LeishMANIAdb: an integrative structural resource of Leishmania proteins

Gábor E. Tusnády[1], András Zeke[1], Zsófia E. Kálmán[2], Marie Fatoux[3*], Sylvie Ricard-Blum[3], Toby Gibson[4], Laszlo Dobson[1,4]

[1] Protein Bioinformatics Research Group, Institute of Enzymology, Research Centre for Natural Sciences, Magyar Tudósok körútja 2, H-1117 Budapest, Hungary
[2] Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter u. 50/A, Budapest 1083, Hungary
[3]Current address: ICBMS UMR CNRS 5246, University Lyon 1, Lyon, Villeurbanne 69622, France (*The work was done in UMR CNRS 5086)
[4] Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany

Abstract

Leishmaniasis is a detrimental disease causing serious changes in quality of life and in some forms death. The disease is spread by the pathogen *Leishmania* transmitted by sandfly vectors and their primary hosts are vertebrates (including humans). The pathogen penetrates host cells and secretes proteins (the secretome) to repurpose them for pathogen growth and to rewire cell signaling via host-pathogen Protein-Protein Interactions (PPIs). Here we present LeishMANIAdb, a database specifically designed to understand how *Leishmania* virulence factors may interfere with host proteins. Since the complete and precise secretome of different Leishmania species is still incomplete, we collected various experimental evidence and computational predictions that may indicate protein secretion, so users can inspect different proteins and see all information unified in a user-friendly web resource. In addition to the secretome, we manually annotated the localization/function of hundreds of transmembrane (TM) *Leishmania* proteins and included annotations of host-pathogen interactions from experiments. We also enriched proteins with structural and functional predictions that can open prospects to understand the molecular mechanisms of infection. Our database, available at http://leishmaniadb.ttk.hu/ may provide novel insights into host-pathogen interactions and help to identify new molecular targets for this neglected disease.

Introduction

Leishmaniasis is a neglected tropical disease causing severe symptoms, affecting around 1 million new people yearly, with annual deaths estimated to be around 60,000 [1]. Although over 90% of cases occur in poor regions south of the Equator, due to climatic changes it appears in new areas, and it has already shown up in Mediterranean European countries [2] or Texas, USA [3]. To this date no human vaccination is available and treatment only helps at an early stage of the infection. *Leishmania* are trypanosomes, belonging to the class Kinetoplastea. Upon infection, the amastigote stage pathogen is detected by phagocytes, where it ends up in a stable parasitophorous vacuole that protects it [4]. The parasite secretes proteins that enter various parts of the cell [5]. The secreted virulence factors then

interfere with cell signaling by interacting with the host proteins: they increase glycolytic metabolism [6], perturb microbicidal pathways [7], escape the innate immune response, and repurpose macrophages for parasite replication [8] by interfering with cellular protein-protein interactions (PPIs). Interestingly, these mechanisms are somewhat unique to *Leishmania* among trypanosomes, which are usually extracellular pathogens and do not enter host cells.

The host targeted interactions are often mediated via Short Linear Motifs (SLiMs) [9]. SLiMs are flexible protein segments composed of a restricted number of residues (between 3-10), that usually bind to ordered protein domains. Their short length and structural flexibility enable them to bind to a wide range of domains. Cellular SLiMs typically bind with low micromolar affinity. These weak and transient interactions enable SLiMs to work in cooperative regulatory systems [10]. Pathogens mimic host SLiMs to interact with host cell proteins [9]. Pathogen SLiMs often bind with higher affinities than the cellular ones, outcompeting the native interactions, permanently re-wiring the host regulation. A few modulatory SLiMs have already been discovered in eukaryotic pathogens: the *Toxoplasma gondii* MapK docking motif [11], stage-specific (promastigote-amastigote) phosphorylation motifs from *Leishmania [12]*. In a recent study, several putative SLiMs were detected in *Leishmania* such as heparin-binding sequences or RGD integrin-binding motifs but their function has not been confirmed yet [13].

Numerous studies investigated *Leishmania* secretomes: most of them expose promastigotes to a heat shock and pH change then analyze the *Leishmania* conditioned medium by liquid chromatography or difference gel electrophoresis combined with mass spectrometry to identify secreted proteins [14], or detect mRNA levels and protein abundance [15]. While high-throughput experiments inherently suffer from a certain level of noise, experiments on individual proteins may be more reliable - in the case of *Leishmania* the vast majority focus on leishmanolysin (GP63) [16,17]. Notably, data were collected on different *Leishmania* species/strains identified via names and identifiers from different sources, making a unified overview challenging. Another key step towards understanding the infection mechanism is the identification of *Leishmania* surface proteins that can mediate the attachment of the pathogen to the host cell. Surfaceome experiments were carried out on related species, but not on *Leishmania* - although human host proteins binding to 24 strains of intact *Leishmania* have been identified [18]. Beside the characterization of *Leishmania* secretomes, the identification of host-*Leishmania* PPIs is needed to narrow down virulence factors perturbing the host cell regulation to modules interfering with host proteins. SLiMs have low information content and simply scanning them in *Leishmania* secretomes may yield many false positives. Their structural and functional context, such as accessibility, conservation and localization, are all key elements to successfully identify those that may have a role in rewiring the host cell regulation. Notably, SLiMs also play a key role in maintaining housekeeping processes in *Leishmania*, therefore to find SLiM rewiring the host regulation, we need to discriminate SLiMs packed on proteins that reach the host cytoplasm or nucleus - unfortunately we only have limited information about these proteins at this date.

No resource collects all this data in a unified and user-friendly way so far. We developed LeishMANIAdb to expedite *Leishmania* research by unifying scattered information from the literature. We collected high-throughput experiments, predictions and enriched them with manual annotation and structural predictions. Although TriTrypDB [19] also offers some of

the datasets we display, it is a genomic resource where individual genes or chromosomal spans can be inspected in their genomic context rather than protein structure and motif information.

Results

Selection of *Leishmania* proteomes and homology mapping of various Trypanosoma proteins

We selected 5 *Leishmania* species (*L. brazliensis*, *L. donovani*, *L. infantum*, *L. major, L. mexicana*) and several related species (*Bodo saltans, Leptomonas seymouri, Trypansoma brucei, Trypansoma cruzi, Trypansoma rangeli, Trypanosoma theileri*) based on the quality (i.e. number of fragments and missing genes) of their proteome from UniProt [20]. All proteins are also cross-referenced to TriTrypDB [19]. Notably, we noticed that around 30% of the proteins have different sequences deposited into these resources, and in most cases the difference is the position of the initiating Methionine - for data compatibility we always use the UniProt sequence version but the conflicts are highlighted in LeishMANIAdb. We also performed a similarity search between these proteins and linked close homologs (see Methods) so annotations and predictions can be easily compared between them. All manual annotations and experimental data from different sources were mapped to these proteins. Altogether LeishMANIAdb contains 40 537 searchable *Leishmania* proteins and 68 924 other *Trypanosome* proteins to strengthen predictions.

Manual annotation of host-pathogen PPIs and TM protein localization

We manually curated hundreds of proteins, using two strategies. The first type of annotation is the collection of host-pathogen PPI experiments on individual proteins, with the majority of them involving leishmanolysin. We collected 54 papers reporting interactions 99 PPIs involving 19 *Leishmania* proteins. Although experiments were mapped back to specific proteins, the results are also displayed on close homologs (with a note that the experimental data is derived from a different protein). The second type of manually curated data is the localization and functional annotation of TM proteins. The aim was to find surface proteins that may facilitate the infection, but we annotated hundreds of other TM proteins with their localizations too. For this task, we used all close homologous proteins defined in the previous step and as that are shared between 3 799 proteins (which is 47% percent of the predicted TM proteomes of the 5 species combined).

The definition of *Leishmania* secretome and protein localization is still incomplete

Leishmania not only exploits host-secretory pathways to distribute effectors but also utilizes an unusual mechanism to deliver proteins to the cytosol of infected cells by releasing exosomes, which can fuse with the vesicular membrane to release their protein content [21]. Therefore computational methods based on signal peptides and localization predictions are not sufficient to predict *Leishmania* secretomes, and we used high-throughput experiments [14,22–25] to increase the coverage of *Leishmania* secretomes. Strikingly, the number of proteins in these secretomes varies to a large extent, and some proteins cannot be identified by mass spectrometry. Other datasets include proteins found in glycosomes, unique to

Leishmania [26], stage-dependent (promastigote/amastigote) phosphoproteomics [12], housekeeping gene localizations [27], exosome content [21], protein and mRNA abundance data [15,28]. When we mapped back secretome and abundance experiments to *Leishmania infantum* (and paralogs) the number of identified proteins ranges from 10 to 2,000 (Figure 1/A), even when experimental conditions were similar. For example, pioneer secretome studies only yielded a few hundred hits, while the latest ones are more inclusive. This can be likely explained by both the quality of sample preparation, MS experimental techniques or the sequences in the background databases.

Gene duplication is often acting on protein families responsible for host-pathogen PPIs, therefore we also collected proteins that are highly expanded in *Leishmania infantum*. In this case we can discriminate proteins with many paralogs within *Trypanosomes* and *Leishmania*-exclusive proteins. We found a distinct amino acid transporter family with many proteins in all Trypanosomes including *Leishmania*. In contrast, amastins, leishmanolysin, 3'A2-related proteins, kinase-containing proteins (and several uncharacterized proteins) seemed to be highly abundant in *Leishmania* proteomes compared to all Trypanosomes (Figure 1/B).

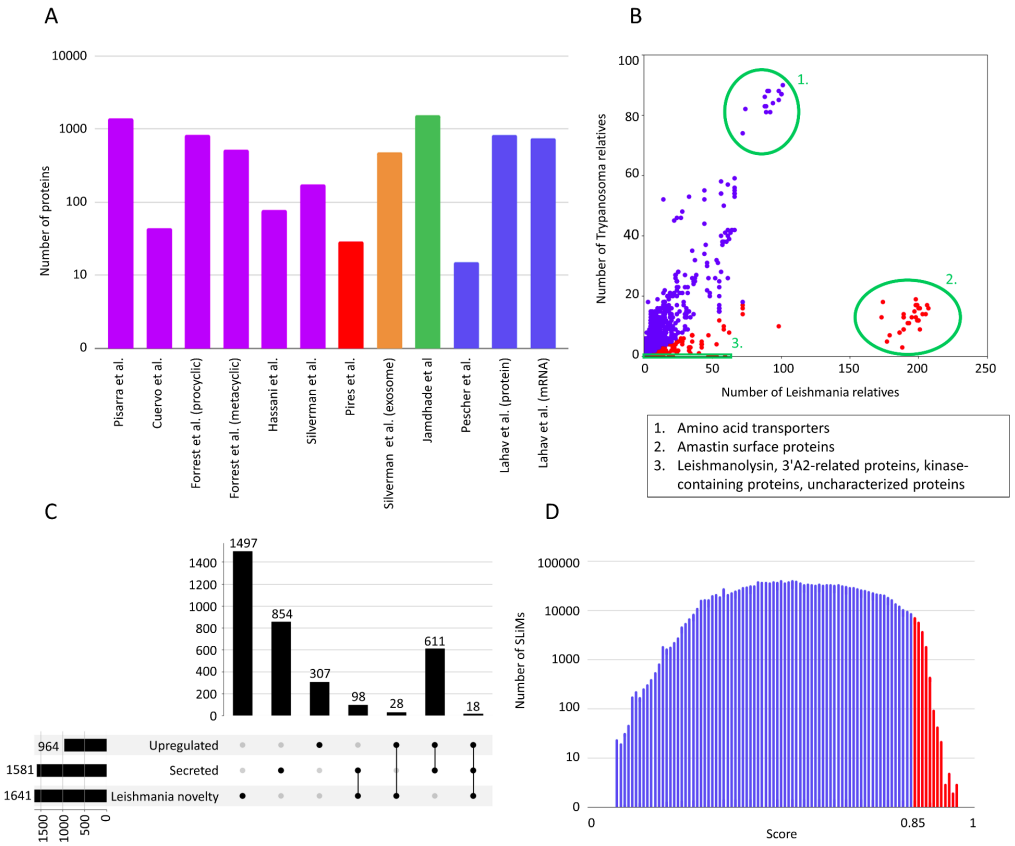

Figure 1: Database content. All data were calculated on Leishmania infantum. A: Number of proteins in different proteomic datasets (purple: promastigote secretome, red: amastigote secretome, orange: exosome, green: housekeeping genes, blue: higher protein abundance level upon infection). B: Number of Trypanosoma and Leishmania close *homologs. Each dot represents a protein (red: at least 80% of* close *homologs are in Leishmania, blue: other proteins). Circles represent distinctive groups. C: Overlap between upregulated, secreted and 'Leishmania novelty' proteins (for more detail see text). D: Distribution of detected SLiMs with different scores. Red marks motifs above 0.85 cutoff (for more details see text).*

AlphaFold2 enables us to detail structural features

We used different methods to predict the structural features of proteins. Classical sequence-based methods can detect globular domains, TM regions and intrinsically disordered regions (IDRs). However, the use of AlphaFold2 (AF2) [29] provides additional structural information. In LeishMANIAdb we used structures available in the AlphaFold database [30] (however we could not map 3192 proteins (~6%)). We not only display the predicted 3D structure of the proteins, but also information derived from the AF2 models such as the secondary structures and the position of the lipid bilayer for membrane proteins using the method introduced in the TMAlphaFold database [31]. Although AF2 was originally built to predict protein structure, the scientific community quickly realized it is as much (if not more) efficient to predict protein disorder [31]. To analyze IDRs we display predicted local distance difference test (pLDDT) values and relative surface accessibility from AF2. For IDR prediction in TM proteins, we tailored MemDis [32] to incorporate features from AF2 instead of sequence-based predictors (see Methods).

Short linear motif candidates that may hijack cell regulation (SLiMs)

We scanned Leishmania proteins for SLiMs using the regular expressions stored in the Eukaryotic Linear Motif (ELM) resource [33]. Scanning SLiMs alone would mostly yield false positive hits, so we developed a scoring system that ranges from 0 to 1, that takes into account most information we collected: conservation, protein disorder, localization, expanded proteins, secretion data, protein expression and homology to housekeeping genes (see Methods, Supplementary Material). The weights of these features were different and somewhat arbitrary, however, in the case of *Leishmania* SLiMs it is not possible to construct a benchmark set to evaluate motif scores. In *Leishmania infantum* we detected over a million motifs, from which ~1% (19 584) had a score above 0.85 (Figure 1/D).

The LeishMANIAdb web resource (webpage)

To visualize all the collected and calculated information we developed an open-access resource. In LeishMANIAdb users can search for proteins using their UniProt ID, UniProt Accession, gene name and protein name. We also provide several protein sets as examples to help users browsing the database. Currently proteins are sorted based on 1) species: *L. braziliensis*, *L. donovani*, *L. infantum*, *L. major,* and *L. mexicana*; 2) manual curation data; 3) experimental data: secreted proteins - if contained in any of the secreted sets, protein abundance/mRNA level data, proteins with any kind of experimental data listed above; 4) computationally predicted information: proteins expanded in Leishmania (score > 0.8 - see Methods, Supplementary Material), transmembrane proteins, proteins with high disordered content (at least 70% of disorder), proteins with high-scoring linear motifs (score > 0.85) and novel *Trypanosoma* proteins (proteins without SwissProt close homologs or Pfam domains). After searching (or selecting a protein set) users can further narrow their selection by choosing any other criteria (Figure 2/A). The entry page for proteins consists of 10 sections (they are only visible if they contain data). The “Quick info” displays the protein name, species, cross-references and its number of amino acid residues. Data curation we made is visible under the “Annotations” section. PPIs, localizations and functions are mirrored from close homologous proteins. We also display functional annotations by Jardim et al. [27]. The

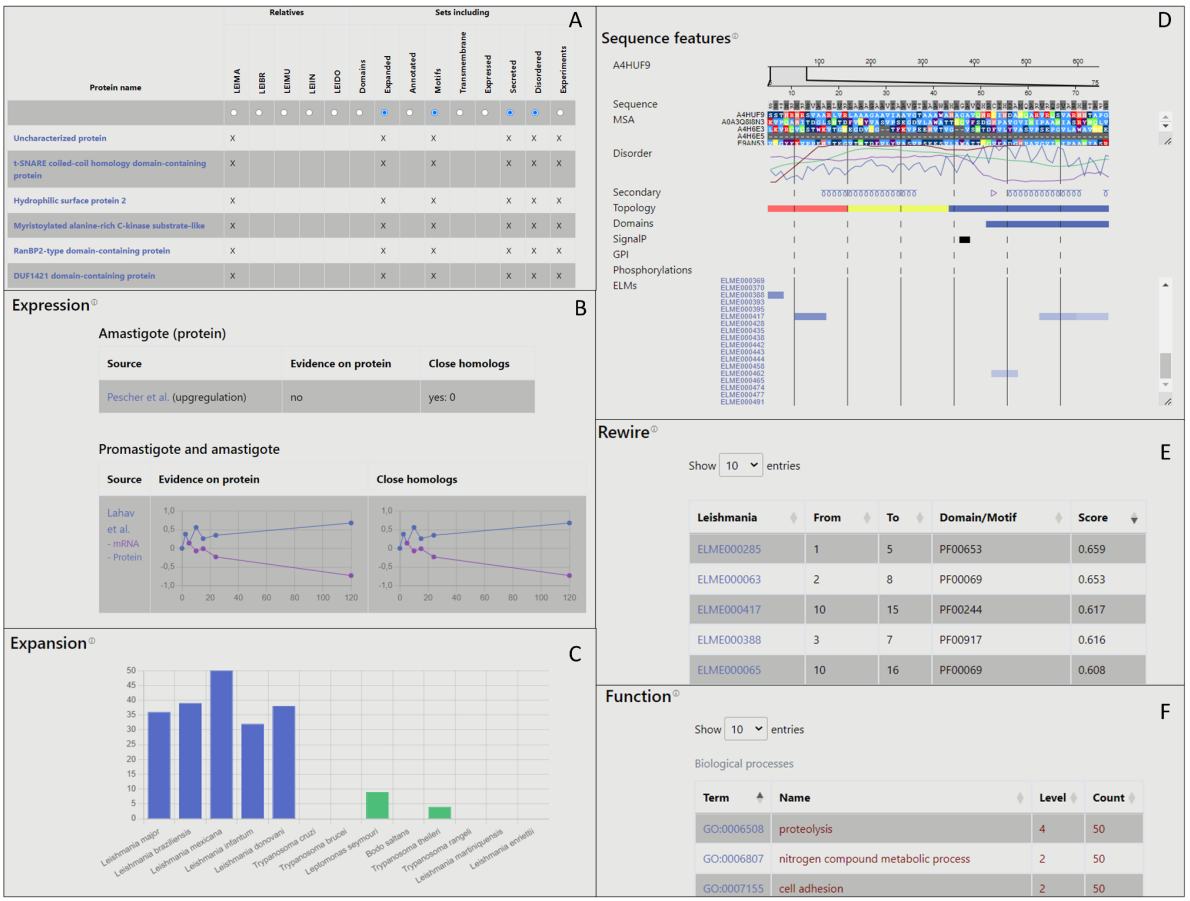

Figure 2: Layout of LeishMANIAdb: A: the search/browse result menu. B: Expression section of the entry page. C: Expansion section. D: Sequence features section. E: Rewire section. F: Function section.

“Localization” section contains all high-throughput experiment data - promastigote and amastigote secretion, an exosome experiment and the glycosome - and predictions are also displayed here. Since the reliability of both predictions and experiments may vary, we also display all this data for close homologous proteins, so users can quickly check the robustness of information. We collected GeneOntology [34] annotations for cellular compartments. In this case, the specificity of the term (how deep it is on the tree) is shown in the “level” column. GO annotations are collected for all close homologous proteins too, and the number of occurrences of each term is displayed. We highlighted terms that are associated with the inspected protein displayed on the page. The “Expression” module can display the expression experiment: static/single point (upregulated or not) or time-course experiments (e.g., mRNA and protein levels available for 7 timesteps across 120 hours, Figure 2/B). In the “Expansion“ all close homologs are displayed by species with a color-code to identify *Leishmania* intracellular and extracellular/free-living relatives (Figure 2/C). The “Sequence features” displays various information (Figure 2/D). On the top, the gapless multiple sequence alignment is visible (the original one can be downloaded) where gaps from the entry protein were removed. Protein disorder, secondary structures, transmembrane topology prediction, domains, signal peptides and glycosylphosphatidyl (GPI) anchors and stage-dependent phosphorylation are also displayed. Predicted SLiMs are shown with a color-coded score (see Methods, Supplementary Material). In the “Structure” section the AF2 predicted structure is available (with the position of the

membrane domain for TM proteins). The “Rewire” section is the table format version of SLiMs from the “Sequence features” module (Figure 2/E). The function section contains GeneOntology Molecular Function and Biological Process terms. As done for the Cellular Component, terms are mirrored from close homologous proteins, and they can be sorted based on their specificity (how deep they are located in the tree) and occurrences considering homologs (Figure 2F). Finally, each result from BLAST search against SwissProt and Trypanosoma relatives are listed in the “Homologs” table.
For each protein (individually, or in batch downloaded for the whole database) the full multiple sequence alignment, high-throughput experiments, annotation and predicted sequential features can be downloaded.

Methods

Resources

Protein sequences were retrieved from UniProt [20,27] and from TriTrypDB [19] based on the UniProt cross-references. Homologs in *Trypanosomes* and in SwissProt were searched with BLAST using e-value: 10^{-5}; sequence identity>20%; coverage>50%. In the “Homologs” section all results are displayed, however for all other sections (and calculation) we only used BLAST results until the first non-Trypanosome SwissProt hit considering sequence identity (termed as “close-homologs”); Further similar proteins were therefore considered as a different homology group. High-throughput experiments were first mapped to the corresponding protein using the identifier provided in the original paper, then mirrored to close Trypanosome homologs based on homology search.
IDRs were predicted using IUPred3 [35] and using the AF2 models’ pLDDT and accessibility values (the latter was calculated by DSSP, normalized using maximum values calculated by Tien et al. [36], threshold defined as suggested by Rost et al. [37]). In the case of TM proteins, IDRs were also predicted by MemDis [32]. In this modified version the PSSMs were generated using Trypanosome sequence library and secondary structure and accessibility were calculated using AlphaFold2. Topology was predicted by CCTOP [38], however to avoid sporadic erroneous predictions, after an initial prediction we performed a constrained iteration where the topologies of homologous proteins were used as a constraint. Using this approach, closely related proteins will always have similar topology. Secondary structure elements were derived from AF2 structures. Pfam domains were identified using InterPro [39]. Protein localization was assigned by GeneOntology [34], predicted by DeepLoc [40] and SignalP6.0 [41]. NetGPI [42] was used to predict GPI-anchors.
To detect SLiMs that may rewire host cell regulation, we used the regular expressions from ELM [33] on all *Leishmania* sequences. We defined different context filters and merged them into a single score to rank motifs (for more details see Supplementary Material): 1) Disordered: The score is the average of the IUPred3, AF2-based pLDDT and accessibility values. All scores were transformed so they range from 0 to 1, with 0.5 being the threshold; 2) Conservation of the motif was checked among close homologs with some permission for slight misalignment, and penalizing motifs that are present across all Trypanosomes; 3) Localization: we used a simplified (intracellular/extracellular) distinction. Motif localization was determined using ELM GO annotations, secretion information and CCTOP prediction, while the domain localization was determined from TOPDOM [43] We looked for pairs where the motif and the domain have the same localization; 4) mRNA level: Using experiments about expression data; 5) protein level: from experiments about protein abundance; 6)

Secretion score based on secretome experiments; 7) Expansion score: reflecting how much the protein is expanded in *Leishmania* species compared to all Trypanosomes; 8) Outgroups score favoring proteins without homologs in SwissProt.

Structure data reflects the version_3 state of the AlphaFold database. All other data was downloaded in October, 2022 from the source databases.

Manual curation

We manually curated hundreds of proteins, using two strategies. First, we searched Google scholar for "*Leishmania* host-pathogen protein interaction" and manually processed the results. Each protein in the experiments was mapped to the corresponding UniProt entry. Then we mapped interaction data to the 5 *Leishmania* proteomes. When the experiment was performed on a protein from different species, we mirrored it to the closest one in LeishMANIAdb, and we also indicated on the webpage that the experiment is from a different protein.
Next we searched for possible surface proteins. For this task we considered all homologs. We collected topology prediction, GPI-anchor prediction, GeneOntology terms and also searched for homologous proteins that were measured [44,45] or predicted to be on the surface. We manually processed the entries using this approach. Notably, this process yielded hundreds of other (than surface) membrane localization annotations.

Web site design

The LeishMANIAdb website is written in PHP (v8.0) using the Laravel (v9.19) framework. All downloaded, predicted or calculated data are stored in a local MySQL (v8.0) database. To visualize sequence features over amino acid sequences, we developed a javascript package using React (18.2), while 3D structures are visualized using the original (for non-TM proteins) or a locally modified version of Mol* [46] for TM proteins. The modified version can display the membrane as two planes around the investigated TM protein using the results of TMDET [47].

Conclusion

The database was developed to be automatically and easily updated when further Leishmania proteomes and secretomes are released. We plan to add new features to the database. Users and readers are welcome to provide feedback and comments on how we could make LeishMANIAdb more useful in their work.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101028908; Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund [K132522]. The authors would like to thank the help of the Gibson team at the European Molecular Biology Laboratory, especially Jesus Alvarado-Valverde for helpful suggestions and discussions, and Philippe Esterre (Pasteur Network, Institut Pasteur, Paris) for his help in selecting a number of Leishmania interactions to be curated.

References

1. Torres-Guerrero E, Quintanilla-Cedillo MR, Ruiz-Esmenjaud J, Arenas R. Leishmaniasis: a review. F1000Res. 2017;6: 750.

2. Gianchecchi E, Montomoli E. The enemy at home: leishmaniasis in the Mediterranean basin, Italy on the focus. Expert Rev Anti Infect Ther. 2020;18: 563–577.

3. McIlwee BE, Weis SE, Hosler GA. Incidence of Endemic Human Cutaneous Leishmaniasis in the United States. JAMA Dermatology. 2018. p. 1032. doi:10.1001/jamadermatol.2018.2133

4. Arango Duque G, Jardim A, Gagnon É, Fukuda M, Descoteaux A. The host cell secretory pathway mediates the export of Leishmania virulence factors out of the parasitophorous vacuole. PLoS Pathog. 2019;15: e1007982.

5. Atayde VD, 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.

6. Ohms M, Ferreira C, Busch H, Wohlers I, Guerra de Souza AC, Silvestre R, et al. Enhanced Glycolysis Is Required for Antileishmanial Functions of Neutrophils Upon Infection with Leishmania donovani. Front Immunol. 2021;12: 632512.

7. Matheoud D, Moradin N, Bellemare-Pelletier A, Shio MT, Hong WJ, Olivier M, et al. Leishmania evades host immunity by inhibiting antigen cross-presentation through direct cleavage of the SNARE VAMP8. Cell Host Microbe. 2013;14: 15–25.

8. Atayde VD, Hassani K, da Silva Lira Filho A, Borges AR, Adhikari A, Martel C, et al. Leishmania exosomes and other virulence factors: Impact on innate immune response and macrophage functions. Cell Immunol. 2016;309: 7–18.

9. Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. Trends in Biochemical Sciences. 2011. pp. 159–169. doi:10.1016/j.tibs.2010.10.002

10. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, et al. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. Chem Rev. 2014;114: 6733–6778.

11. Pellegrini E, Palencia A, Braun L, Kapp U, Bougdour A, Belrhali H, et al. Structural Basis for the Subversion of MAP Kinase Signaling by an Intrinsically Disordered Parasite Secreted Agonist. Structure. 2017;25: 16–26.

12. Tsigankov P, Gherardini PF, Helmer-Citterich M, Späth GF, Zilberstein D. Phosphoproteomic analysis of differentiating Leishmania parasites reveals a unique stage-specific phosphorylation motif. J Proteome Res. 2013;12: 3405–3412.

13. Peysselon F, Launay G, Lisacek F, Duclos B, Ricard-Blum S. Comparative analysis of Leishmania exoproteomes: Implication for host–pathogen interactions. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 2013. pp. 2653–2662. doi:10.1016/j.bbapap.2013.09.015

14. Cuervo P, De Jesus JB, Saboia-Vahia L, Mendonça-Lima L, Domont GB, Cupolillo E. Proteomic characterization of the released/secreted proteins of Leishmania (Viannia) braziliensis promastigotes. J Proteomics. 2009;73: 79–92.

15. Lahav T, Sivam D, Volpin H, Ronen M, Tsigankov P, Green A, et al. Multiple levels of gene regulation mediate differentiation of the intracellular pathogen Leishmania. FASEB J. 2011;25: 515–525.

16. Guay-Vincent M-M, Matte C, Berthiaume A-M, Olivier M, Jaramillo M, Descoteaux A. Revisiting Leishmania GP63 host cell targets reveals a limited spectrum of substrates. PLoS Pathog. 2022;18: e1010640.

17. Gregory DJ, Godbout M, Contreras I, Forget G, Olivier M. A novel form of NF-kappaB is induced by Leishmania infection: involvement in macrophage gene expression. Eur J Immunol. 2008;38: 1071–1081.

18. Fatoux-Ardore M, Peysselon F, Weiss A, Bastien P, Pratlong F, Ricard-Blum S. Large-scale investigation of Leishmania interaction networks with host extracellular matrix by surface plasmon resonance imaging. Infect Immun. 2014;82: 594–606.

19. Shanmugasundram A, Starns D, Böhme U, Amos B, Wilkinson PA, Harb OS, et al. TriTrypDB: An integrated functional genomics resource for kinetoplastida. PLoS Negl Trop Dis. 2023;17: e0011058.

20. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51: D523–D531.

21. Silverman JM, Clos J, de'Oliveira CC, Shirvani O, Fang Y, Wang C, et al. An exosome-based secretion pathway is responsible for protein export from Leishmania and communication with macrophages. J Cell Sci. 2010;123. doi:10.1242/jcs.056465

22. Hassani K, Antoniak E, Jardim A, Olivier M. Temperature-induced protein secretion by Leishmania mexicana modulates macrophage signalling and function. PLoS One. 2011;6. doi:10.1371/journal.pone.0018724

23. Forrest DM, Batista M, Marchini FK, Tempone AJ, Traub-Csekö YM. Proteomic analysis of exosomes derived from procyclic and metacyclic-like cultured Leishmania infantum chagasi. J Proteomics. 2020;227. doi:10.1016/j.jprot.2020.103902

24. Silverman JM, Chan SK, Robinson DP, Dwyer DM, Nandan D, Foster LJ, et al. Proteomic analysis of the secretome of Leishmania donovani. Genome Biol. 2008;9. doi:10.1186/gb-2008-9-2-r35

25. Pissarra J, Pagniez J, Petitdidier E, Séveno M, Vigy O, Bras-Gonçalves R, et al. Proteomic Analysis of the Promastigote Secretome of Seven Leishmania Species. J Proteome Res. 2022;21. doi:10.1021/acs.jproteome.1c00244

26. Jamdhade, Pawar H, Chavan S, Sathe G, Umasankar PK, Mahale KN, et al. Comprehensive proteomics analysis of glycosomes from Leishmania donovani. OMICS. 2015;19. doi:10.1089/omi.2014.0163

27. Jardim A, Hardie DB, Boitz J, Borchers CH. Proteomic Profiling of Leishmania donovani Promastigote Subcellular Organelles. J Proteome Res. 2018;17: 1194–1215.

28. Pescher P, Blisnick T, Bastin P, Späth GF. Quantitative proteome profiling informs on phenotypic traits that adapt Leishmania donovani for axenic and intracellular proliferation. Cell Microbiol. 2011;13. doi:10.1111/j.1462-5822.2011.01593.x

29. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589.

30. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022;50: D439–D444.

31. Dobson L, Szekeres LI, Gerdán C, Langó T, Zeke A, Tusnády GE. TmAlphaFold database: membrane localization and evaluation of AlphaFold2 predicted alpha-helical transmembrane protein structures. Nucleic Acids Res. 2023;51: D517–D522.

32. Dobson L, Tusnády GE. MemDis: Predicting Disordered Regions in Transmembrane Proteins. Int J Mol Sci. 2021;22. doi:10.3390/ijms222212270

33. Kumar M, Michael S, Alvarado-Valverde J, Mészáros B, Sámano-Sánchez H, Zeke A, et al. The Eukaryotic Linear Motif resource: 2022 release. Nucleic Acids Res. 2022;50: D497–D508.

34. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019;47: D330–D338.

35. Erdős G, Pajkos M, Dosztányi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. Nucleic Acids Res. 2021;49: W297–W303.

36. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. Maximum Allowed Solvent Accessibilites of Residues in Proteins. PLoS ONE. 2013. p. e80635. doi:10.1371/journal.pone.0080635

37. Rost B, Sander C. Conservation and prediction of solvent accessibility in protein families. Proteins. 1994;20: 216–226.

38. Dobson L, Reményi I, Tusnády GE. CCTOP: a Consensus Constrained TOPology prediction web server. Nucleic Acids Res. 2015;43: W408–12.

39. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. Nucleic Acids Res. 2023;51: D418–D427.

40. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics. 2017;33: 3387–3395.

41. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40: 1023–1025.

42. Gíslason MH, Nielsen H, Armenteros JJA, Johansen AR. Prediction of GPI-Anchored proteins with pointer neural networks. doi:10.1101/838680

43. Varga J, Dobson L, Tusnády GE. TOPDOM: database of conservatively located domains and motifs in proteins. Bioinformatics. 2016;32: 2725–2726.

44. Langó T, Róna G, Hunyadi-Gulyás É, Turiák L, Varga J, Dobson L, et al. Identification of Extracellular Segments by Mass Spectrometry Improves Topology Prediction of Transmembrane Proteins. Sci Rep. 2017;7: 42610.

45. Bausch-Fluck D, Goldmann U, Müller S, van Oostrum M, Müller M, Schubert OT, et al. The in silico human surfaceome. Proceedings of the National Academy of Sciences. 2018. doi:10.1073/pnas.1808790115

46. Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, et al. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. Nucleic Acids Res. 2021;49: W431–W437.

47. Tusnády GE, Dosztányi Z, Simon I. TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. Bioinformatics. 2005;21: 1276–1277.