Protein abundance in the midgut of wild tsetse flies (Glossina palpalis palpalis) naturally infected by Trypanosoma congolense s.l.

Jean Marc Tsagmo, Flobert Njiokou, Alexis Dziedziech, Valerie Rofidal, Sonia Hem, Anne Geiger

Received: 28 December 2022 | Accepted: 8 June 2023

DOI: 10.1111/mve.12676

ORIGINAL ARTICLE

Medical and Veterinary Entomology

Protein abundance in the midgut of wild tsetse flies (*Glossina palpalis palpalis*) naturally infected by *Trypanosoma congolense* s.l.

Jean Marc Tsagmo[1,2,3] | Flobert Njiokou[2] | Alexis Dziedziech[4] | Valerie Rofidal[5] | Sonia Hem[5] | Anne Geiger[1,2,6]

[1]INTERTRYP, Institut de Recherche pour le Développement, University of Montpellier, Montpellier, France

[2]Faculty of Science, University of Yaoundé I, Yaounde, Cameroon

[3]Trypanosome Transmission Group, Trypanosome Cell Biology Unit, Department of Parasites and Insect Vectors and INSERM U1201, Institut Pasteur, Paris, France

[4]Biology of Host-Parasite Interactions Unit, Department of Parasites and Insect Vectors, Institut Pasteur, Paris, France

[5]IPSiM, Univ Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France

[6]Center for Research on Filariasis and Other Tropical Diseases (CRFilMT), Yaounde, Cameroon

Correspondence
Jean Marc Tsagmo, Trypanosome Transmission Group, Trypanosome Cell Biology Unit, Department of Parasites and Insect Vectors and INSERM U1201, Institut Pasteur, Paris, France.
Email: jean-marc.tsagmo-ngoune@pasteur.fr

Funding information
ARTS fellowship from the Institut de Recherche pour le Développement (IRD); Région Languedoc-Roussillon-Appel d·Offre «Chercheur d·Avenir 2011

Abstract

Tsetse flies (*Glossina* spp.) are major vectors of African trypanosomes, causing either Human or Animal African Trypanosomiasis (HAT or AAT). Several approaches have been developed to control the disease, among which is the anti-vector Sterile Insect Technique. Another approach to anti-vector strategies could consist of controlling the fly's vector competence through hitherto unidentified regulatory factors (genes, proteins, biological pathways, etc.). The present work aims to evaluate the protein abundance in the midgut of wild tsetse flies (*Glossina palpalis palpalis*) naturally infected by *Trypanosoma congolense* s.l. Infected and non-infected flies were sampled in two HAT/AAT foci in Southern Cameroon. After dissection, the proteomes from the guts of parasite-infected flies were compared to that of uninfected flies to identify quantitative and/or qualitative changes associated with infection. Among the proteins with increased abundance were fructose-1,6-biphosphatase, membrane trafficking proteins, death proteins (or apoptosis proteins) and SERPINs (inhibitor of serine proteases, enzymes considered as trypanosome virulence factors) that displayed the highest increased abundance. The present study, together with previous proteomic and transcriptomic studies on the secretome of trypanosomes from tsetse fly gut extracts, provides data to be explored in further investigations on, for example, mammal host immunisation or on fly vector competence modification via para-transgenic approaches.

KEYWORDS
differential protein expression, natural fly infection, trypanosomiasis, tsetse fly, vector competence control

INTRODUCTION

Human and Animal African Trypanosomiasis [respectively HAT and AAT (also called Nagana)], also known as sleeping sickness, are parasitic diseases caused by curved flagellate protozoa of the genus *Trypanosoma*. Human African Trypanosomiasis, in its chronic form, accounts for 97% of the reported cases and is caused mainly by the species *Trypanosoma brucei gambiense* (Tbg). It is transmitted by several tsetse fly species from the group Palpalis (including *Glossina palpalis gambiensis* (Gpg) and *G. fuscipes fuscipes*) found in 24 countries in West and Central Africa. The acute form, accounting for about 3% of reported cases, is caused by the species *Trypanosoma brucei rhodesiense* (Tbr), and transmitted by tsetse flies belonging to the group Morsitans (including *G. morsitans morsitans*; *G. pallidipes*) found in

13 East African countries. While HAT is caused by Tbg and Tbr, the animal disease, AAT, is caused by three other species: *Trypanosoma brucei brucei* (Tbb), *Trypanosoma vivax* (Tv) and *Trypanosoma congolense* (Tc) transmitted mainly by *G. palpalis palpalis* (Gpp); the disease induces major economic losses to agriculture estimated to be around 5 billion US Dollars per year and therefore represents a huge obstacle to the development of endemic countries (WHO, 2017).

Completion of the parasite cycle requires successive infection in two hosts, a vertebrate host (including humans) and the tsetse fly *Glossina* sp. (which is both host and vector to the parasite). The parasite undergoes several stages of development in these different hosts before reaching the fly salivary glands where the metacyclic stage, the only infectious form to vertebrates, occurs. Thus, a method to control the disease could consist of preventing the fly's ability to acquire and host the parasite or interrupting a step of the life cycle necessary for parasite development in the fly. This anti-vector approach could be an alternative or complementary approach to existing control methods. These methods include insecticide-impregnated traps (tiny targets) and screens used in HAT control, as well as the sterile insect technique (SIT)—releasing γ-irradiated male *Glossina* flies—used in AAT control (Vreysen et al., 2014). Moreover, SIT presents limits since during the insect's lifespan; these irradiated vectors may still acquire the parasite (when taking a blood meal on a trypanosome-infected animal) allowing transmission and development of disease in the vertebrate host. Thus, SIT applications could benefit from methods that prevent the development of trypanosomes in these released sterile male flies.

It was shown that, in the field as well as in experimental designs, most of the tsetse flies are naturally refractory to trypanosome infection (Aksoy et al., 2003; Farikou et al., 2010; Njiokou et al., 2004, 2006; Rio et al., 2004; Simo et al., 2008; Tsagmo Ngoune et al., 2017). In the frame of an anti-vector strategy that would consist of reducing the fly vector competence, a molecular approach has been developed to decipher, at the midgut level, the molecular dialogue occurring in this tripartite system—the fly, its symbionts and the trypanosome. The objective was to identify genes for which expression was modified—and consequently the abundance of the proteins they encode—in response to fly infection. Comparing transcriptomes of insectary-raised *G. palpalis gambiensis* (Gpg) flies experimentally infected or not infected with *T. brucei gambiense* (Tpg) allowed for the identification of a number of differentially expressed genes (DEGs) from the fly and its symbionts (Hamidou Soumana, Loriod, et al., 2014; Hamidou Soumana et al., 2015). Several differentially secreted proteins were characterised (Geiger et al., 2015).

Since previous investigations were performed on insectary flies, we sought to determine what was happening in the wild by comparing the protein secretion of naturally trypanosome infected versus uninfected field flies. This study was carried out on the midgut of Gpp flies that were either infected with *Trypanosoma congolense* s. l. (Tc) or noninfected, and flies were collected and sampled from two HAT/AAT foci, Campo and Bipindi in South Cameroon (Tsagmo Ngoune et al., 2017). Our biological model was adopted for field fly investigations given that Gpp was the only species highly represented among the tsetse fly populations in the two foci (up to 99.3% prevalence in Bipindi). Despite three majors differences between the Gpg/Tbg and Gpp/Tc couples, including (1) *Trypanosoma brucei gambiense* (Tbg) is responsible for human trypanosomiasis, whereas *Trypanosoma congolense* (Tc) is responsible for the animal one; (2) the mature infective forms of trypanosomes are located in the salivary gland for *T. gambiense* while they are located in the proboscis for *T. congolense* (Peacock et al., 2012); and (3) *Glossina palpalis palpalis* (Gpp) is a forest fly while *Glossina palpalis gambiensis* (Gpg) is a river fly (Geiger et al., 2005; Salou et al., 2012), many important biological characteristics are shared, including a number of pathways governing the parasitic process (Tsagmo Ngoune et al., 2017).

Here we report on 3291 proteins that were identified with special emphasis on those for which abundance was significantly modified following fly infection. Finally, despite the data being generated from different host/parasite couples and experimental designs, we attempted to normalise and search for similarities between the results from our field-collected flies and those previously reported on insectary-reared flies (Geiger et al., 2015).

MATERIALS AND METHODS

Sampling areas, fly dissection and midgut storage

The tsetse flies used to perform the present investigation were part of those collected in May and June 2015 in two HAT/AAT foci, Campo and Bipindi (including the villages of Ipono, Mabiogo and Campo-Beach; and Lambi, Bidjouka and Ebimin-bang, respectively), located in the southern region of Cameroon (Figure 1). The Campo focus (2°20′ N, 9°52′ E) is located on the Atlantic coast and extends along the Ntem river, and includes some forest areas. The Bipindi focus (3°2′ N, 10°22′ E) has a typical forest bioecological environment, including an equatorial forest and farmland along roads and around villages. Tsetse flies were captured using pyramidal traps placed in suitable tsetse fly biotopes (Lancien, 1981). The collected tsetse flies were identified at the species level on the basis of morphological criteria and adapted taxonomic keys (Grébaut et al., 2004). Next, the samples were separated into two groups of teneral and non-teneral flies. The non-teneral Gpp flies were dissected in a drop of sterile 0.9% saline solution, according to the midgut dissection protocol developed by Penchenier and Itard (1981). Their guts were individually transferred into tubes containing RNA later (Ambion; Texas) and stored at −80°C until DNA and protein extraction processing. All tools were carefully cleaned after the dissection of each fly to prevent cross-contamination. The sample size rationale and calculations are presented in Supporting Information S1.

DNA and protein extraction

Samples stored at −80°C were thawed at room temperature, dried and extracted from RNA later. The midguts were treated with the Nucleospin TriPrep extraction kit (MACHEREY-NAGEL; Düren,

FIGURE 1 The geographical localisation of the Campo and Bipindi HAT/AAT (South of Cameroon) foci where tsetse samples were collected.

Germany) according to the manufacturer's protocol, which allowed for the extraction of both DNA and proteins.

One hundred microliters of DNA Elute solution were used to recover extracted DNA from each sample. DNA quantity and quality were inspected using a nanodrop (Thermo Fisher Scientific Inc, Waltham, MA).

The protein pellet was suspended in 100 µL of Laemmli buffer (Laemmli, 1970) prepared with Complete Mini Protease Inhibitor (Roche; Basel, Switzerland) and Pefabloc SC (Sigma-Aldrich, Saint-Louis, Missouri). Total protein fractions were recovered by centrifugation at 14,000*g* for 10 min at room temperature, and protein concentrations were measured using a 2D Quant kit (GE Healthcare, Chicago, Illinois). The different products obtained were stored at −80°C until use.

Molecular screening of trypanosomes in flies' gut samples

Infected tsetse flies were identified by normal PCR using parasite species-specific primers, as described by Herder et al. (2002). Briefly, a denaturing step at 94°C for 5 min was followed by 44 amplification cycles. Each cycle included a denaturing step at 94°C for 30 s, annealing at 55 °C for 30 s and an extension step at 72°C for 1 min. A final extension was performed at 72°C for 10 min. The amplified products were separated on 2% agarose gel containing ethidium bromide and visualised under UV illumination. Positive (2 ng of reference DNA) and negative controls were included in each set of PCR amplification experiments. PCR amplifications giving a positive result were repeated

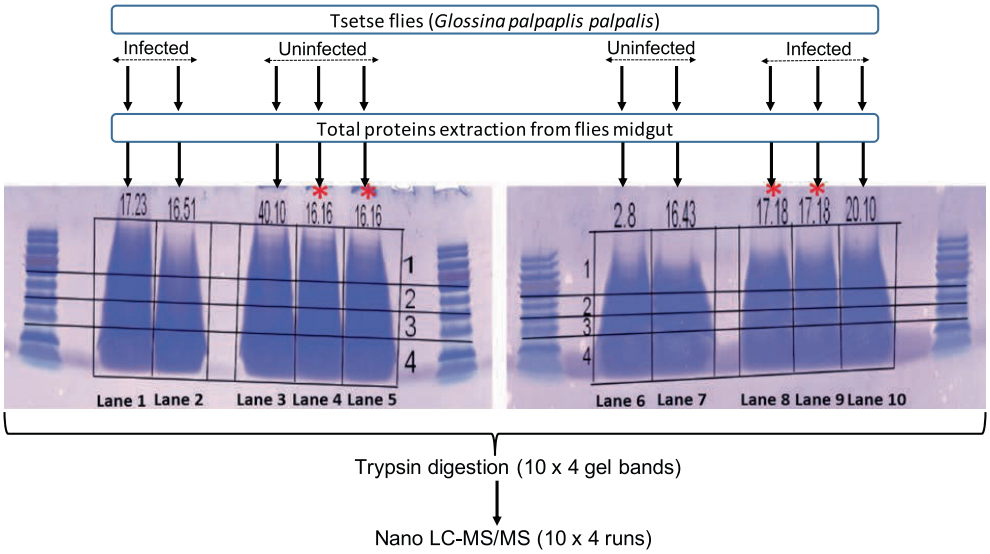

FIGURE 2 Separation of proteins by one-dimensional SDS-PAGE. Lanes 1, 2, 8, 9 and 10: midgut protein replicates from infected flies. Lanes 3, 4, 5, 6 and 7: midgut protein replicates from uninfected flies. Replicates 16.16 (from uninfected flies) and 17.18 (from infected flies) were duplicated; one of each was used to perform preliminary analytic tests.

once for confirmation. As a positive control, we used DNA extracted from parasite species amplified in mice. For negative controls, and to ensure we did not have any contaminations, we used: (1) normal master mix + nuclease-free water (BLANK control); (2) mix without primers + positive control DNA (NO-primer control); and (3) Mix without enzyme + positive control DNA (NO-Taq control).

This process allowed the discrimination of infection status of the fly's gut extracts as either, trypanosome infected or non-infected. Among the extracts from infected flies, only those from *T. congolense* infected flies were used. Finally, four biological replicates were constituted for both *T. congolense* infected flies and non-infected flies. Each replicate was a mix of four randomly chosen midgut extracts from either infected or non-infected flies.

Protein preparation for LC–MS/MS analysis

Twenty micrograms of proteins from each of the 8 replicates were heated at 90°C for 5 min and centrifuged for 5 min at 14,000*g* prior to separation by one-dimensional SDS-PAGE using 10 × 7 cm Tris/glycine PAGE gels (12% acrylamide Mini-PROTEAN TGX precast gels, Biorad; California).

Further processing was similar to that described by Geiger et al. (2015). Briefly, after a 1.5 cm migration into the resolving gel, proteins were fixed in the gel and visualised with Coomassie brilliant blue R-250. Gel pictures were obtained with a high-resolution scanner (Amersham Biosciences; Little Chalfont, UK; Figure 2). Each lane was transversally manually cut into four bands that were individually transferred to 1.5 mL Eppendorf tubes. Bands were first washed with 1 mL of water followed by 1 mL of 25 mM NH_4HCO_3. Destaining was performed twice in the presence of 1 mL of 50% acetonitrile in 25 mM NH_4HCO_3. Gel bands were dehydrated twice in 1 mL of 100% CH_3CN and finally dried at room temperature. Destaining was followed by reducing disulfide bridges with 250 µL of 10 mM DTT at 56°C for 45 min; the supernatant was then removed and cysteine groups were alkylated with 250 µL of 55 mM iodoacetamide for 30 min on a vortex in the dark. Gel bands were washed twice with 1 mL of 50% acetonitrile in 25 mM NH_4HCO_3. Bands were subsequently dehydrated in 1 mL of 100% CH_3CN and dried at room temperature. Twenty microliters of a trypsin solution (Sequencing Grade Modified Trypsin, Promega, Madison) were added to each gel piece at a concentration of 0.0125 µg/µL in 25 mM NH_4HCO_3 and maintained on ice for 15 min. Twenty microliters of 25 mM NH_4HCO_3 were added, and the samples were maintained for another 15 min at room temperature. Protein digestion was performed overnight at 37°C and stopped by the addition of 100 µL of 2% formic acid with sonication in an ultrasonic bath for 10 min. Supernatants containing trypsic peptides were transferred into a 0.1 mL glass insert. The remaining trypsic peptides were extracted twice from bands by the addition of 100 µL of 80% acetonitrile in 2% formic acid. Extracted peptides were pooled in glass inserts and then dried under a vacuum. Peptides were then resuspended in 8 µL of a 2% formic acid solution before LC–MS/MS analysis.

LC–MS/MS analysis

The LC–MS/MS experiments were performed using an Ultimate 3000 nanosystem (Thermo Fisher Scientific Inc; Waltham, MA) interfaced online with a nano easy ion source and a Q-Exactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific Inc, Waltham, MA). The samples were analysed in data-dependent acquisition (DDA).

A volume of 1.6 µL of peptides was first loaded onto a precolumn (Thermo Scientific PepMap 100 C18, 5 µm particle size,

100 Å pore size, 300 µm i.d. × 5 mm length) from the Ultimate 3000 autosampler with 0.1% FA in 2% acetonitrile at a flow rate of 20 µL/min. After a 3-min loading period, the column valve was switched to allow the elution of peptides from the pre-column onto the analytical column. Then, peptides were separated by reverse-phase column (Thermo Scientific PepMap C18, 3 µm particle size, 100 Å pore size, 75 µm i.d. × 25 cm length) at a flow rate of 300 nL/min using a three-step gradient—the amount of solvent B was increased (0.1% FA in 90% acetonitrile.) by 5%–15% in 70 min, then from 15% to 25% between 70 and 85 min, followed by 25%–50% from 85 to 96 min.

Peptides were transferred to the gaseous phase with positive ion electrospray ionisation at 1.7 kV, and the top 10 precursors were acquired between 350 and 1500 m/z with a 2 Th (Thomson, San Diego, CA) selection window, dynamic exclusion of 60 s, normalised collision energy (NCE) of 27 and resolutions of 70,000 for MS and 17,500 for MS2. Spectrum was recorded with Xcalibur software (3.0.63; Thermo Fisher Scientific Inc, Waltham, MA).

Protein identification and quantification

The *.raw* files were analysed with MaxQuant (Tyanova et al., 2016) version v 1.5.0.0 using default settings. The minimal peptide length was set to seven. The criteria 'Trypsin/P' (which means C-terminus peptides of 'K/R' unless followed by 'P': 'K/R' followed by 'P' cannot be a cleavage site) was chosen as the digestion enzyme. Carbamidomethylation of cysteine was selected as a fixed modification and oxidation of methionine and acetylation (protein N terminus) as variable modifications. Up to two missed cleavages were allowed. The mass tolerance for the precursor was 20 and 4.5 ppm for the first and the main searches, respectively, and for the fragment ions was 20 ppm. The files were searched and a global database was built (759,499 entries) compiled from *Drosophila*, *Ceratitis*, *Glossina morsitans*, *Glossina palpalis*, *Trypanosoma brucei gambiense*, *Trypanosoma brucei brucei*, *Trypanosoma congolense*, *Wolbachia*, *Sodalis*, *Wigglesworthia*, *Spiroplasma*, *Homo sapiens*, *Sus scrofa*, *Bos Taurus* and *Ovis aeris* protein databases from UniProtKb (2022-02).

Identified proteins were filtered according to the following criteria: at least two different trypsin peptides with at least one unique peptide, an *E* value below 0.01 and a protein *E* value smaller than 0.01 were required. Using the above criteria, the rate of false peptide sequence assignment and false protein identification were lower than 1%.

Proteins were quantified by label-free method with MaxQuant software using unique and razor peptide intensities (Cox & Mann, 2008). Statistical analyses were carried out using RStudio package software.

The protein intensity ratio (protein intensity in infected fly/protein intensity in uninfected fly) was tested for significant differences in protein abundance using a *t*-test. Hits were retained if they were quantified in at least three of the four replicates in at least one experiment. Proteins with a significant ($p < 0.05$) quantitative ratio (above 1.2 or below 0.8) were considered as significantly upregulated and downregulated, respectively.

Functional annotation

The Gene Ontology (GO) annotation of proteins was performed using the UNIPROT Mass Spectrometry Data Analysis Site (The UniProt Consortium, 2017). Protein denomination homogenising and alphabetical classifying allowed an easy visualisation of the isoforms of given proteins—they are distinguished from each other by distinct accession numbers. Further, this classifying process allowed discrimination of subunits for a given protein (e.g., for the subunits of the eukaryotic translation initiation factor 3), or of proteins involved in multiproteic complexes (such as proteasome), or in the structure of biological particles (such as ribosomes).

RESULTS

Epidemiological data

The overall epidemiological data were presented by Tsagmo Ngoune et al. (2017). A total of 1991 flies were trapped (775 in Campo; 1216 in Bipindi); most of them belonged to the *G. p. palpalis* species (95.61% of the flies trapped in Campo; 99.33% of those from Bipindi). The other fly species identified in Campo were composed of *Glossina caliginea* (2.06%), *Glossina palicera* (1.87%) and *Glossina nigrofusca* (0.52%), while the only other fly species identified in the Bipindi focus was *G. palicera* (0.66%). Over 99% of the identified species were non-teneral. Out of the 1991 flies, 1245 were dissected, the others were either teneral (10) or desiccated (736). Among 1245 dissected flies, a total of 421 were screened by traditional PCR. Out of the 337 Campo flies analysed, 25 (7.41%) were infected by the *Trypanosoma congolense* 'forest type', 16 (4.74%) by the *Trypanosoma congolense* 'savanah type', and 14 (4.15%) by both parasites. In contrast, Bipindi flies only carried the *Trypanosoma congolense* 'forest' type (8.33%). During this epidemiological field work, we also searched for other species of trypanosomes (*Trypanosoma vivax* and *Trypanosoma brucei*) using specific primers for each of them, but neither of these species were identified in the processed flies.

Identified proteins

Based on mass spectrometry raw data analyses, 3291 proteins have been identified. As described in the Materials and Methods section, the proteins extracted from the flies' gut were identified with reference to the proteins listed in our compiled database that associates several UniProt organism-specific data. Table 1 shows the number of identified proteins with reference to the proteins listed in different databases. As expected, most of the proteins (2728) matched with reference proteins from the *Glossina* species; 297 and 104 proteins matched with *Drosophila* and *Ceratitis* reference proteins, respectively. Some proteins matched with *Wigglesworthia* reference proteins (*Wigglesworthia* being the obligate *Glossina* symbiont), and only a few with

TABLE 1 Number of proteins identified with reference to the different databases and percentage of the total number of proteins.

Databases	Number of proteins identified (%)
Glossina	2728 (82.89)
Drosophila	297 (9.02)
Ceratitis	104 (3.16)
Bos taurus	43 (1.31)
Homo sapiens	42 (1.28)
Ovis aries	23 (0.70)
Sus scrofa	17 (0.51)
Wigglesworthia	17 (0.51)
Trypanosoma	11 (0.33)
Spiroplasma	4 (0.12)
Sodalis	3 (0.09)
Wolbachia	2 (0.06)
Total	3291

Spiroplasma, *Sodalis*, and *Wolbachia* (non-obligate *Glossina* symbionts) reference proteins. Finally, some proteins were identified with reference to four mammals (*Bos Taurus*, *Homo sapiens*, *Ovis aries*, and *Sus scrofa*) chosen because tsetse flies feed frequently on these mammals, thus ingesting their blood and the proteins they contain.

Table S1 shows the 3291 identified proteins and their quantification (Label Free Quantification [LQF]) in each of the eight replicates (four from infected and four from non-infected flies). Proteins were alphabetically classified after previous homogenization of their nomenclature to avoid artefacts due to synonyms. Column C provides the name of the organisms for which reference proteins were listed and their corresponding database. This allowed the identification of proteins extracted from the tsetse flies. Overall, a very large number and diversity of proteins could be identified. However, among these 3291 proteins, 1862 (56.5%) were uncharacterized. Most of these proteins are referred to as 'uncharacterized proteins', some as 'hypothetical proteins', and some others still are identified with an alphanumerical identifier.

To get a better insight into the protein diversity, non-characterised proteins were discarded; the 1429 characterised proteins are shown in Table S2. Among the 1429 characterised proteins, 114 (8.1%) were mitochondrial. The number of proteins, including their isoforms if any, involved in biological structures, in multifunctional processes or even in a given single catalytic activity appeared to be highly diverse. For example, we identified 103 (7.2%) ribosomal proteins, including 30 mitochondrial (29.1% of the overall proteins identified as mitochondrial); 30 proteolytic enzymes; 26 eukaryotic translation initiation factor subunits (including 2 isoforms of each of the A, D, F, G, H and M subunits); 17 proteasome subunits, including 10 and 7 isoforms of the alpha and the beta subunits, respectively; 16 isoforms of the peptidyl-prolyl isomerase; 8 serpins (serine protease inhibitors). Finally, we identified 231 protein 'species' (i.e. displaying a specific biological function such as a given enzyme activity) displaying 2–16 isoforms, for a total of 653 isoforms; 694 proteins did not share any isoform.

Protein quantification

Among the 3291 identified proteins, only 1818 passed the different quantification filters and thus were significantly quantified (Table S3). Proteins from the flies' symbionts, *Sodalis glossinidius*, *Wigglesworthia glossinidia* and *Wolbachia*, as well as from trypanosomes were poorly represented, and most of them could not be quantified. Also, the haemoglobin was poorly represented.

Differentially expressed proteins

The analysis was performed on the 1818 quantified proteins, both on characterised and uncharacterized proteins. After statistical analysis, the proteins whose abundance was significantly (Student's *t*-test, *p*-value <0.05) increased or decreased in infected versus non-infected flies were identified and the ratio 'infected/non-infected' was calculated for each protein. After the application of an additional filter to consider only proteins whose ratio variations were either ≤0.8 (i.e., proteins with decreased abundance in trypanosome-infected flies as compared to non-infected flies) or ≥1.2 (proteins with increased abundance in infected flies), 236 proteins (7.2% and 13% of the identified and quantified proteins, respectively) were identified with significant quantitative changes (Table S4). Among these proteins, 175 displayed decreased abundance and 61 increased abundance. However, 61 out of the 236 (25.8%) differentially expressed proteins were uncharacterized [46/175 (26.3%) underexpressed and 15/61 (24.6%) overexpressed proteins, respectively]. The corresponding overexpressed and underexpressed proteins which were characterised are listed in Table S5.

Figure 3 provides an overview of the different protein groups and how they are built.

The group of 129 characterised proteins with decreased abundance include RNA binding proteins, kinases, actin, ribosomal proteins, endocytosis proteins, oxidoreductases, coronin, Yolk and Ras proteins.

The 46 characterised proteins with increased abundance were shown to be mainly involved in transport, metabolism of carbohydrates and protein degrading activities (proteases and other lytic enzymes). Membrane trafficking proteins have been identified (Troponin, translocase, etc.), as well as several phosphorylases, phosphatases and oxidoreductases.

Functional annotation of proteins with significant abundance variations

Functional annotation was carried out to determine the functions associated with the differentially expressed proteins, particularly regarding fly infection. We classified the differentially expressed proteins according to two main categories: Molecular function and

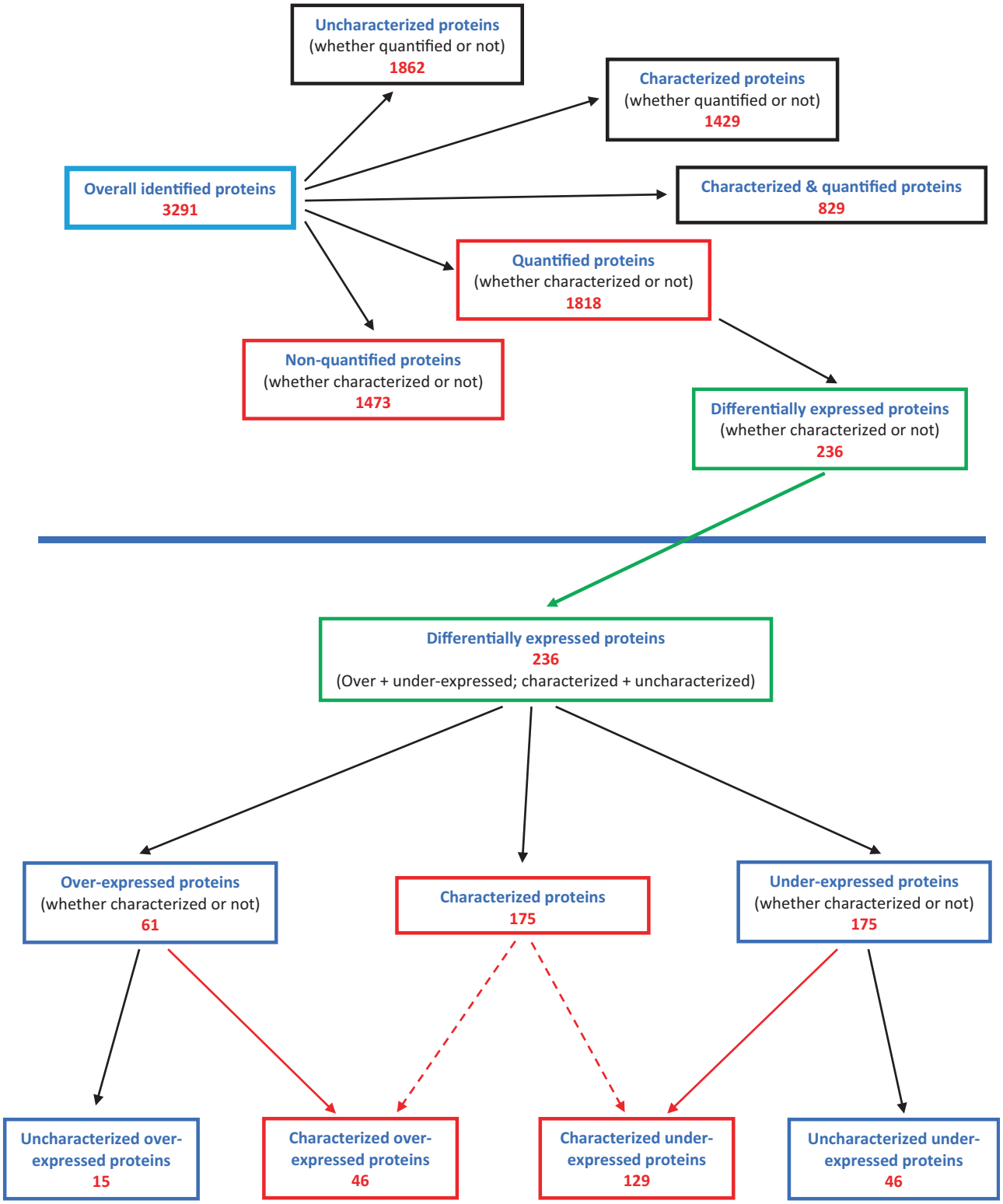

FIGURE 3 Repartition of the proteins extracted from guts of trypanosome-infected or uninfected tsetse flies into different groups and subgroups: the overall proteins, whether or not they were quantified or characterised, detected in the tsetse fly gut extracts; where available they are identified by name and accession number. *Quantified proteins*: proteins whose abundance met the threshold for the different quantification filters. *Characterised proteins*: those that could be identified with reference to proteins listed in diverse databases. *Differentially expressed proteins*: quantified proteins whose abundance in the four replicates from infected flies differed significantly from those recorded in the four replicates from non-infected flies, and whose abundance ratio of infected versus uninfected flies is higher than 1.2 or less than 0.8.

TABLE 2 Functional annotation of the differentially expressed proteins.

Functions	Protein groups					
	All proteins (236)	All characterised proteins (175)	Overexpressed proteins		Underexpressed	
			All (61)	Characterised (46)	All (175)	Characterised (129)
Molecular function						
Unknown	13.2	17.9	22.1	23.4	10	12.4
Binding activity	40.4	37.5	35.8	33.8	42	41.3
Catalytic activity	8.2	9.1	9.5	10.4	7.8	7.8
Enzyme activity	7.1	6.7	5.3	5.2	7.8	8.3
Peptidase activity	7.1	6.2	6.3	6.5	7.4	6
Transcription/translation	6	7.7	9.5	10.4	4.8	5
Redox activity	4.9	4.2	3.2	3.9	5.6	4.6
Transport activity	4.7	4	3.2	2.6	5.2	5.5
Structural activity	4.1	3	–	–	5.6	6
Immune activity	3.3	2.9	3.2	2.6	3.3	3.2
Motor activity	0.8	0.6	2.1	1.3	0.4	–
Biological processes						
Unknown	12	12.8	11.6	13.9	12.1	12.4
Translation	18	18.8	9.3	11.1	21.5	21.6
Biosynthetic process	14	15.8	9.3	11.1	15.9	17.5
Metabolic process	13.3	15	11.6	13.9	14	15.5
Alternative pathway	11.3	12.8	18.6	22.2	8.4	9.3
Transport	8	3.8	14	2.8	5.6	4.1
Protein folding	6	4.5	9.3	8.3	4.7	3.1
Cellular process	5.3	5.3	–	–	7.5	7.2
Cellular organisation	4.7	3.8	4.7	2.8	4.7	4.1
Stress response	4	3.8	7	8.3	2.8	2.1
Redox process	2.7	3	2.3	2.8	2.8	3.1
Proteolysis	0.7	0.8	2.3	2.8	–	–

biological process. The annotation was performed separately on six groups of proteins: (1) the overall 236 differentially expressed proteins (characterised or not); (2) the overall, characterised differentially expressed proteins; (3) the overexpressed proteins (characterised or not); (4) the characterised overexpressed proteins; (5) the underexpressed proteins (characterised or not); (6) the characterised underexpressed proteins. Table 2 shows, for each group, the percentage of proteins involved in different activities or processes.

Regarding molecular function, the annotation of the 236 differentially expressed proteins showed binding activity to be highly prominent since around 40% of the proteins were involved in this activity; 8.2% displayed catalytic activity and 7.1% demonstrated peptidase or enzyme activity. Finally, the function of 13.2% of the proteins remained unknown. This group included proteins whose abundance was either increased or decreased following fly infection. Therefore, we have performed the functional annotation separately on these two groupings and compared the results.

Clear differences were detected regarding biologically relevant activities. While no function could be assigned to 13.2% of the proteins from the entire group of differentially expressed proteins, the functions of 22.1% of the overexpressed proteins and 10% of the underexpressed proteins were unknown. Large differences are also observed between the overexpressed and the underexpressed protein groups regarding binding activity (35.8% vs. 42%) and translation/transcription activity (9.5% vs. 4.8%). In contrast, reduced differences are noted between the 'All' (characterised + uncharacterized proteins) and the 'Characterised' subgroups.

Similarly, regarding the biological processes; 9.3% of all (characterised + uncharacterized) overexpressed proteins were involved in translation, 21.5% of the proteins of the corresponding subgroup of underexpressed proteins were also involved in biological processes. In contrast, the percentages of proteins involved in an alternative pathway and the protein folding processes were much lower in the underexpressed than in overexpressed proteins: 8.4% versus 18.6% and 4.7% versus 9.3%. Here also, there were only a few differences between the two subgroups of the same group, whether overexpressed or underexpressed proteins. These findings are illustrated in Figure 4.

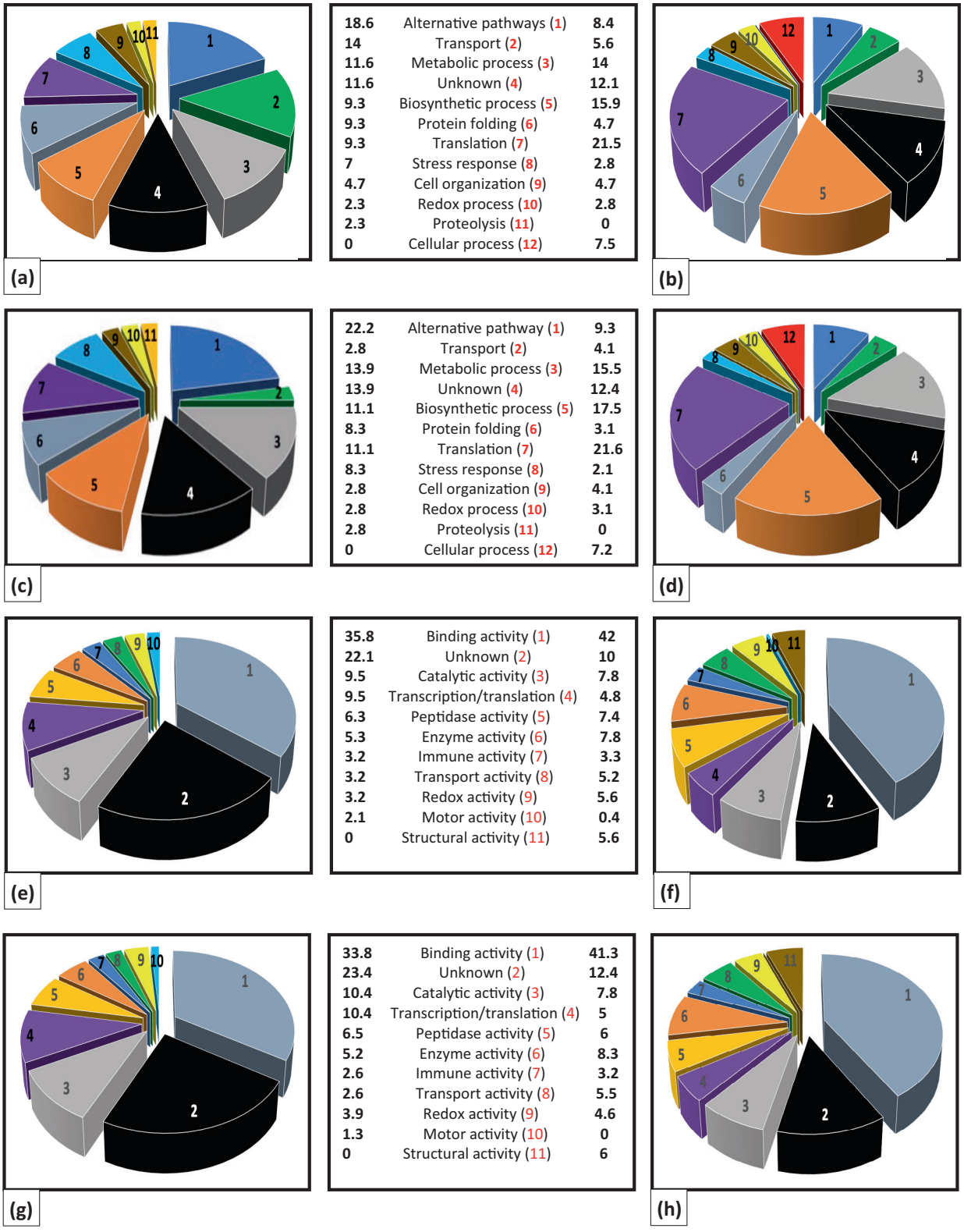

FIGURE 4 Functional annotation (Biological process and Molecular function) performed on four groups of proteins: (a) biological process of all overexpressed proteins; (b) biological process of all underexpressed proteins; (c) biological process of characterised overexpressed proteins; (d) biological process of characterised underexpressed proteins; (e) molecular function of all overexpressed proteins; (f) molecular function of all underexpressed proteins; (g) molecular function of characterised overexpressed proteins; (h) molecular function of characterised underexpressed proteins.

DISCUSSION

The work presented here is a part of the global investigation regarding vector control strategies with aims to understanding their limitations and enhance their efficacy. The goal of this alternative strategy is not only to reduce or eradicate tsetse fly populations but also to reduce or possibly eliminate the fly's vector competence, i.e. its ability to transmit the parasite to humans and animals. Within a population of tsetse flies, about 80% of the individuals are naturally refractory to trypanosome infection, which renders them unable to transmit the parasite. Thus, the question what makes these refractory flies different from their non-refractory neighbours? Differences in environmental conditions cannot explain this difference in competence given the shared environment between refractory and non-refractory flies. This study carried out a molecular approach to identify differences, if any, in gene expression and, consequently, differences in the abundance of the corresponding encoded proteins, between trypanosome-infected and uninfected flies. The objective was to identify genes/proteins associated with fly infection that could become targets in the frame of anti-vector strategies to control trypanosomiasis.

The analysis of events occurring under an uncontrolled system presents several difficulties, which make interpretation of data constricted and subject to unknown external variables. Several factors were unknown such as the age of the sampled flies, the number of blood meals they had before trapping, the duration of infection, and the number of trypanosomes ingested. Further, we were unable to validate 'non-infected' individuals being truly uninfected, or detect newly infected individuals being that early physiological/molecular changes may not have been detectable. One major caveat of this study is that some individuals in the non-infected group could have been refractory to trypanosome infection while others could have been susceptible to infection but were unable to be infected. Regarding this major caveat, the difficulty of identifying a fly's status in a field condition was hindered by our inability to tag individual flies and thus track their status after an initial feeding on an infectious blood meal (to confirm the contact with parasites), and subsequently, follow up on their status 14 days post-infected meal to determine if they were still infected (susceptible) or non-infected (refractory or self-cured). However, transcriptional/translation comparison of infected and non-infected flies helped us to distinguish what was being repressed or upregulated in infected flies during the infection. Therefore, more variability in the results recorded on these field tsetse flies may be expected than in the recorded experimentally controlled experiments. Consequently, we recognise that more robust data with a higher number of replicates per condition would be necessary in the future to consolidate the following conclusions.

Nevertheless, as reported in the results section, 3291 proteins (including isoforms) were identified, that is, about 6-fold more than those identified in extracts from the guts of insectary-reared flies (Geiger et al., 2015). The difference is likely due to the refinement of analytical techniques as well as database updates between the two studies.

Regarding the overall identified proteins, special attention has been paid to those from trypanosomes, the fly-infecting parasite and from *Wigglesworthia*, *Sodalis*, and *Wolbachia*, the three tsetse fly symbionts involved in fly survival, fly infection by trypanosomes causing HAT or AAT and fly reproduction, respectively. Surprisingly, the 1818 quantified proteins included only one trypanosome protein, one protein from *Sodalis*, 8 from *Wigglesworthia* and none from *Wolbachia*. While the absence of these proteins may be plausible for *Sodalis* and *Wolbachia* given their non-obligatory nature as tsetse flies symbionts, it is confounding for *Wigglesworthia*, given its role as the primary symbiont of the tsetse fly, described as essential in every individual fly. There is no clear explanation for the absence of *Wigglesworthia* proteins in our field-collected tsetse flies. This being the first study quantifying the protein secretion of wild tsetse flies (and the symbionts/parasites they harbor), it could be that *Wigglesworthia* doesn't secrete as many proteins in wild flies as they do in the lab-reared ones, or that their secreted proteins were too low and thus below the detection threshold. Confirmation of this observation is thus required from future studies, through, for example, sampling and quantifying the protein secretion of these insects in comparison to the symbionts/parasites they harbor. The question of whether or not *Wigglesworthia* are present in wild caught tsetse flies was previously addressed in our previous study (Tsagmo Ngoune et al., 2019) through the use of metagenomics. We confirmed the presence of *Wigglesworthia* at the expected proportion thus highlighting the question of secretion in our current study. Future studies should investigate the proteome of tsetse flies and their symbionts/parasites, in relation to the detection of *Wigglesworthia* proteins, to confirm the sequencing findings but upstream at transcriptomic level. Examination of the 1473 proteins, whose abundance is too low to be reliably quantified, allows a more complete assessment of the infection versus non-infection scenarios. Finally, a total of 16 *Wigglesworthia*, 10 trypanosomes, 4 *Sodalis* and 2 *Wolbachia* proteins were identified, numbers considered low given their symbiotic importance, especially regarding *Sodalis*, compared to numbers previously recorded on insectary flies (Geiger et al., 2015; Table S6). Previous studies on insectary flies showed that all flies (*G. p. gambiensis*) harboured *Sodalis*, and experimental infection was carried out by feeding the flies on mice displaying a high parasitaemia, allowing the flies to ingest a high number of trypanosomes (Geiger et al., 2005, 2007). In contrast, in the wild, *Sodalis* prevalence is highly variable (Farikou et al., 2010, 2011) as is the parasitemia of the infected host on which the flies may feed.

Other proteins of interest are poorly represented such as caspases (cysteine proteases involved in apoptosis), elongation factors, endonucleases, translation initiation factors, chaperones and heat shock proteins. In addition, we noted the near absence of haemoglobin, and of alpha- and beta-globin in the guts of field flies, although two enzymes (uroporphyrinogen decarboxylase and coproporphyrinogen oxidase) involved in the haemoglobin metabolism were identified. In contrast, haemoglobin and alpha/beta-globin were represented in insectary flies (Geiger et al., 2015). Field flies suck blood on various hosts and it may be that the ingested haemoglobin was almost entirely digested by the time the flies were trapped.

In the frame of the present work, it was of major interest to identify among the 1818 quantified proteins those whose abundance differed significantly depending on whether they were extracted from the guts of infected or uninfected flies. The corresponding 236 proteins are listed in Table S4 (and in Table S5 where the uncharacterized proteins were discarded).

Among the proteins that are most downregulated in infected flies versus uninfected, we found RNA and DNA binding proteins, ribosomal proteins and several translation initiation factors that may disturb normal fly cellular functions necessary in the fly immune response against the invading trypanosome (Aitken & Lorsch, 2012; Jackson et al., 2010). In addition, and in contrast to the previous study (Geiger et al., 2015), glutathione S-transferase (ratio = 0.63), involved in cellular detoxifying (Wojtkowiak-Giera et al., 2011), was shown to be downregulated in infected field flies compared to uninfected flies.

Among the highest upregulated characterised proteins were AhpC (alkyl hydroperoxide reductase C, a thiol-specific peroxidase that catalyses the reduction of hydrogen peroxide), organic hydroperoxides to water and alcohols (involved in cell protection against oxidative stress by detoxifying peroxides; Wang et al., 2013) and a serpin (serine protease inhibitor). Further, six eukaryotic initiation factors (eIF) proteins involved in the initiation of mRNA translation into proteins (Aitken & Lorsch, 2012; Jackson et al., 2010) were characterised of which four showed decreased abundance (eIF1, eIF4F, eIF5A and eIF5C) while two showed increased abundance (eIF3B and eIF3D).

In our study, Serpin A3-5 was the highest oversecreted protein (×8.5). Serpins form a group of serine proteases inhibitors, which act in defence against serine proteases. The serine proteases are considered to be, together with other proteases, trypanosome virulence factors (Bossard et al., 2013). Thus, overexpression of serpins may result from the activation of the host immune response against serine proteases secreted in vivo by the invading parasite as shown by transcriptomic analyses (Hamidou Soumana, Tchicaya, et al., 2014; Matetovici et al., 2016; Ooi et al., 2015). Other proteins are also highly overrepresented, such as xanthine dehydrogenase (×6.6), alkyl hydroperoxide reductase (×4.7), eukaryotic translation initiation factor 3 (×3), proteasome subunit beta (×3) and membrane trafficking protein (×1.7), a key regulator of intracellular membrane trafficking, from the formation of transport vesicles to their fusion with the cellular membrane (Chatterjee & Major, 2001; Ferguson, 1999). Table S4 also revealed that several uncharacterized proteins are underexpressed or overexpressed; they are therefore likely associated with infection of the tsetse fly by trypanosomes. Over time, these proteins will likely be characterised and further annotation of the present dataset may be possible in the near future.

Functional annotation performed on the six groups of differentially secreted proteins showed that both 'Molecular function' (for binding activity, transcription/translation activity and structural activity) and 'Biological process' (for translation, alternative pathway, protein folding, cellular, stress response and proteolysis processes) were differently enriched between overexpressed and underexpressed proteins. The interpretation of these results was difficult given that a protein can be involved in different molecular functions and several biological processes. However, trypanosome infection demonstrated a shift and an imbalance in certain biological functions and pathways. When compared to uninfected flies, more than 40% of the 175 underexpressed proteins in the infected flies were involved in binding (molecular function), providing strong evidence for molecular changes occurring and highlighting the importance of protein binding in infected individuals, likely suppressing metabolic or other processes that are energetically demanding to meet the energetic demand of fighting an infection. In contrast, other proteins involved in this same binding activity were overexpressed; this was the case for 35% of the 61 overexpressed proteins in infected flies versus non-infected, likely aiding in the identification and degradation of foreign pathogens and virulent factors. This observation applies to other functions or biological processes as well signalling either a well or poorly-formed immune response. Regarding alternative pathway processes (biological processes), 8% of the underexpressed proteins were involved compared to 20% of the overexpressed proteins. It is important to note that the abundance of a protein does not necessarily predict the overall effective level of its activity; multiple physicochemical factors may be involved, such as, for an enzyme, the availability (concentration) of its substrate. Moreover, some proteins play key roles which may open the possibility of alternative metabolic pathways; this may occur when two or more enzymes differing in their catalytic specificity compete for the same substrate leading to different products entering different metabolic pathways. The variation in the abundance of a protein is dependent on the variation in its level of biosynthesis, under the control of its corresponding gene(s) whose expression is, in turn, regulated by possibly several factors including epigenetic and pleiotropic effects. Thus, further investigation is necessary to determine the role the involved proteins may play in the framework of the infectious process.

In summation, a comparative analysis of the protein abundance in gut extracts from insectary-reared *G. palpalis gambiensis* flies artificially infected (or uninfected) with *T. brucei gambiense* was previously carried out (Geiger et al., 2015). This offered the opportunity to find some similarities between results recorded under controlled (insectary flies) and uncontrolled (field-collected flies) conditions. The results are only indicative (and not quantitative) since the '2015 study' was performed according to a different experimental design compared to our present study. Table S7 shows the list of characterised proteins, including isoforms (uncharacterized excluded) from field and insectary flies, including their infection status. Finally, Table S8 shows 364 proteins (isoforms excluded) identified from the guts of the field flies that have a homologous protein in the gut extracts from insectary flies. They include a number of proteasome and ribosomal proteins, transporters, binding proteins, oxidoreductases, glutathione S-transferases, chaperones, heat shock proteins, elongation factors, eukaryotic translation initiation factors, proteases (amino peptidases, metalloproteases, serine proteases, etc.) and serpins. This study provides data on several proteins shown to be either underexpressed or overexpressed in infected field flies and associates these proteins with the susceptibility of tsetse flies to trypanosome infection.

CONCLUSION

To conclude, an important objective of the present study was to verify if the molecular events recorded during experimental infections are representative of those occurring in field flies. Our results validate exploring an alternative vector control strategy for fighting sleeping sickness. This approach would consist of reducing, if not altogether eliminating, the vector competence of the tsetse fly. These results provide further data and nuanced information from which to draw for future investigations including in the exploration of mammalian host immunisation or fly vector competence modification. Targeting the vector as a point of control via para-transgenic approaches is a promising direction given the compulsory relationship between the bacteria and the fly. Thus, using *Sodalis* as an in vivo expression vector, we could deliver important transgenic modified immune proteins, such as serpins, and trypanocidal compounds into the fly's gut or compounds stimulating the tsetse fly's immune defences, leading to the increase of the fly refractory rate and/or reduction of the fly's vector competence. Such approaches have been attempted recently, with focus, for example, on the effect of the translationally controlled tumour protein (Bossard et al., 2017, 2021) and provide encouraging results for global efforts against trypanosomiasis.

AUTHOR CONTRIBUTIONS

Jean Marc Tsagmo: Methodology; validation; formal analysis; investigation; writing – original draft; writing – review and editing; visualization; project administration. **Flobert Njiokou:** Methodology; validation; investigation; writing – original draft; supervision. **Alexis Dziedziech:** Writing – original draft; writing – review and editing. **Valerie Rofidal:** Methodology; software; validation; formal analysis; investigation; writing – original draft. **Sonia Hem:** Methodology; software; formal analysis; investigation; resources; data curation; writing – original draft; writing – review and editing. **Anne Geiger:** Conceptualization; resources; validation; writing – original draft; writing – review and editing; visualization; supervision; project administration; funding acquisition.

ACKNOWLEDGEMENT

The authors thank the Mass spectrometry facilities of the Montpellier Proteomics Platform (PPM, BioCampus Montpellier) where the experiments were carried out.

FUNDING INFORMATION

This research was funded by the Région Languedoc–Roussillon–Appel d'Offre 'Chercheur d'Avenir 2011' and the ARTS fellowship from the Institut de Recherche pour le Développement (IRD). The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD034619.

ORCID

Jean Marc Tsagmo https://orcid.org/0000-0001-5327-5313

REFERENCES

Aitken, C.E. & Lorsch, J.R. (2012) A mechanistic overview of translation initiation in eukaryotes. *Nature Structural & Molecular Biology*, 19(6), 568–576. Available from: https://doi.org/10.1038/nsmb.2303

Aksoy, S., Gibson, W.C. & Lehane, M.J. (2003) Interactions between tsetse and trypanosomes with implications for the control of trypanosomiasis. *Advances in Parasitology*, 53, 1–83.

Bossard, G., Bartoli, M., Fardeau, M.-L., Holzmuller, P., Ollivier, B. & Geiger, A. (2017) Characterization of recombinant *Trypanosoma brucei gambiense* translationally controlled tumor protein (r *Tbg* TCTP) and its interaction with *Glossina* midgut bacteria. *Gut Microbes*, 8, 413–427. Available from: https://doi.org/10.1080/19490976.2017.1331833

Bossard, G., Cuny, G. & Geiger, A. (2013) Secreted proteases of *Trypanosoma brucei gambiense*: possible targets for sleeping sickness control? *BioFactors*, 39, 407–414. Available from: https://doi.org/10.1002/biof.1100

Bossard, G., Rodrigues, V., Tour, E. & Geiger, A. (2021) Mice immunization with *Trypanosoma brucei gambiense* translationally controlled tumor protein modulates immunoglobulin and cytokine production, as well as parasitaemia and mice survival after challenge with the parasite. *Infection, Genetics and Evolution*, 87, 104636.

Cairns, D.A., Barrett, J.H., Billingham, L.J., Stanley, A.J., Xinarianos, G., Field, J.K. et al. (2009) Sample size determination in clinical proteomic profiling experiments using mass spectrometry for class comparison. *Proteomics*, 9, 74–86. Available from: https://doi.org/10.1002/pmic.200800417

Chatterjee, S. & Major, S. (2001) The GPI-anchor and protein sorting. *Cellular and Molecular Life Sciences*, 58, 969–987. Available from: https://doi.org/10.1007/PL00000831

Conroy, R.M. (2018) The RCSI Sample size handbook. https://doi.org/10.13140/RG.2.2.30497.51043

Cox, J. & Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26, 1367–1372. Available from: https://doi.org/10.1038/nbt.1511

Farikou, O., Njiokou, F., Simo, G., Asonganyi, T., Cuny, G. & Geiger, A. (2010) Tsetse fly blood meal modification and trypanosome identification in two sleeping sickness foci in the forest of southern Cameroon. *Acta Tropica*, 116, 81–88. Available from: https://doi.org/10.1016/j.actatropica.2010.06.002

Farikou, O., Thevenon, S., Njiokou, F., Allal, F., Cuny, G. & Geiger, A. (2011) Genetic diversity and population structure of the secondary symbiont of tsetse flies, *Sodalis glossinidius*, in sleeping sickness foci in Cameroon. *PLoS Neglected Tropical Diseases*, 5, e1281. Available from: https://doi.org/10.1371/journal.pntd.0001281

Ferguson, M.A. (1999) The structure, biosynthesis and functions of glycosylphosphatidyl-inositol anchors, and the contributions of trypanosome research. *Journal of Cell Science*, 112(Pt 17), 2799–2809. Available from: https://doi.org/10.1242/jcs.112.17.2799

Geiger, A., Hamidou Soumana, I., Tchicaya, B., Rofidal, V., Decourcelle, M., Santoni, V. et al. (2015) Differential expression of midgut proteins in

Trypanosoma brucei gambiense-stimulated vs. non-stimulated *Glossina palpalis gambiensis* flies. *Frontiers in Microbiology*, 6, 444. Available from: https://doi.org/10.3389/fmicb.2015.00444

Geiger, A., Ravel, S., Frutos, R. & Cuny, G. (2005) Sodalis glossinidius (Enterobacteriaceae) and vectorial competence of *Glossina palpalis gambiensis* and *Glossina morsitans morsitans* for *Trypanosoma congolense savannah* type. *Current Microbiology*, 51, 35–40. Available from: https://doi.org/10.1007/s00284-005-4525-6

Geiger, A., Ravel, S., Mateille, T., Janelle, J., Patrel, D., Cuny, G. et al. (2007) Vector competence of *Glossina palpalis gambiensis* for *Trypanosoma brucei* s.l. and genetic diversity of the symbiont *Sodalis glossinidius*. *Molecular Biology and Evolution*, 24, 102–109. Available from: https://doi.org/10.1093/molbev/msl135

Grébaut, P., Mbida, J.A.M., Kondjio, C.A., Njiokou, F., Penchenier, L. & Laveissiére, C. (2004) Spatial and temporal patterns of human african trypanosomosis (HAT) transmission risk in the bipindi focus, in the forest zone of Southern Cameroon. *Vector–Borne and Zoonotic Diseases*, 4, 230–238.

Hamidou Soumana, I., Klopp, C., Ravel, S., Nabihoudine, I., Tchicaya, B., Parrinello, H. et al. (2015) RNA-seq de novo assembly reveals differential gene expression in *Glossina palpalis gambiensis* infected with *Trypanosoma brucei gambiense* vs. non-infected and self-cured flies. *de novo*, 6, 1259.

Hamidou Soumana, I., Loriod, B., Ravel, S., Tchicaya, B., Simo, G., Rihet, P. et al. (2014) The transcriptional signatures of *Sodalis glossinidius* in the *Glossina palpalis gambiensis* flies negative for *Trypanosoma brucei gambiense* contrast with those of this symbiont in tsetse flies positive for the parasite: possible involvement of a Sodalis-hosted prophage in fly Trypanosoma refractoriness? *Infection, Genetics and Evolution*, 24, 41–56. Available from: https://doi.org/10.1016/j.meegid.2014.03.005

Hamidou Soumana, I., Tchicaya, B., Chuchana, P. & Geiger, A. (2014) Midgut expression of immune-related genes in *Glossina palpalis gambiensis* challenged with *Trypanosoma brucei gambiense*. *Frontiers in Microbiology*, 5, 609. Available from: https://doi.org/10.3389/fmicb.2014.00609

Herder, S., Simo, G., Nkinin, S. & Njiokou, F. (2002) Identification of trypanosomes in wild animals from Southern Cameroon using the polymerase chain reaction (PCR). *Parasite*, 9, 345–349. Available from: https://doi.org/10.1051/parasite/2002094345

Jackson, R.J., Hellen, C.U. & Pestova, T.V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews. Molecular Cell Biology*, 11(2), 113–127. Available from: https://doi.org/10.1038/nrm2838

Kanshin, E. & Thibault, P. (2014) Efficient sample processing for proteomics applications—are we there yet? *Molecular Systems Biology*, 10, 758. Available from: https://doi.org/10.15252/msb.20145760

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227, 680–685.

Lancien, J. (1981) Description du piège monoconique utilisé pour l'élimination des glossines en République Populaire du Congo. *Ent. Méd. Parasitol.*, 19, 235–238.

Matetovici, I., Caljon, G. & van den Abbeele, J. (2016) Tsetse fly tolerance to *T. brucei* infection: transcriptome analysis of trypanosome-associated changes in the tsetse fly salivary gland. *BMC Genomics*, 17, 971. Available from: https://doi.org/10.1186/s12864-016-3283-0

Njiokou, F., Laveissière, C., Simo, G., Nkinin, S., Grébaut, P., Cuny, G. et al. (2006) Wild fauna as a probable animal reservoir for *Trypanosoma brucei gambiense* in Cameroon. *Infection, Genetics and Evolution*, 6, 147–153. Available from: https://doi.org/10.1016/j.meegid.2005.04.003

Njiokou, F., Simo, G., Nkinin, S.W., Laveissière, C. & Herder, S. (2004) Infection rate of *Trypanosoma brucei* s.l., *T. vivax*, *T. congolense* "forest type", and *T. simiae* in small wild vertebrates in South Cameroon. *Acta Tropica*, 92, 139–146. Available from: https://doi.org/10.1016/j.actatropica.2004.04.011

Ooi, C.-P., Haines, L.R., Southern, D.M., Lehane, M.J. & Acosta-Serrano, A. (2015) Tsetse GmmSRPN10 has anti-complement activity and is important for successful establishment of trypanosome infections in the Fly Midgut. *PLoS Neglected Tropical Diseases*, 9, e3448. Available from: https://doi.org/10.1371/journal.pntd.0003448

Peacock, L., Cook, S., Ferris, V., Bailey, M. & Gibson, W. (2012) The life cycle of *Trypanosoma* (*Nannomonas*) *congolense* in the tsetse fly. *Parasites & Vectors*, 5, 109. Available from: https://doi.org/10.1186/1756-3305-5-109

Penchenier, L. & Itard, J. (1981) Une nouvelle technique de dissection rapide des glandes salivaires et de l'intestin de glossines. *Ent. Méd. Parasitology*, 55–57.

Rio, R.V.M., Hu, Y. & Aksoy, S. (2004) Strategies of the home-team: symbioses exploited for vector-borne disease control. *Trends in Microbiology*, 12, 325–336. Available from: https://doi.org/10.1016/j.tim.2004.05.001

Salou, E., Rayaissé, J.B., Laveissière, C., Sanon, A. & Solano, P. (2012) Interactions comportementales et rythmes d'activité de *Glossina palpalis gambiensis* et *G. tachinoides* (Diptera: Glossinidae) en galerie forestière au Burkina Faso. *Parasite*, 19, 217–225. Available from: https://doi.org/10.1051/parasite/2012193217

Simo, G., Njiokou, F., Mbida Mbida, J.A., Njitchouang, G.R., Herder, S., Asonganyi, T. et al. (2008) Tsetse fly host preference from sleeping sickness foci in Cameroon: epidemiological implications. *Infection, Genetics and Evolution*, 8, 34–39. Available from: https://doi.org/10.1016/j.meegid.2007.09.005

The UniProt Consortium. (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 45, D158–D169. Available from: https://doi.org/10.1093/nar/gkw1099

Tsagmo Ngoune, J.M., Njiokou, F., Loriod, B., Kame-Ngasse, G., Fernandez-Nunez, N., Rioualen, C. et al. (2017) Transcriptional profiling of Midguts prepared from Trypanosoma/*T. congolense*-positive *Glossina palpalis palpalis* collected from two distinct Cameroonian foci: coordinated signatures of the Midguts' remodeling as *T. congolense*-supportive niches. *Frontiers in Immunology*, 8, 876. Available from: https://doi.org/10.3389/fimmu.2017.00876

Tsagmo Ngoune, J.M., Reveillaud, J., Sempere, G., Njiokou, F., Melachio, T.T., Abate, L. et al. (2019) The composition and abundance of bacterial communities residing in the gut of Glossina palpalis palpalis captured in two sites of southern Cameroon. *Parasites & Vectors*, 12, 151. Available from: https://doi.org/10.1186/s13071-019-3402-2

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T. et al. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, 13, 731–740. Available from: https://doi.org/10.1038/nmeth.3901

Vreysen, M.J.B., Saleh, K., Mramba, F., Parker, A., Feldmann, U., Dyck, V.A. et al. (2014) Sterile insects to enhance agricultural development: the case of sustainable tsetse eradication on Unguja Island, Zanzibar, using an area-wide integrated Pest management approach. *PLoS Neglected Tropical Diseases*, 8, e2857. Available from: https://doi.org/10.1371/journal.pntd.0002857

Wang, H.-W., Chung, C.-H., Ma, T.-Y. & Wong, H.-C. (2013) Roles of alkyl hydroperoxide reductase subunit C (AhpC) in viable but non culturable *Vibrio parahaemolyticus*. *Applied and Environmental Microbiology*, 79(12), 3734–3743. Available from: https://doi.org/10.1128/AEM.00560-13

Wojtkowiak-Giera, A., Wandurska-Nowak, E., Michalak, M. & Derda, M. (2011) Changes in the activity and kinetics of mouse intestinal glutathione transferase during experimental trichinellosis. *Parasitology Research*, 108, 1309–1313. Available from: https://doi.org/10.1007/s00436-010-2205-3

World Health Organization. (2006) The world health report : 2006 : working together for health. *World Health Organization*. 209p.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Supplementary Material S1: Sample size calculation.

Table S1. Listing of the proteins identified in gut extracts from trypanosome-infected and non-infected field-collected tsetse flies. Overall, 3291 proteins were identified among which 1429 were characterised with reference to databases; 1862 proteins were uncharacterized. Quantification (LFQ_Label Free Quantification) was performed on 4 replicates of extracts from infected and non-infected flies. *p*-value results from comparing protein abundance in the 4 replicates from infected versus the 4 replicates from uninfected flies. Among the 3291 proteins, 1818 crossed the quantifying filters (red fonts) and were significantly quantified while 1473 low abundant proteins could not be significantly quantified (black fonts).

Table S2. Listing of the 1429 proteins characterised in gut extracts from field-collected tsetse flies. This Table is extracted from Table S1 with reference to the uncharacterized proteins that were suppressed.

Table S3. Listing of the 1818 quantified proteins from gut extracts from field-collected tsetse flies. This Table is extracted from Table S1 from which low abundant proteins have been discarded.

Table S4. Differentially expressed proteins. The table lists all the differentially expressed proteins. They include 175 underexpressed proteins (129 characterised and 46 uncharacterized) and 61 overexpressed proteins (46 characterised and 15 un-characterised). Proteins are considered as differentially expressed when the abundance ratio (abundance in infected tsetse flies vs. abundance in non-infected flies) is either ≤0.8 (underexpressed proteins) or ≥1.2 (overexpressed proteins). Proteins are classified according to their abundance ratio.

Table S5. Listing of the characterised differential expressed proteins. This Table is extracted from Table S4 from which the un-characterised proteins have been discarded.

Table S6. Comparing trypanosome and symbiont proteins extracted from guts of field (black) and insectary flies (blue). These data are extracted from Table S2.

Table S7. Comparing the identified and characterised proteins extracted from guts of field-collected (black) and insectary-reared (blue) tsetse flies. The proteins have been alphabetically classified. At first, all identified proteins, whether quantified as high or low, were analysed; in a second step 'uncharacterized' and 'hypothetical' proteins, as well as proteins identified with an alpha-numerical identifier, were discarded before sorting. The proteins from insectary flies are those (after elimination of the uncharacterized ones) that were previously identified by Geiger et al. (2015).

Table S8. Proteins that are common to both field-collected (black) and insectary-reared (blue) tsetse flies. The data from this Table have been extracted from Table S7. Isoforms have been discarded.

How to cite this article: Tsagmo, J.M., Njiokou, F., Dziedziech, A., Rofidal, V., Hem, S. & Geiger, A. (2023) Protein abundance in the midgut of wild tsetse flies (*Glossina palpalis palpalis*) naturally infected by *Trypanosoma congolense* s.l.. *Medical and Veterinary Entomology*, 1–14. Available from: https://doi.org/10.1111/mve.12676