron transport chain, the proteins decreased in the \textit{Δfut1} parasites are shown in blue, those not detected in our data set are shown in white, and those that were unchanged are shown in gray. It should be noted that trypanosomatids possess a single mitochondrion per cell whose structure differs from the typical metazoan mitochondrion shape depicted here. This figure was created with BioRender.com.
speculate that the cross-mutant effects arise from competition for GDP-fucose synthesized in the cytosol (28, 29), for transport and use by secretory pathway fucosyltransferases (dependent on LPG2) or the mitochondrion (FUT1).

**DISCUSSION**

This work presents a high-coverage *L. major* proteomic data set consisting of new mass spectrometry-based evidence for nearly 3,500 proteins beyond the 3,600 proteins typically identified experimentally in several *Leishmania* species, including *L. major*, by others (36). For quantitative analysis, the data were parsed to yield a high-confidence data set of 5,841 proteins. Comparison showed that the proteomes of WT, Δlpg2−, and Δfut1− parasites were highly congruent, but with significant quantitative variation (58). Over half of the significantly changing proteins were uncharacterized hypothetical proteins.

**FUT1 and mitochondrial dysfunction.** In addition to severely delayed growth, the Δfut1− mutant exhibits profound mitochondrial dysfunction, including loss of membrane potential, bloated cristae, presence of large aggregates, loss of kDNA compactness, and complete loss of the kDNA network in some parasites (30). Benefiting from the high coverage of mitochondrial proteins in our data set (>80%) (Fig. 3), we were able to survey the impact of FUT1 deletion on these (Table 2; Table S5; Fig. 4). The Δfut1− mutant had significantly decreased levels of key components of mitochondrial respiratory chain complexes III (MMP-β) and IV (cytochrome c oxidases COX8 and COX10), (59) as well as the alternative NADH dehydrogenase (NDH2) (58, 60). Δfut1− alterations in several maxicircle (mitochondrion)-encoded components of the respiratory chain could not be assessed, as they were absent in our data sets. Collectively, downregulation of this pathway matches the changes in membrane potential and growth seen in Δfut1− parasites (30).

Δfut1− parasites showed downregulation of kinetoplast-associated proteins (KAPs), histone-like proteins attributed to packaging mitochondrial DNA (kDNA) in trypanosomatids (61). Downregulation of these has been associated with rearrangement of the kinetoplast structure, parasite growth defects, and shrinkage and complete loss of the kDNA network (62, 63), all phenotypes seen in Δfut1− parasites (30).

Enzymes impacted in Δfut1− parasites include 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) and proline dehydrogenase (ProDH) (Table 2; Table S5; Fig. 5). HMGR is a key enzyme in the mevalonate pathway, yielding isoprenoids important for viability and proliferation (64). ProDH mediates the first step in a pathway that leads to production of α-ketoglutarate entering the tricarboxylic acid cycle (65) and generates FADH2, which transfers electrons to coenzyme Q and the electron transport pathways (65). ProDH transcripts have been found to be present in higher levels in drug-resistant *Leishmania* (66).

Several proteins associated with stress responses were also impacted in Δfut1− parasites, including mitochondrial heat shock proteins HSP60 and HSP110 (67). HSP60 acts as one of the principal mediators of protein homeostasis in the mitochondrial matrix and in response to damaged or misfolded mitochondrial proteins after oxidative stress (67, 68). The inositol phosphosphingolipid phospholipase C-like (ISCL) protein, which contributes to parasite survival in acidic environments, was increased in Δfut1− parasites (69).

**A first-pass *L. major* glycoproteome, lacking detectable fucosylation.** The FUT1 protein shows motifs characteristic of fucosyltransferases within the GT11 family, and recombinant *Trypanosoma brucei* and *L. major* FUT1 proteins were able to fucosylate various glycans in vitro (30, 31). Moreover, LmjFUT1 was shown to fucosylate several peptides in vitro (30 [TbrFUT1 was not tested]). The in vivo substrate of FUT1 has proven elusive in both parasite species, and the broad specificity of LmjFUT1 prompted us to search for protein O-fucosylation as part of this work. While glycopeptides were not detected with high confidence in the total proteomic studies, potentially reflecting technical considerations (70, 71), following HILIC enrichment we were able to identify 58 glycopeptides representing 49 *Leishmania* proteins, many of which were known or
anticipated to be glycosylated but many of which were novel (Table 1). Several other potential modifications were identified computationally, such as pyrophosphorylation (72), which will require experimental verification. In Trypanosoma cruzi, HILIC enrichment similarly successfully identified a variety of N- and O-glycosylated proteins (73). Comprehensive studies of the Leishmania glycoproteome were lacking (50), with previous studies being mostly limited to a few glycoproteins such as GP63 (leishmanialysin). Despite the utility of the HILIC enrichment, many known glycosylated proteins of Leishmania were not detected, and thus, the proteins in Table 1 should be regarded as a first-pass (preliminary) glycoproteome, confirmation of which will be needed in any event.

Notably, we did not find evidence for peptide O-fucose or fucosylated glycans. Although glycopeptide enrichment using HILIC can be utilized for O-glycoproteome characterization, it may bias for peptides with large glycans or multiple glycosylations (74). It is possible that any potential FUT1-dependent protein fucosylation may be in a form recalcitrant to the methods used, or that other parasite stages could yield different results. It is also possible that further proteomic searches extending our ultradeep effort might eventually yield proteins or modifications (including fucosylation) not identified here. Alternatively, while it is difficult to prove a negative, the possibility that the in vivo substrate of FUT1 may be glycans or glycolipids rather than proteins or glycoproteins seems at least as likely from the absence of detectable peptide fucosylation. Resolution of this question ultimately will depend on identification of the relevant FUT1-dependent glycoconjugate in the future.

Summary and perspective. Proteins showing altered expression in Δlpg2 parasites could contribute to the persistence-without-pathology phenotype seen in this mutant; however, as these parasites grow normally in culture, we were challenged to compose a predictive query of the 186 differentially expressed candidates for those most likely to contribute to the phenotype. These data also raised the possibility that changes in the expression of one or more of these proteins could contribute to the recovery of amastigote virulence in Δlpg2 revertants occasionally found in infected animals (23). Experimental tests of the role(s) of the LPG2-dependent proteome in the loss or recovery of amastigote virulence will be required to resolve this. Similarly, Δfut1 parasites showed 249 differentially expressed proteins, including many known or potential mitochondrial proteins, as expected from its strong impact on mitochondrial function, as well as a number of nonmitochondrial proteins. While our studies raise many hypotheses for the roles of both FUT1 and LPG2 in parasite biology, genetic confirmation will be required, as well as studies to establish whether the effects seen are direct or indirect consequences of gene ablation. Fortunately, with the advent of high-throughput knockout screening via CRISPR technology (75), probing the importance of this panoply of genes will be increasingly feasible.

MATERIALS AND METHODS

Cell culture. All parasites were derivatives of L. major strain Fn (MHOM/IL/80/Fn). Parasites were grown as the promastigote form in vitro in complete medium 199 (M199) supplemented with 10% heat-inactivated fetal bovine serum (FBS), 40 mM HEPES (pH 7.4), 100 mM adenine, 1 μg/mL biotin, 5 μg/mL hemin, penicillin-streptomycin, and 2 μg/mL biopetin. Parasites were seeded in 200 mL medium at a density of 10^5/mL in roller bottles and allowed to grow until mid-log phase, corresponding to a density of 1 x 10^9 to 4 x 10^9/mL. Culture vessels were rotated at approximately 2 rpm using a cell production roller apparatus (Bellco Biotechnology). The Δfut1 segregant was described previously (30).

Generation of a Δlpg2 mutant line by CRISPR/Cas9 mutagenesis. Parasites were engineered to express a human codon-optimized Streptococcus pyogenes Cas9 and an LPG2 sgRNA (plasmid p63Phleo-HispCas9 or B7521; single guide RNA [sgRNA] synthesized from template B7617) expressed from a U6 promoter as described previously (76). To replace the LPG2 ORF, homology-directed repair (HDR) templates were designed containing a hygromycin B (HYG) resistance gene flanked by sequences matching the 30 nucleotides flanking the LPG2 ORF. A total of 10 μg of HDR DNA was transfected into the L. major HSpCas9-B7617 parasites, which were then plated on semisolid M199 medium containing 50 μg/mL hygromycin B. Numerous hygromycin B-resistant (HYG') colonies were obtained, of which 26 were tested and appeared to be WT/Δlpg2:HYG heterozygotes. Subsequently, parasites were transfected with a blasticidin (BSD) resistance gene HDR template, bearing the 30 nucleotides (nt) of LPG2 flanking sequence as before. The HSpCas9-B7617-HYG' parasites were transfected and plated onto semisolid
medium containing 50 μg/mL hygromycin B and 10 μg/mL blasticidin. A total of 22 colonies were tested and confirmed to be Δpg2/HYD/Δpg2.850 knockouts by PCR tests (Fig. S1), and preliminary data showed them to be LPG deficient, as judged by their failure to agglutinate with monoclonal anti-PG antibody WIC79.3. One clone was selected for further study (SpCas9-R7617 B+H-c14.2), referred to here as the Δpg2 Z clone.

**Generation of parasite lysates for proteomic analysis.** Whole-parasite lysates were prepared as described previously (77) with some modifications. After 3 washes with ice-cold phosphate-buffered saline (PBS), parasites were resuspended in 1 mL of ice-cold lysis buffer [6 M guanidinium chloride, 100 mM Tris (pH 8.5), 10 mM Tris(2-carboxyethyl)phosphine (TCEP), 40 mM 2-mercaptoethanol (CMA)] supplemented with a protease inhibitor tablet (Roche) and 0.2 mg/mL 1,10-phenanthroline. Samples were immediately boiled at 95°C in a ThermoMixer at 1,500 rpm. After cooling samples on ice for 10 min, an additional boiling step was done at 95°C. One milligram of protein was collected from four bio-waste bottles. The supernatant was removed, samples were air dried while covered with a Kimwipe and stored at −80°C until ready for analysis.

**Digestion of L. major proteome samples.** Precipitated protein pellets were resuspended in 50% trichloroethanol and then heated at 50°C for 10 min with shaking at 1,000 rpm. Resuspended samples were then reduced/alkylated in a single step with 20 mM Tris(2-carboxyethyl)phosphine and 40 mM chloroacetamide for 1 h in the dark. Samples were then diluted 10-fold with 100 mM triethylammonium bicarbonate and digested with trypsin (1/100 [wt/wt]) overnight with shaking at 800 rpm. Digested samples were then reduced/alkylated in a single step with 20 mM Tris(2-carboxyethyl)phosphine and 40 mM chloroacetamide for 1 h in the dark. Samples were then diluted 10-fold with 100 mM triethylammonium bicarbonate and digested with trypsin (1/100 [wt/wt]) overnight with shaking at 800 rpm. Digested samples were then acidified to a final concentration of 0.5% formic acid and desalted with 50-mg C18 Sep-Pak columns (Waters Corporation, Milford, MA, USA) according to the manufacturer’s instructions. Briefly, C18 Sep-Pak columns were conditioned with buffer B (0.1% formic acid [FA], 80% acetonitrile [ACN]) and washed with 10 volumes of buffer A (0.1% trifluoroacetic acid [TFA], 2% ACN), samples were loaded, columns were washed with 10 volumes of buffer A, and bound peptides were eluted with buffer B and then dried by vacuum centrifugation.

**High-pH fractionation.** Proteome samples were fractionated by basic reverse-phase chromatography according to the protocol of Batt and Olsen (78) with minor modifications. Briefly, peptides were resuspended in 1 mL of buffer A (5 mM ammonium formate, pH 10.5) and separated using a 1100 series high-performance liquid chromatograph (HPLC) (Agilent Technologies, CA) using a Gemini NX C18 column (4.6 by 250 mm, 5 μm) according to the manufacturer’s instructions. Brieﬂy, C18 Sep-Pak columns were conditioned with buffer B (0.1% formic acid [FA], 80% acetonitrile [ACN]) and washed with 10 volumes of buffer A (0.1% trifluoroacetic acid [TFA], 2% ACN); samples were loaded, columns were washed with 10 volumes of buffer A, and bound peptides were eluted with buffer B and then dried by vacuum centrifugation.

**Proteomics analysis using reverse-phase LC-MS.** Pooled basic reverse-phase fractions were resuspended in buffer A* (2% ACN, 0.1% TFA) and separated using a two-column chromatography setup composed of a PepMap100 C18, 20-mm by 75-μm trap and a PepMap C18, 300-mm by 75-μm analytical column (Thermo Fisher Scientific, CA) using a Gemini NX C18 column (4.6 by 250 mm, 5 μm) Phenomenex, CA) at a flow rate of 1 mL/min. Separation was accomplished using a 90-min gradient with samples loaded on the column at 2% buffer B (5 mM ammonium formate, 90% ACN [pH 10.5]) for 6 min. The concentration of buffer B was then ramped from 2% to 28% over 45 min, from 28% to 40% over 5 min, and from 40% to 80% over 5 min. The gradient was held at 80% B for 2 min, and then the column was regenerated by being returned to 2% B over 10 min and held at 2% B for 10 min. Sixty 1-min fractions were collected. Each group of five fractions was combined to generate a total of 12 pooled fractions, which were concentrated by vacuum centrifugation, desalted using C18 stage tips, and then subjected to mass-spectrometric analysis.

**Proteome samples were processed using MaxQuant (v1.6.3.4) (79) and searched against the Leishmania major MHOM (UniProt accession no. UPQ000000542, containing 6,038 proteins) and TriTrypDB Leishmania major Fc (TrypDB 37) databases. The reference proteome database was supplemented with predicted proteins from the kinetoplast maxicircle (mitochondrion) of L. major (LmjF00.0040, NADH dehydrogenase 7 [MURF3, ND7]; LmjF00.0050, cytochrome oxidase 3; LmjF00.0060, Cyb; LmjF00.0070, MURF4 [A6]; LmjF00.0080, MURF1; LmjF00.0100, NADH dehydrogenase subunit 1 [ND1]; NADH dehydrogenase subunit 4, NADH dehydrogenase subunit 5, LmjF00.0110 cytochrome oxidase 2 [CO2]; LmjF00.0120, MURF2; LmjF00.0130, cytochrome oxidase 1 [CO1]; LmjF00.0150, NADH dehydrogenase 4 [ND4]; LmjF00.0180, NADH dehydrogenase 5 [ND5]; due to extensive pan-RNA editing, other protein products that could not be reliably predicted were not included. Searches were undertaken using trypsin enzyme specificity with carbamidomethylation of cysteine as a fixed modification. Oxidation of methionine and dHex modifications of serine/threonine residues were included as variable modifications, and a maximum of 2 missed cleavages was allowed. To attempt to identify any potential complex fucosylation events, dependent peptide sequencing was enabled. To ensure the inclusion of only high-quality peptide spectral matches (PSMs), a PSM FDR of 0.1% was set, while an FDR of 1% was allowed at the protein level. To enhance the identification of peptides between samples, the “match between runs” option was enabled, with a precursor match window set to 2 min and an alignment window of 20 min with the LFQ option enabled (80).
The result file was then uploaded to Perseus (V1.6.0.7) (81) for statistical analysis. Potential contaminants, proteins only identified by site, and reverse decoys were removed. For LFQ comparisons, values were log2-transformed and biological replicates were grouped. Data were parsed further by removing proteins not identified in at least two biological replicates in at least one cell line. Missing values were inferred based on the observed total peptide intensities with a range of 0.3× and a downshift of 1.8× using Perseus. A multiple-sample (ANOVA) test with a permutation-based FDR set at 0.05 and the artificial within groups parameter S0 set at 1 was performed to identify significantly differentially abundant proteins between Δfgt1, WT, and Δpg2- lines. A heat map of significantly different proteins was constructed after Z-score-based normalization and Euclidean clustering of the transformed LFQ values (82).

**Glycopeptide enrichment by ZIC-HILIC.** Two hundred fifty micrograms of whole-cell lysates that had been digested, C18 Sep-Pak cleaned, and dried was resuspended in 80% acetonitrile for sharing preliminary data on the LPG status and mouse infectivity of Leishmania major proteins between Dclal within groups parameter S0 set at 1 was performed to identify significantly differentially abundant proteins between Δfgt1, WT, and Δpg2- lines. A heat map of significantly different proteins was constructed after Z-score-based normalization and Euclidean clustering of the transformed LFQ values (82).

**LC-MS analysis of ZIC-HILIC-enriched samples.** ZIC-HILIC-enriched samples were resuspended in buffer A” and separated using a two-column chromatography setup composed of a PepMap100 C18 20-mm by 75-μm trap and a PepMap C18 500-mm by 75-μm analytical column (Thermo Fisher Scientific) coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific). ZIC-HILIC-enriched samples were analyzed using 185-min gradients. Separation gradients were run for each sample, altering the buffer composition from 2% buffer B to 28% B over 106 or 166 min depending on the run length, from 28% B to 40% B over 9 min, and then from 40% B to 80% B over 3 min; the composition was held at 80% B for 2 min, dropped to 2% B over 2 min, and held at 2% B for another 3 min. The Lumos mass spectrometer was operated in a data-dependent mode with a single Orbitrap MS scan (350 to 1,800 m/z), maximal injection time of 50 ms, an AGC maximum of 4 × 10^5 ions, and a resolution of 120,000) acquired every 3 s followed by Orbitrap MS/MS HCD scans of precursors (NCE, 30%; maximal injection time of 100 ms; AGC set to a maximum of 1.0 × 10^4 ions; and a resolution of 15,000). HCD scans containing the oxonium ions (204.0867; 138.0545 and 366.1396 m/z) triggered two additional product-dependent MS/MS scans of potential glycopeptides; an Orbitrap electron transfer–higher-energy collision dissociation (ET-CID) scan (NCE, 15%; maximal injection time of 250 ms; AGC set to a maximum of 2 × 10^4 ions with a resolution of 30,000) and an ion trap collision-induced dissociation (CID) scan (NCE, 35%; maximal injection time of 40 ms; AGC set to a maximum of 5 × 10^4 ions). Data files were searched using MSFragger (v15) (46) using the Leishmania major MHOM proteome (UniProt accession no. UP000000542). Open database searches were performed, allowing modifications between −150 and 1,000 Da on the deep proteome analysis, with global.modsummary.tsv used to assess the presence of fucosylated modifications. For the detection of glycoforms within ZIC-HILIC enrichments, open database searches were performed, allowing modifications between −150 and 2,000 Da. The results from the ZIC-HILIC open searches were combined within R, and only assignments with an MSFragger expectation value of <0.001 and a delta mass of >140 Da were retained for analysis.

**Data availability.** All MS data, search results, and R scripts have been deposited in the PRIDE ProteomeXchange Consortium repository (83, 84) with identifiers PXD015966 and PXD035738.

**SUPPLEMENTAL MATERIAL**

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, XLSX file, 1.6 MB.

**SUPPLEMENTAL FILE 2**, XLSX file, 0.2 MB.

**SUPPLEMENTAL FILE 3**, XLSX file, 1.4 MB.

**SUPPLEMENTAL FILE 4**, XLSX file, 0.5 MB.

**SUPPLEMENTAL FILE 5**, PDF file, 0.4 MB.

**ACKNOWLEDGMENTS**

We acknowledge Igor Almeida for discussions and Deborah Dobson and Suzanne Hickerson for sharing preliminary data on the LPG status and mouse infectivity of L. major Fn Δpg2- . We thank the Melbourne Mass Spectrometry and Proteomics Facility of The Bio21 Molecular Science and Biotechnology Institute at The University of Melbourne for the support of mass spectrometry analysis. We thank Greg Matlashewski for providing plasmids used in CRISPR/Cas9 editing and the members of our group for discussions.

Funding was provided by R01 AI031078 as well as the NIAID R01 research supplement to promote diversity in health-related research (G.P.). This work was supported by National Health and Medical Research Council of Australia (NHMRC) project grants awarded to N.E.S. (APP1100164).

November/December 2022 Volume 10 Issue 6 10.1128/spectrum.03052-22 13
71. Illiano A, Pinto G, Melchiorre C, Carpentieri A, Faraco V, Amoresano A. 2020. Protein glycosylation investigated by mass spectrometry: an overview. Cells 9:1986. https://doi.org/10.3390/cells9091986.

72. Azevedo C, Livermore T, Saiardi A. 2015. Protein polyphosphorylation of lysine residues by inorganic polyphosphate. Mol Cell 58:71–82. https://doi.org/10.1016/j.molcel.2015.02.010.

73. Alves MJM, Kawahara R, Viner R, Mattos EC, Thaysen-Andersen M, Larsen MR, Palmsano G. 2017. Comprehensive glycoprofiling of the epimastigote and trypomastigote stages of Trypanosoma cruzi. J Proteomics 151:182–192. https://doi.org/10.1016/j.jprot.2016.05.034.

74. Riley NM, Bertozzi CR, Pitteri SJ. 2021. A pragmatic guide to enrichment strategies for mass spectrometry-based glycoproteomics. Mol Cell Proteomics 20:100029. https://doi.org/10.1074/mcp.R120.002277.

75. Beneke T, Demay F, Hookway E, Ashman N, Jeffery H, Smith J, Valli J, Becvar T, Myskova J, Shafig S, Sadlova J, Wheeler RJ, Gluenz E. 2019. Genetic dissection of a Leishmania flagellar proteome demonstrates requirement for directional motility in sand fly infections. PLoS Pathog 15:e1007828. https://doi.org/10.1371/journal.ppat.1007828.

76. Zhang W-W, Matlashewski G. 2015. CRISPR-Cas9-mediated genome editing in vivo insulin signaling dynamics. Nat Biotechnol 33:990–995. https://doi.org/10.1038/nbt.3327.

77. Batth TS, Olsen JV. 2016. Offline high pH reversed-phase peptide fractionation for deep phosphoproteome coverage. Methods Mol Biol 1355:179–192. https://doi.org/10.1007/978-1-4939-3049-4_12.

78. Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372. https://doi.org/10.1038/nbt.1511.

79. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. 2014. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13:2513–2526. https://doi.org/10.1074/mcp.M113.031591.

80. Tyanova S, Temu T, Sinitsyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 13:731–740. https://doi.org/10.1038/nmeth.3901.

81. Tyanova S, Cox J. 2018. Perseus: a bioinformatics platform for integrative analysis of proteomics data in cancer research. Methods Mol Biol 1711:133–148. https://doi.org/10.1007/978-1-4939-7493-1_7.

82. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Perez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tione S, Cox J, Audain E, Walzer M, Jamuzczak AF, Ternes T, Bazzoni A, Vizcaino JA. 2019. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res 47:D442–D450. https://doi.org/10.1093/nar/gky1106.

83. Vizcaino JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternes T, Xu QW, Wang R, Hermjakob H. 2016. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res 44:D447–D456. https://doi.org/10.1093/nar/gkv1145.