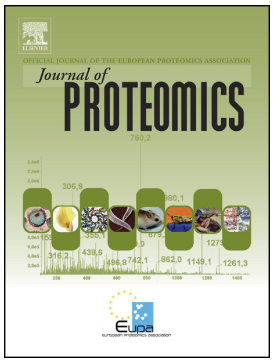
Accepted Manuscript



Revisiting Polybia paulista wasp venom using shotgun proteomics – Insights into the N-linked glycosylated venom proteins

Caroline Lacerra de Souza, José Roberto Aparecido dos Santos-Pinto, Franciele Grego Esteves, Amilcar Perez-Riverol, Luís Gustavo Romani Fernandes, Ricardo de Lima Zollner, Mario Sergio Palma

PII: S1874-3919(19)30093-4
DOI: https://doi.org/10.1016/j.jprot.2019.03.012
Reference: JPROT 3345

To appear in: *Journal of Proteomics*

Received date: 14 January 2019
Revised date: 26 February 2019
Accepted date: 20 March 2019

Please cite this article as: C.L. de Souza, J.R.A. dos Santos-Pinto, F.G. Esteves, et al., Revisiting Polybia paulista wasp venom using shotgun proteomics – Insights into the N-linked glycosylated venom proteins, Journal of Proteomics, https://doi.org/10.1016/j.jprot.2019.03.012

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Research article

Revisiting *Polybia paulista* wasp venom using shotgun proteomics - insights into the N-linked glycosylated venom proteins

Caroline Lacerra de Souza[a], José Roberto Aparecido dos Santos-Pinto[a,*], Franciele Grego Esteves[a], Amilcar Perez-Riverol[a], Luís Gustavo Romani Fernandes[b], Ricardo de Lima Zollner[b] and Mario Sergio Palma[a,*]

[a]Center of the Study of Social Insects, Department of Biology, Institute of Biosciences of Rio Claro, São Paulo State University, Rio Claro, SP, Brazil 13500; [b]Laboratory of Translational Immunology, Faculty of Medicine, University of Campinas (UNICAMP), Cidade Universitária "Zeferino Vaz", Campinas, SP, Brazil 13083887.

*Corresponding authors: Prof. Dr. Mario Sergio Palma, CEIS-IBRC- UNESP, Av. 24A, nº 1515, Bela Vista - Rio Claro, SP, Brazil, CEP 13506-900, E-mail: mspalma@rc.unesp.br and Dr. José Roberto Aparecido dos Santos-Pinto, CEIS - UNESP, Av. 24A, nº 1515, Bela Vista - Rio Claro, SP, Brazil, CEP 13506-900, E-mail: jrbio08@gmail.com

Keywords: Hymenoptera venom; toxins; neurotoxins; allergy; glycosylation; mass spectrometry

Abstract

ACCEPTED MANUSCRIPT

The partial proteome of *Polybia paulista* wasp venom was previously reported elsewhere using a gel-dependent approach and resulted in the identification of a limited number of venom toxins. Here, we reinvestigated the *P. paulista* venom using a gel-free shotgun proteomic approach; the highly dynamic range of this approach facilitated the detection and identification of 1,673 proteins, of which 23 venom proteins presented N-linked glycosylation as a posttranslational modification. Three different molecular forms of PLA1 were identified as allergenic proteins, and two of these forms were modified by N-linked glycosylation. This study reveals an extensive repertoire of hitherto undescribed proteins that were classified into the following six different functional groups: (i) typical venom proteins; (ii) proteins related to the folding/conformation and PTMs of toxins; (iii) proteins that protect toxins from oxidative stress; (iv) proteins involved in chemical communication; (v) housekeeping proteins; and (vi) uncharacterized proteins. It was possible to identify venom toxin-like proteins that are commonly reported in other animal venoms, including arthropods such as spiders and scorpions. Thus, the findings reported here may contribute to improving our understanding of the composition of *P. paulista* venom, its envenoming mechanism and the pathologies experienced by the victim after the wasp stinging accident.

ACCEPTED MANUSCRIPT

1. Introduction

ACCEPTED MANUSCRIPT

Wasp venoms have been described as natural chemical sources of several bioactive components, including low molecular mass compounds, highly abundant peptides, and toxic/allergenic proteins. These components may be used to defend the colony against predators/intruders, as well as for preventing infections with pathogenic microorganisms [1-3]. Since these bioactive components may function as toxins, allergens, hormones, antibiotics, and defensins, many studies have been performed to investigate clinically relevant wasp venoms. The venom components may interact with different pharmacological targets, causing inflammation, blood pressure changes, cardiac arrhythmia, pain and neurotoxicity, among other toxic effects [1,3,4].

The broad range of pathophysiological symptoms experienced by the victims during the envenoming process caused by wasp stinging accidents have increasingly been reported [2,5-7]. Because of this scientific interest, the structural and functional characterization of the wasp venom components may potentially contribute to the rational design of antivenoms [8], novel biopharmaceuticals and biopesticides [2,9].

Currently, a few of the less abundant components, such as neurotoxins, which are commonly described in the venoms of spiders, scorpions and solitary Hymenoptera insects [10,11], have been described in social wasp venoms. Neurotoxic molecules are present in a wide range of animal venoms to aid in prey immobilization/killing; they act on several molecular targets, mainly ion channels, with varying selectivity, specificity and efficiency [3], and display great potential in pharmacological applications. Additionally, omic approaches have been widely used to reveal the toxin arsenal in wasp venoms and have even identified the less abundant molecules, as proposed in the current investigation with *P. paulista* wasp venom.

P. paulista is a clinically relevant neotropical social wasp from Southeastern Brazil. Similar to most social insects in the order Hymenoptera, its venom consists of a complex mixture of bioactive toxins that range from low molecular mass compounds to peptides and proteins [4,12]. The composition of *P. paulista* venom has been investigated; for this purpose, the use of proteomic and peptidomic profiling strategies have facilitated the identification and isolation of a series of novel toxins with pharmacological potential for biotechnological applications [2,4,12,13]. The proteomic profile of *P. paulista* venom was the first to be characterized among the social wasps, and it was determined using a combination of 2-DE, with

ACCEPTED MANUSCRIPT

in-gel digestion, followed by a mass spectrometry analysis [12]. Only 84 proteins were identified, somewhat limiting our understanding of the general envenoming mechanism.

The presence of carbohydrates in protein structures seems to be relatively common among wasp venom toxins [14,15], since glycosylation constitutes a secondary immunogenic determinant for human IgE [16,17]. However, in these studies, only the most abundant components were identified and to date, no systematic study has been conducted to characterize the posttranslational modifications (PTMs), such as glycosylation [18], in the venom proteome of social wasps. Thus, in the current study, we are reporting the results of a gel-free shotgun proteomic study of *P. paulista* venom, including the detection of proteins displaying N-linked glycosylation and evaluated the immunoreactivity of the glycosylated proteins to IgE. The findings of the current investigation may contribute to improving our understanding of the envenoming mechanism caused by the wasp stinging accident and the development of new, specific immunotherapy strategies against wasp venom allergies.

2. Materials and methods

2.1. Polybia paulista venom

Polybia paulista wasps were collected from a georeferenced site (22°23'43.6" S and 47°32'38.9" W) at São Paulo State University at Rio Claro, SP, Southeast Brazil and were immediately frozen and dissected. The venom reservoirs were removed, gently squeezed and washed with a solution containing a 0.2% (v/v) protease inhibitor cocktail (Sigma Fast™ protease inhibitors, Saint Louis, USA) to extract the venom. The samples were processed using methods described by dos Santos-Pinto et al. [15]. The protein concentration was estimated using the Qubit® Protein Assay kit (Life Technologies - Qubit® 3.0 fluorometer), according to the manufacturer's instructions.

2.2. Shotgun proteomic approach

In-solution digestion was used for the shotgun strategy. For this purpose, protein samples (200 µg) were solubilized in 50 mM ammonium bicarbonate, pH 7.9, containing 7.5 M urea, for 2 h to denature the proteins. Samples were processed as described in the study by dos Santos-Pinto et al. [19]. The digested

ACCEPTED MANUSCRIPT

samples were stored at -80 ºC until further analysis; the tryptic peptides were solubilized in 50% ACN/0.5% TFA and submitted to mass spectrometry.

2.3. N-linked Glycosylated protein enrichment

The N-glycosylated proteins were enriched using the Glycoprotein Isolation Kit, ConA (Thermo Scientific™), according to the manufacturer's instructions. Protein samples (1 mg) were solubilized in a solution containing wash buffer and a 0.2% (v/v) cocktail of protease inhibitors (Sigma Fast™ protease inhibitors) and added to spin column containing the ConA lectin resin. The glycoproteins were eluted from the spin column with elution buffer and then submitted to in-solution digestion with trypsin (40 ng/µL, Promega) as mentioned above and described in the study by dos Santos-Pinto et al. [19].

2.4. Mass spectrometry-based proteomic analyses

Peptide analyses were performed using an HPLC system (Shimadzu) coupled to a µLC-ESI-micrOTOF-Q-III (Bruker Daltonics Bremen, Alemanha). Prior to the analyses, the mass spectrometer was calibrated with a Tune-Mix Electrospray Calibrant (Agilent) solution. A C18 column (3 mm id x 100 mm, 2.2 µm, Shim-pack XR-ODS, Shimadzu), was used to separate the peptides with a gradient of 5-95% ACN for 120 min. HyStar v3.2 software (Bruker Daltonics) was used to control the data acquisition and analysis over the mass range of m/z 100-3000. MS spectra were recorded, followed by the acquisition of five data-dependent CID MS/MS spectra generated from the three highest intensity precursor ions. MS/MS spectra were interpreted and peak lists were generated using DataAnalysis 4.1 (Bruker Daltonics).

2.5. Database search of the proteomic data

Database searches were performed using the MASCOT 2.2.06 (Matrix Science, London, UK) against the most recent Hymenoptera protein sequences deposited in the NCBInr database (http://blast.ncbi.nlm.nih.gov, on March 26, 2018). The 1,495,071 entries contained in the taxa Hymenoptera were selected. For the proteins that were not able to be identified within this database, searches were performed against proteins from the taxa Arthropoda, which contains 6,448,695 entries, and against proteins from the NCBInr database, which contains 363,947,122 entries. Search parameters

ACCEPTED MANUSCRIPT

were set as follows: taxonomy, enzyme selected as trypsin, 2 maximum missing cleavage sites allowed, peptide mass tolerance was 0.2 Da for the MS and 0.2 Da for the MS/MS spectra, carbamidomethyl (C) was specified as a fixed modification, while methionine oxidation was specified in MASCOT as variable modifications. Afterwards, the identified proteins were subjected to additional filtering using Scaffold 4.3.2 (Proteome Software Inc., Portland, OR) to validate peptide identification. A false discovery rate (FDR) of less than 1% was calculated by requiring significant matches to at least 2 different sequences. Using the Scaffold Local FDR algorithm, the peptide probability identification was set to a minimum of 99%, while the protein probability identification was set to 95%.

2.6. N-linked glycosylation search

PTM searches were performed using the BioTools 3.2 (Bruker Daltonics) and Peaks studio 7 (complete software for proteomics) software packages. The protein identification and modification data returned by the searchers were manually inspected and filtered to obtain confirmed protein identification and modification lists. A list of 27 common N-glycosylation modifications were selected and applied to virtually cleaved and fragmented peptides that were compared with experimentally obtained MS/MS spectra. The selected modifications were Hexose, Hexosamine, Hex(1)HexNac(1)dHex(1), Hex(1)HexNac(1)NeuAc(1), Hex(1)HexNac(1)NeuAc(2)(N), Hex(1)HexNac(2), Hex(1)HexNac(2)dHex(1), Hex(1)HexNac(2)dHex(1)Pent(1), Hex(1)HexNac(2)dHex(2), Hex(1)HexNac(2)Pent(1), Hex(2)HexNac(2), Hex(2)HexNac(2)dHex(1), Hex(2)HexNac(2)Pent(1), Hex(3), Hex(3)HexNac(1)Pent(1), Hex(3)HexNac(2), Hex(3)HexNac(2)Pent(1), Hex(3)HexNac(4), Hex(4)HexNac(4), Hex(5)HexNac(2), Hex(5)HexNac(4), HexNac, HexNac(1)dHex(1), HexNac(1)dHex(2), HexNac(2), HexNac(2)dHex(1) and HexNac(2)dHex(2).

2.7. Functional enrichment analysis and molecular biological interpretation

The list of identified proteins was analyzed and classified according to Gene Ontology terms (http://geneontology.org/) to understand and interpret the proteomic data. Only the nonredundant proteins were considered for functional analyses, and the proteins classified as uncharacterized, hypothetical and putative were not considered. Based on the functional classification using the GO terms (cellular

ACCEPTED MANUSCRIPT

component, biological process and molecular function) for these proteins, general mechanisms and functions for the identified proteins were proposed.

2.8. Immunoblotting

Sera from 10 *P. paulista*-sensitized patients were obtained at the Anaphylaxis Outpatient Clinic Hospital, Medical Sciences College from Campinas State University - UNICAMP, São Paulo, Brazil, with the approval of the University Ethical Board - CAAE number: 80822817.3.0000.5404. Serum samples were collected at the time of the wasp allergy diagnosis and prior to the administration of any treatment to each patient. Serum samples were stored at -80 °C until use. The entire procedure was performed using the methods reported by Perez-Riverol et al. [20].

2.9. Bioinformatics analysis tools

Sequence alignments were performed using the multiple sequence alignment tool MUSCLE [21] and the B-cell linear epitopes were predicted using Bcepred Prediction Server - Prediction of linear B-cell epitopes, based on the physico-chemical properties [22]. The SWISS-MODEL server [23] was used to produce representative images of the three-dimensional molecular model.

3. Results

Eighty-four proteins were identified in a previous analysis of *P. paulista* wasp venom using a bottom-up gel-based proteomic approach [12], of which 53 proteins were described as typical venom proteins; some of these proteins were glycosylated. However, the carbohydrate moieties of these modifications were not chemically identified. Moreover, some of these proteins were recognized by allergen-specific IgE (sIgE) in sera from patients with a wasp venom allergy [12]. Due to the technical limitations of the bottom-up gel-based approach used in the previous publication, only the most abundant proteins were identified, whereas the rare proteins and those with smaller masses remained undetected.

Considering the limitations of the method that was previously applied to explore the proteome of *P. paulista* venom [12], we decided to reinvestigate this biological system using a gel-free shotgun proteomic approach and determined the glycoprotein profile of this venom. After mass spectrometry

ACCEPTED MANUSCRIPT

analyses and data processing, 1,673 proteins were identified in the *P. paulista* venom. The Supplementary Tables S1 to S12 list all the proteins identified in the present study, while Tables 1 to 3 contain some proteomic data that was specifically extracted from the tables mentioned above to emphasize some proteins of interest in this study.

A specific protein database is not available for any social wasp species in the NCBI database; therefore, the protein searches were initially performed against the Hymenoptera database. For the proteins that were not identified in this database, new searches were performed against Arthropoda and NCBInr databases. All proteins were initially classified according to their functions, as determined using the Gene Ontology (GO) database (cellular component, biological process and molecular function). The results are shown in Figure 1A-C. Only the most abundant GO terms were listed; uncharacterized proteins were not considered in the analyses, since these proteins are designated as presenting "unknown functions". Because the analyzed sample is a secretion from venom glands, the search for functional roles of proteins in GO databases predicts molecular function (when the information is available) based on reports of intracellular roles of these proteins. The roles of secreted proteins are poorly predicted by GO enrichment searches; secreted proteins generally have individual functions in the extracellular environment since these proteins do not typically interact in a manner similar to proteins involved in the steady-state metabolism in the cytoplasm or organelles [19]. Since the GO enrichment search does not reliably predict the functions of intracellular proteins after being secreted into the extracellular medium, we used the molecular functions of the proteins predicted by the GO analyses. Thus, the proteins were grouped into six different functional groups: (i) typical venom proteins; (ii) proteins related to the folding/conformation and posttranslational modifications (PTMs) of toxins; (iii) proteins related to protecting toxins from oxidative stress; (iv) proteins involved in chemical communication; (v) housekeeping proteins from venom glands; and (vi) uncharacterized proteins. Due to the large amount of data generated, the functional annotation of the identified proteins in the GO database (cellular component, biological process and molecular function) and the complete list of functional groups are presented in Supplementary Tables S1-S6. Figure 1D shows the total protein distribution identified in *P. paulista* venom according to their functions. Overall, 7% of the total proteins are typical venom proteins; 4% are involved in the protecting toxins from oxidative stress, 5% in the folding/conformation and PTMs of toxins; 2% are

ACCEPTED MANUSCRIPT

involved in chemical communication; 20% are housekeeping proteins; and 62% of the proteins were identified as “uncharacterized proteins” present in the venom gland.

The first group, the (i) typical venom proteins, comprises toxins, allergens and other proteins that have already been described in the venoms of other social Hymenoptera species. In addition to these, proteins with low abundance, such as some neurotoxin-like proteins described in venoms from solitary arthropods (scorpions, spiders and solitary wasps), were identified and located in this group. One hundred and nine proteins were identified in this group (i); 71 of these proteins, which were identified in the Hymenoptera database, are shown in Table 1 (Table S10 presents all proteomic data), while 38 proteins identified in the Arthropoda and NCBInr databases as being "similar” to the typical venom proteins from other animal species are shown in Table 2 (Table S11 presents all proteomic data). This functional group, which is presented in Tables 1 and 2, was divided into 5 different subgroups and will be discussed according to the functions of the proteins during the envenoming mechanism caused by *P. paulista* stinging accidents: (a) proteins that promote damage to tissue; (b) proteins that promote venom diffusion through the tissues and to the blood; (c) proteins that promote inflammation; (d) allergens; and (e) neurotoxins. Based on the composition of toxins in venom and the clinical importance of *P. paulista* wasp, in the present study, we will focus the discussion on the (i) typical venom protein functional group; thus, this group will be discussed in detail later.

The second functional group (ii) comprises 92 proteins related to the folding/conformation and posttranslational modifications (PTMs) of toxins, such as heat shock proteins, peptidylprolyl isomerase, aminopeptidase, serine/threonine protein kinase, alpha-mannosidase, alpha-glucosidase, beta-galactosidase, protein glycosyltransferase, 3-beta glucuronosyltransferase and alkaline phosphatase. Table S2 presents the complete list of all identified proteins, along with information obtained from the proteomic and GO analyses. These proteins have been described in Hymenoptera venoms and are involved in protein transport among different cellular compartments, the degradation of unstable proteins, protein folding, stability and maintenance of the integrity of toxins, the preservation of structural conformation of these proteins within the venom gland and their functional activities, and proteins related to the PTMs [12,15,24].

ACCEPTED MANUSCRIPT

The third group (iii) included 62 proteins related to protecting toxins from oxidative stress, such as cytochrome oxidase, cytochrome P450, NADH dehydrogenase, NADH-ubiquinone oxidoreductase and thioredoxin peroxidase. Table S3 shows the complete list of all the identified proteins, along with information obtained from the proteomic and GO analyses. These proteins may be responsible for protecting the venom toxins from oxidative stress in the venom gland and have been described in honeybee, wasp and ant venoms [12,24,25]. According to Santos et al. [12], the presence of these proteins in the *P. paulista* wasp venom is also due to the rupture of the venom reservoir during protein extraction and processing.

The fourth group (iv) comprises 32 proteins involved in chemical communication, such as odorant binding proteins, odorant receptor proteins, chemosensory proteins and carboxylic ester hydrolases. Table S4 presents the complete list of all identified proteins, along with information obtained from the proteomic and GO analyses. These proteins were identified for the first time in *P. paulista* venom, and their physiological functions have not been completely elucidated. However, the odorant binding proteins (OBPs) have been described in the venoms of the solitary wasps *Nasonia vitripennis*, *Leptopilina heterotoma*, and *Pteromalus puparum* [26-28]. In the social wasp *Polistes dominulus*, the OBPs and chemosensory proteins are also expressed in antennae, wings and legs [29]. OBPs, chemosensory proteins and pheromone-binding proteins were detected in the venom of the ant *Solenopsis invicta* using a proteomic analysis [25]. These proteins may be related to the mechanisms regulating the complex social behaviors of these ants, since they disperse their venom in nests or along the trails as an orientation cue for foraging [25]. Another protein class identified in this group was the carboxylic ester hydrolases. A group of ester hydrolases was previously reported by Esteves et al. [30] to function as alarm pheromone (medium chained ester) components of secretions from Dufour's gland into the wasp venom.

The fifth group (v) includes 329 housekeeping proteins, which are probably derived from the venom gland cells during venom gland dissection. Table S5 shows the complete list of all the identified proteins, along with information obtained from the proteomic and GO analyses. Although the venom glands mainly contain housekeeping proteins (20% of total proteins), this functional group does not play any apparent role in the envenoming mechanism caused by *P. paulista* stinging accidents. These proteins are related to metabolic functions associated with energy production, possibly to support the intense

ACCEPTED MANUSCRIPT

protein synthesis occurring in the venom gland, and may be contaminants from the material derived from tissues/muscles delivered to the venom during the dissections [12,25].

The last group (vi) included 1,049 uncharacterized proteins. Table S6 presents the complete list of all identified proteins, along with information obtained from the proteomic and GO analyses. These proteins were identified when the protein searches were performed in the Hymenoptera database, and all uncharacterized proteins were identified as belonging to the solitary wasp *Nasonia vitripennis* and the honeybee *Apis mellifera*, which are insects from the order Hymenoptera for which genomic data are available [31,32]. Thus, GO information may be available for these proteins; however, they are referred to as "uncharacterized proteins" due to the incomplete data [33]. Based on our findings, a very large number of proteins in the *P. paulista* wasp venom still have an unknown function.

Considering the importance of glycosylation in animal venoms [17], we determined the glycoprotein profile of the *P. paulista* venom. In the present study, the Glycoprotein Isolation kit, ConA (Thermo Scientific ™), was used to isolate glycoproteins (N-linked glycosylation) that display a strong affinity for concanavalin-A lectin (ConA lectin) immobilized on agarose. In the present study, our proteomic findings identified (Supplementary Table S2) a series of proteins related to glycosylation, such as alpha-mannosidase, alpha-glucosidase, beta-galactosidase, protein glycosyltransferase, 3-beta glucuronosyltransferase and aldose 1-epimerase. These proteins may be involved in N-linked glycosylation of *P. paulista* venom toxins.

After the enrichment of proteins displaying N-linked glycosylation, mass spectrometry analyses and data processing, 546 N-glycosylated proteins were identified. Due to the large amount of data generated, all information obtained from the proteomic analysis and the functional annotation of all the identified proteins in the GO database (cellular component, biological process and molecular function) is provided in Supplementary Table S7. Since the venom composition from *P. paulista* wasp is very relevant and is the focus of this study, we will present glycosylation results focusing on the (i) typical venom protein functional group; thus, this group will be discussed in detail later. Twenty-three N-linked glycosylated proteins that represent typical toxins, allergens and other venom proteins that have already been described in the venoms of social Hymenoptera species were identified. These results are presented in Table 3 (Table S12 presents all proteomic data) and reveal the presence of different molecular forms of

ACCEPTED MANUSCRIPT

some proteins, with different accession codes, i.e., proteins that exhibit small differences in their amino acid sequences and/or posttranslational modifications and thus, may have different structural; however, it is necessary future studies to verify if they present different or similar functions. For example, phospholipase A1 was abundantly detected in our analyses and is present in the venom as six different N-linked glycosylated molecular forms (Table 3); it also seems to occur as three different non-N-linked glycosylated molecular forms (Supplementary Table S8). Meanwhile, Supplementary Table S8 shows all non-N-linked glycosylated proteins that were typically identified in animal venoms (subgroup "i") i.e., venom proteins that have no affinity for ConA lectin immobilized on agarose.

Phospholipase A1 has been identified as one of the major allergens in wasp venoms; thus, we will focus on its N-linked glycosylated molecular forms. Figure 2A-B shows two representative mass spectra that present the assignment of N-linked glycosylation sites for 2 different molecular forms of phospholipase A1 identified in *P. paulista* wasp venom. Thus, figure 2A shows the CID spectrum of the tryptic peptide PFN*EYDILFFVYTR used to assign the N-linked glycosylation as Hex2HexNac2 (Δ = +730.26), with UniProtKB accession number A2VBC4, while figure 2B shows the CID spectrum of the tryptic peptide FN*EYDILFFVYTR used to assign the N-linked glycosylation as Hex2HexNac2Pent1 (Δ = +862.31), with UniProtKB accession number E7D4M5. In mass spectrometry-based proteomics the use of trypsin digestion potentially cleavages exclusively C-terminal to arginine or lysine amino acids. However, in our results as we can note in the Figure 2A-B, were generated a semitryptic peptides due a nonspecific tryptic cleavage in the proline residue, characterizing a chemical fragmentation occured up on collision-induced dissociation MS/MS.

Importantly, the mass spectrometry analysis was performed using CID (collision-induced dissociation) fragmentation, which facilitates the detection of carbohydrates bound to the side chain of an asparagine residue in the case of N-linked carbohydrates. In this case, the spectra do not present the peaks corresponding to the internal fragmentations of the carbohydrate moiety. In addition, the results presented in Table 3 show that the N-linked glycosylation site was not able to be identified for all the glycoproteins detected. These N-linked glycosylated proteins showed affinity for the ConA lectin resin during enrichment, and thus carbohydrates attached to the side chain of some asparagine residues in the sequences of these proteins. However, some of these carbohydrates may have been lost during the mass

ACCEPTED MANUSCRIPT

spectrometry analyses, since the N-linked glycosylation is unstable to moderately stable during MS^n fragmentation [34,35].

Although, conventionally, N-linked glycosylation occurs through the recognition of the consensus sequon, Asn-X-Ser/Thr, it has been reported that protein N-linked glycosylation can also occurs in non-consensus sequons; and may also play significant roles in biological functions [36]. In our scientific outcomes, demonstrated in Table 3, it is possible to observe that the N-linked glycosylation is present in different sequence motifs such as Asn-Ser-Leu, Asn-Cys-Pro, Asn-Lys-Cys, Asn-Leu-Lys, Asn-Thr-Ala and Asn-Leu-Gly, as also reported by Duta et al., [36]. Since, the protein glycosylation may be governed by additional diversifications in sequence motifs and structure, which are yet to be fully explored [36], our results suggest the presence of novel glycosylation motifs in the *P. paulista* venom proteins. Indeed, for a better understanding, further investigations are necessary to address the significance of these atypical motifs.

Our findings also revealed a large number of uncharacterized proteins (Supplementary Tables S6 and S7) identified in *P. paulista* venom possessing some type of N-linked carbohydrate in their sequences; in addition, 36 peptide sequences were detected in the *P. paulista* venom that did not identify a protein in the database search. These sequences were obtained using the *de novo* sequencing approach with the Peaks study 7 software (complete software for proteomics and PTMs) and present different types of N-linked glycosylation, as shown in Supplementary Table S9. As an example, figure 3A and 3B shows two representative mass spectra and the assignment of N-linked glycosylation sites for a peptide from an uncharacterized protein (Figure 3A) and for a peptide from nonidentified protein (Figure 3B). In summary, the *P. paulista* venom may present an even more complex composition, containing proteins that have still unknown functions within the venom gland, or even after they are secreted.

Considering the well-known wasp venom allergenicity to humans, the glycoproteomic profile obtained from the enrichment for N-linked glycosylated proteins in the 1-DE was submitted to immunoblotting using sera from patients with a *P. paulista* venom allergy. Figure 4A shows the electrophoretic profile of proteins obtained from the crude venom prior to enrichment (lane 1), the profile of the enriched N-linked glycosylated proteins from the venom (lane 2), and the profile of the non-N-linked glycosylated protein (lane 3), i.e., the fraction that did not bind to the ConA lectin immobilized on agarose.

ACCEPTED MANUSCRIPT

Figure 4B shows the immunoblotting profile for the proteins that were immunoreactive to human IgE in the crude venom (lane 1). At least seven proteins bands corresponded to allergenic proteins; the N-linked glycosylated protein fraction that was immunoreactive to human IgE is shown in the lane 2, where two protein bands corresponding to N-glycosylated proteins were reactive to human IgE. Meanwhile, lane 3 shows the profile of non-N-linked glycosylated proteins that were immunoreactive to human IgE. This profile revealed two protein bands corresponding to allergenic proteins. The bands of the two the N-linked glycosylated allergenic proteins (Figure 4B, lane 2) were excised from the gel shown in Figure 4A at the corresponding positions (lane 2), subjected to an in-gel with trypsin digestion, and submitted to a proteomic analysis (data not shown). Two PLA1 proteins with an MW of ~35 kDa were identified; the protein in the upper band was identified with the code A2VBC4, and the protein in the lower band was identified with the code E7D4M5. The allergenic protein corresponding to the upper band from lane 3 in the immunoblot of non-N-linked glycosylated proteins (indicated with a white arrow in Figure 4B) was also identified as PLA1 with the accession code A2VBC4. Therefore, three molecular forms of allergenic PLA1s were identified: two (accession codes A2VBC4 and E7D4M5) were identified as glycoproteins and one (accession code A2VBC4) was identified as a non-N-linked glycosylated protein. This result corroborates the data published previously by Santos et al. [12], where it was reported the detection of different molecular forms of phospholipases and some of these molecular forms were glycosylated and immunoreactive to human IgE and others were only immunoreactive to human IgE.

The immunoreactivity pattern comprising two bands for phospholipase A1 molecular forms with very similar mass values was also reported by Perez-Riverol et al. [37] in studies of crude *P. paulista* venom. Tables 1 and 3 show a series of phospholipases A1 molecular forms with MWs ranging from 33-39 kDa that were identified in *P. paulista* venom and are deposited in databases with different accession codes. Thus, the bands present in the N-linked glycosylated protein fraction (Figure 4B, lane 2), which proved to be immunoreactive, are N-linked glycosylated phospholipase A1 molecular forms with different mass values and migrate with a mass of ~35 kDa in the gel. Wasp venom proteins can present a small polymorphic variation even for the same species. The polymorphic variations may be crucial since it provides proteins with different conformational structures. This can suggest that PLA1, for example, may

ACCEPTED MANUSCRIPT

occurs as a series of different forms and/or posttranslationally modified forms with a variety of sequence variations in the different wasp venoms [12].

Figure 4C shows the alignment between the highly conserved amino acid sequences of both molecular forms (accession numbers A2VBC4 and E7D4M5) and the position of each carbohydrate moiety identified to facilitate a comparison of the two phospholipases A1 molecular forms identified in the present investigation. A careful observation of both molecular form sequences revealed that they contained three different types of N-linked glycosylation moieties: the form identified by the code E7D4M5 presented the PTMs Hex2HexNac2Pent1, hexosamine, and hexose, while the form identified by A2VBC4 presented only the moiety Hex2HexNac2. Because both sequences are highly conserved, the molecular form presenting the longer sequence (accession code A2VBC4) was modeled to identify the position of each carbohydrate moiety in its 3-D structure, as shown in Figure 4D. The sequence of the form with the code A2VBC4 was used in the algorithm Bcepred Prediction Server - Prediction of linear B-cell epitopes [22] to simulate the linear epitopes for B-cells. As shown in Figure 4D, all glycosylated positions corresponded to PLA1 linear epitopes (with N-linked carbohydrates) for human IgG and/or IgE.

4. Discussion

In the current investigation, proteins were searched against Arthropoda and NCBInr databases, as well as the Hymenoptera database. This approach allowed us to identify venom toxin-like proteins that are commonly reported in animal venoms such as spiders, scorpions, and even marine organisms, as shown in Tables 2 and S11. As mentioned above, all identified proteins were organized into six different functional groups: (i) typical venom proteins; (ii) proteins related to the folding/conformation and posttranslational modifications (PTMs) of toxins; (iii) proteins related to protecting toxins from oxidative stress; (iv) proteins involved in chemical communication; (v) housekeeping proteins from venom glands; and (vi) uncharacterized proteins. The present manuscript will focus on discussing the functional group of (i) proteins that are typical of animal venoms. Proteins in this group was classified into five different subgroups according to the functions of the proteins during the envenoming mechanism caused by *P. paulista* stinging accidents: (a) proteins that promote damage to tissues in general; (b) proteins that

ACCEPTED MANUSCRIPT

promote inflammation; (c) proteins that promote venom diffusion through the tissues and to the blood; (d) allergens, and (e) neurotoxins.

The first two subgroups (a and b - proteins that promote damage to tissues in general and inflammation) included proteins such as phospholipases (PLA1s and PLA2s), disintegrins/metalloproteinases and protonectin. Phospholipases have been described in Hymenoptera venoms, particularly in wasp and bee venoms [12,20,25,38-40], and usually occur in two subtypes, PLA1s and PLA2s. These proteins exhibit inflammatory activity in tissues, i.e., they disrupt the cell membrane lipid bilayer by catalyzing the specific hydrolysis of ester bonds, converting these substrates into their corresponding lyso compounds through the release of fatty acids, which stimulate tissue damage. The phospholipid packing in several types of biological membranes is disrupted by these enzymes, leading to pore formation, cell lysis, tissue damage and inflammation [12,41]. PLA1s and PLA2s do not display sequence homology; apparently, these proteins have distinct functions. PLA2s have been described to function as anticoagulants in snake venoms due the hydrolysis caused by these proteins, although this function remains unclear [41]. Additionally, PLA2s induce local myotoxicity [42]. We identified different phospholipase molecular forms in *P. paulista* venom. Disintegrins/metalloproteinases are proteins involved in the inflammatory process and systemic reactions that have been described in different animal venoms, such as snakes [43], Hymenoptera species such as the ants *Solenopsis invicta* and *Tetramorium bicarinatum* [25,44] and the wasp *P. paulista* [12]. Another finding reported in the present study was the presence of protonectin, a peptide that was previously described in the study by Baptista-Saidemberg et al. [45] as displaying chemotactic activity and the ability to degranulate mast cells. This peptide disrupts the mast cell membranes, leading to the release of histamine and subsequent vasodilation and inflammation [46].

The third subgroup (c) comprises proteins that promote venom diffusion through the tissues and to the blood, such as hyaluronidase, astacin-like metalloproteases, toxin-like proteins, venom chitinase, venom protease-like proteins, and snake venom metalloprotease-like proteins. We identified different hyaluronidase molecular forms in *P. paulista* venom. This protein is an endo-N-acetylhexosaminidase that functions as a “venom spreading factor” in animal venoms by cleaving the β-1,4 glycosidic bond between

ACCEPTED MANUSCRIPT

N-acetyl-glucosamine and D-glucuronic acid of hyaluronic acid in the extracellular matrix, resulting in an increase in tissue permeability and facilitating the diffusion of the toxic venom components [12,38,44].

The astacin-like metalloprotease toxin-like proteins included in this subgroup have been already described in the venoms of *Loxosceles sp*. spiders [47]. These proteins function by facilitating toxin diffusion into the tissue and may be involved in digesting the prey, since these proteins have been reported in the venom and silk glands of the African spider *Stegodyphus mimosarum* [48]. The protein chitinase, which is also included in this subgroup, can been described as a "dissemination factor" because it may facilitates toxin diffusion by hydrolyzing chitin, the extracellular matrix and other diffusion obstacles [49].

The fourth subgroup (d) comprises several proteins that have been already described to function as allergens in Hymenoptera venoms, such as venom allergen 5, hyaluronidase, acid phosphatase, arginine kinase, heat shock proteins, dipeptidyl peptidase, vitellogenin and PLA1s [12,25,38,39,40,44,50,51-54]. Most of these allergens have been described as glycosylated proteins; however, no study using a proteomic approach to characterize this PTM in wasp venom allergens has been conducted. Glycosylation is one of the most common types of PTMs, and might also be a cause of cross-reactivity in patients who are allergic to Hymenoptera venom [16]. The specific roles of the carbohydrates attached to the wasp venom proteins are unclear, but they may constitute secondary immunogenic determinants for IgE, contributing to the biological activity, solubility, stability, and resistance to proteases [17,55]. In addition to other functions, carbohydrates protect glycoproteins from thermal denaturation and enzymatic digestion. Thus, glycosylation during protein biosynthesis may be involved in the formation and conservation of secondary structures, leading to correct protein folding; this process, thus, plays an important role in maintaining the biological activity and the ability of the glycoprotein to interact with human IgE [56]. Therefore, here, we will focus the discussion on some of the most important allergens that have been described as glycosylated proteins and were identified in the present study.

PLA1s have been widely investigated as allergens in Hymenoptera venoms and were abundantly detected in our analyses. In the study by Santos et al. [12], PLA1s present in crude *P. paulista* venom were immunoreactive to human IgE. In the recent study by Perez-Riverol et al. [37], PLA1s from several clinically relevant wasps have conserved epitopes; according to the authors, these epitopes are the main

ACCEPTED MANUSCRIPT

cause of cross-reactivity *in vitro* based on this allergen. PLA1s represent important markers for an accurate diagnosis of a wasp venom allergy [51]; thus, several researchers have attempted to clone and produce these toxins as recombinant proteins for the development of a CRD (component-resolved diagnosis) of the wasp venom allergy [20,37,61,60].

This allergen has been described to be present as different molecular forms in Hymenoptera venoms, indicating that many PTMs may occur in its sequences [12]. In the current study, PLA1 was identified in the both N-linked glycosylated protein fraction and the non-N-linked glycosylated protein fraction from *P. paulista* venom (Table 3, Supplementary Table S8, and Figures 2 and 4A-B). As mentioned above, this finding is due to the existence of molecular forms present in the venom [39]; some of these molecular forms are N-linked glycosylated, whereas others are not, as shown in Figure 4B. According to the studies by Santos et al. [12], some of these PLA1 molecular forms are glycosylated and were immunoreactive to human IgE.

Many insect venom allergens have been described as glycoproteins that exhibit cross-reactive carbohydrate determinants (CCDs) through the addition of α-1,3-fucose. It is an important, highly immunogenic N-glycosylation moiety because it is directly related to the allergy diagnosis [40,60]. These carbohydrates are not detected in humans but are recognized as antigens when they contact the human immune system. Although considered clinically irrelevant, CCDs are related to cross-reactivity in allergy diagnoses and can generate a false positive result in specific IgE detection tests, since some patients produce antibodies against these CCDs but not specific protein antigens in wasp venom [20,40]. However, in the current study, the N-linked glycosylated allergens, such as phospholipases A1, did not present the α-1,3-fucose that characterizes the presence of CCDs. Our findings are consistent with a recent study reporting a lack of CCDs in *P. paulista* venom allergens [63]. A similar result was previously obtained for *Polistes sp.* wasps, which are phylogenetically similar to *P. paulista*.

Venom allergen 5 is recognized as an important and very common allergen in wasp venoms; however, a physiological function has not been identified for this protein [57]. Despite the importance of this protein as allergen, few studies have been conducted to establish the immunological potential of this protein [12,15,58]. Dos Santos-Pinto et al. [58] reported the mapping of B-cell linear epitopes of venom allergen 5 (Poly p 5) from *P. paulista* venom. In this study, nine immunoreactive human IgG epitopes were

ACCEPTED MANUSCRIPT

identified, one of which was also recognized by human IgE antibodies. In the current study, the different molecular forms of venom allergen 5 did not present N-linked carbohydrates in their sequences (Supplementary Table S8). However, in the previous studies from our group, this allergen from *P. paulista* venom was reported to be a glycoprotein [12] with an O-linked glycosylation site at K141, as well as sites for other PTMs such as hydroxylation and phosphorylation [15].

Hyaluronidases have been reported to function as allergens in scorpion, honeybee and wasp venoms [12,38]. Santos et al. [12] identified four hyaluronidase molecular forms in the *P. paulista* venom that were immunoreactive to human IgE. Although no N-glycosylated hyaluronidase was identified in the present investigation, this type of glycosylation has been described in wasp venom hyaluronidases, as well as in *V. vulgaris*, *D. maculata*, *P. anularis* and *P. paulista* [14,50].

Another protein that was identified in our analyses and included in the allergen subgroup was dipeptidyl peptidase. This protein has already been described as a clinically relevant allergen in the *Apis mellifera* (Api m 5) and *Vespula vulgaris* (Ves v 3) venoms as glycosylated molecular forms [40,59,60]. The biological functions of these proteins in insect venoms during the envenoming mechanism remains unknown. However, by analyzing the enzymatic activity, Blank et al. [59] suggested that Api m 5 and Ves v 3 (allergens corresponding to the dipeptidyl peptidases) were potentially involved in the conversion of pro-toxins to their active forms after venom inoculation in the victim. In the current investigation, this protein was identified for the first time in *P. paulista* venom and was present in both the N-linked glycosylated protein fraction (hexose, Δ = +162.05) and the non-N-linked glycosylated protein fraction (Tables 3, S8 and S12).

This allergen subgroup also included heat shock proteins [12], vitellogenins [52], arginine kinases [12,24], venom serine proteases [44], venom acid phosphatases [54], allergen Pen m 2 [64], allergen Cr-PI and tropomyosin-like proteins [65,66] that were previously described as allergens in venoms from wasps and other organisms. The tropomyosin protein is widely described as being present in shrimp muscle, one of the major food allergens reported in the literature [65,66]. In the current investigation, this protein was identified in *P. paulista* venom both in the N-linked glycosylated protein fraction (hexose, Δ = +162.05; Hex1HexNAc1dHex1, Δ = +511.19; hexosamine, Δ = +161.07; Hex1HexNac2, Δ = +568.21) and in the non-N-linked glycosylated protein fraction (Tables 3, S8 and S12).

ACCEPTED MANUSCRIPT

The fifth subgroup (e) included a series of neurotoxins similar to typical neurotoxins present in various animal venoms. Venomous animals use these toxins as part of their chemical arsenals for predation purposes and territory defense [25,67]. Generally, these proteins are used by solitary predators such as scorpions, spiders and solitary wasps to paralyze the prey and induce death. In the current investigation, we identified 19 neurotoxin-like proteins for the first time in the venom of a social wasp (Table 2). Importantly, these proteins were identified by searching of our experimental data against the Arthropoda and NCBInr databases.

We identified neurotoxin-like proteins that originally described and characterized in the venoms from different genera of spiders, such as *omega-theraphotoxin-Hhn1e-like* (*Haplopelma hainanum*), *alpha-latrotoxin-Lm1a-like* (*Latrodectus mactans*), *delta-ctenitoxin-Pn2c-Neurotoxin Pn2-5A-like* (*Phoneutria nigriventer*), *U2-agatoxin-Ao1f-like* (*Agelena orientalis*) and *U13-theraphotoxin-Cg1a-like* (*Chilobrachys guangxiensis*) [68,69], as well as in venoms from different genera of scorpions, as *alpha-toxin Amm3-like* (*Androctonus mauritanicus mauritanicus*), *toxin Cex12-like* (*Centruroides exilicauda*), *potassium channel toxin Hge-beta-KTx-like* (*Hadrurus gertschi*), *neurotoxin LmNaTx3-like* (*Lychas mucronatus*), *potassium channel toxin alpha-KTx 12.1-like (Tityus serrulatus), toxin Tb2-like (P-beta* NaTx5.3)* (*Tityus bahiensis*) and *potassium channel toxin alpha-KTx 6.16-like* (*Opisthacanthus cayaporum*) [70,71].

Our results also revealed the presence of a *Scolopendra 20528.11 Da toxin-like* protein (*Scolopendra angulata*) that was originally described and characterized in centipede venoms. This neurotoxin-like protein possesses several functions in prey paralysis [72]. The presence of a *Scolopendra* neurotoxin-like was also reported in the studies by dos Santos-Pinto et al. [25] of venom from the ant *Solenopsis invicta*. In addition, an interesting finding was the identification of two conotoxin-like proteins, which constitute a group of neurotoxins typical of *Conus* venoms [73]. These proteins were *conotoxin PnMKLT1-0122* (*Conus pennaceus*) and *conotoxin MaIr34* (*Conus marmoreus*). Corroborating our conotoxin-like protein findings in *P. paulista* venom, the glacontryphan-M peptide was detected with 100% sequence coverage in the wings of the butterfly *Hebomoia glaucippe*, which is structurally and functionally identical to the peptide toxins reported in *Conus* venoms [74]. Moreover, transcripts with highly conserved cysteine residues in their sequences were identified in the venom glands of the ant *Dinoponera*

ACCEPTED MANUSCRIPT

quadriceps; these sequences were very similar to some conotoxins, such as Mü-conotoxin MrVIB that was originally described in *Conus marmoreus* venoms [75]. Notably, we can suggest that these conotoxin-like proteins identified in *P. paulista* venom may are present in their active forms in *P. paulista* wasp venom; although future studies designed to confirm these results are prudent, since the genome of *P. paulista* has not been sequenced.

The presence of neurotoxins in the *P. paulista* venom may be an evolutionary remnant from solitary wasps, which use their venoms to paralyze the nervous system of their prey [1,76]. Thus, neurotoxicity must be considered in the envenoming mechanism caused by *P. paulista* stinging accidents, as presented in figure 5. This schematic summarizes the general envenoming mechanism proposed for the *P. paulista* venom, focusing on the typical venom proteins identified in the present study.

5. Conclusions

In summary, in the present investigation using a gel-free shotgun proteomic approach and N-linked glycosylation enrichment method, we are presenting the identification of a large number of proteins in the venom, including novel toxins and allergens, and provide insights into N-linked glycosylated venom proteins. Glycosylation (N- and O-linked carbohydrates) has been shown to be important in the interaction of the epitope with allergen-specific human IgE. Based on our findings, the biological significance of the several N-linked carbohydrates types identified in *P. paulista* venom proteins may be quite extensive. We speculated that this type of posttranslational modification contributes to increasing the stability of the folded glycoprotein and preventing the interaction of the glycosylated toxins with the victim's immune system; in addition, an N-linked glycosylated protein fraction of the venom represents a potential evolutionarily novel adaptation for the venom to escape the effects of different inhibitors present in the different types of prey. Another interesting contribution of this study is the identification of some neurotoxin-like proteins in the *P. paulista* venom, indicating that neurotoxicity must be considered in the envenoming mechanism of this wasp. Thus, our findings may contribute to improving our understanding of the molecular composition of the venom and the envenoming mechanisms caused by the wasp stinging accident.

Acknowledgments

ACCEPTED MANUSCRIPT

This work was supported by grants from FAPESP (Proc. 2013/26451-9, Proc. 2016/16212-5, Proc. 2017/04680-7, Proc. 2017/10373-0, and Proc. 2017/22405-3) and CNPq (Proc. 301656/2013-4 and Proc. 150699/2017-4). M.S.P. and R.L.Z are researchers from the National Research Council of Brazil-CNPq. C.L.S. is undergraduate research fellow from FAPESP at São Paulo State University - UNESP, Rio Claro, Brazil.

Authors' contributions

M.S.P. obtained the major funding, conceived and designed the experiments and supervised the research. J.R.A.S.P. conceived and designed the experiments and supervised the research. J.R.A.S.P. and C.L.S. obtained samples for protein extraction, designed and performed all the proteomic experiments, analyzed the data, prepared the figures and tables. F.G.E. analyzed the data and prepared the tables. A.P.R. and L.G.R.F. performed the immunoblotting analysis. R.L.Z. contributed analytical tools and analyzed the immunoblotting data. M.S.P., J.R.A.S.P. and C.L.S. drafted the text. All authors reviewed the manuscript.

Notes

The authors have no competing financial interests to declare.

References

[1] Palma MS. Hymenoptera insect peptides. Handbook of Biologically Active Peptides, second ed. 2013;58:1942.

[2] Perez-Riverol A, dos Santos-Pinto JRA, Lasa AM, Palma MS, & Brochetto-Braga MR. Wasp venomic: Unravelling the toxins arsenal of *Polybia paulista* venom and its potential pharmaceutical applications. Journal of Proteomics 2017;161:88-103.

[3] dos Santos-Pinto, JRA, Perez-Riverol A, Lasa AM, & Palma MS. Diversity of peptidic and proteinaceous toxins from social Hymenoptera venoms. Toxicon 2018;148:172-196.

[4] Dias NB, de Souza BM, Gomes PC, Brigatte P, Palma, MS. Peptidome profiling of venom from the social wasp *Polybia paulista*. Toxicon 2015;107:290-303.

[5] Bouzid W, Klopp C, Verdenaud M, Ducancel F, Vétillard A. Profiling the venom gland transcriptome of *Tetramorium bicarinatum* (Hymenoptera: Formicidae): The first transcriptome analysis of an ant species. Toxicon 2013;70:70-81.

[6] Perez-Riverol A, Justo-Jacomini DL, Zollner RDL, Brochetto-Braga MR. Facing Hymenoptera Venom Allergy: From Natural to Recombinant Allergens. Toxins 2015;7:2551-2570.

[7] Touchard A, et al. The biochemical toxin arsenal from ant venoms, Toxins (Basel) 2016;8:1-28.

[8] Calvete JJ, Lomonte B. A bright future for integrative venomics. Toxicon 2015;107:159-162.

[9] Touchard A, et al. Elucidation of the unexplored biodiversity of ant venom peptidomes via MALDI-TOF mass spectrometry and its application for chemotaxonomy. Journal of Proteomics 2014;105:217-231.

[10] Schwartz EF, Mourão CB, Moreira KG, Camargos TS, Mortari MR. Arthropod venoms: a vast arsenal of insecticidal neuropeptides. Biopolymers 2012;98(4):385-405.

ACCEPTED MANUSCRIPT

[11] Jiménez-Vargas JM, Possani LD, Luna-Ramírez K. Arthropod toxins acting on neuronal potassium channels. Neuropharmacology 2017;127:139-160.

[12] Santos LD, et al. Profiling the proteome of the venom from the social wasp *Polybia paulista*: A clue to understand the envenoming mechanism. Journal of Proteome Research 2010;9:3867-3877.

[13] Leite NB, et al. PE and PS Lipids Synergistically Enhance Membrane Poration by a Peptide with Anticancer Properties, Biophys. J. 2015;109:936-947.

[14] Kolarich D, Léonard R, Hermmer W, Altmann F. The N-glycans of yellow jacket venom hyaluronidases and the protein sequence of its major isoform in Vespula vulgaris. The FEBS Journal 2005;272:5182-5190.

[15] dos Santos-Pinto JRA, et al. Using proteomic strategies for sequencing and post-translational modifications assignment of antigen5, a major allergen from the venom of the social wasp *Polybia paulista*. Journal of proteome research 2014;13(2):855-865.

[16] Eržen R, Korošec P, Šilar, M, Mušic E, Košnik M. Carbohydrate epitopes as a cause of cross-reactivity in patients allergic to Hymenoptera venom. Wien. Klin. Wochenschr 2009;121:349-352.

[17] Andrade-Silva D, et al. Proteomic and Glycoproteomic Profilings Reveal That Posttranslational Modifications of Toxins Contribute to Venom Phenotype in Snakes. J. Proteome Res. 2016;15:2658-2675.

[18] Schwarz F, Aebi M. Mechanisms and principles of N-linked protein glycosylation. Current opinion in structural biology 2011;21(5):576-582.

[19] dos Santos-Pinto JR, et al. Silkomics: Insight into the Silk Spinning Process of Spiders. J. Proteome Res. 2016;15(4):1179-1193.

[20] Perez-Riverol A, et al. Molecular cloning, expression and IgE-immunoreactivity of phospholipase A1, a major allergen from Polybia paulista (Hymenoptera: Vespidae) venom. Toxicon 2016;124:44-52.

[21] Edgar RC, et al. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792-1797.

[22] Saha S, & Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins: Structure, Function, and Bioinformatics 2006;65(1):40-48.

[23] Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018; 46(W1):W296-W303.

[24] Peiren N, et al. Proteomic analysis of the honey bee worker venom gland focusing on the mechanisms of protection against tissue damage. Toxicon 2008;52:72-83.

[25] dos Santos Pinto JR, et al. Proteomic view of the venom from the fire ant *Solenopsis invicta* Buren. J. Proteome Res 2012;11(9):4643-4653.

[26] de Graaf DC, Aerts M, Brunain M, Desjardins CA, Jacobs FJ, Werren JH, Devreese B. Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies. Insect Mol. Biol. 2010;19:11-26.

[27] Heavner ME, et al. Partial venom gland transcriptome of a *Drosophila parasitoid wasp*, *Leptopilina heterotoma*, reveals novel and shared bioactive profiles with stinging Hymenoptera. Gene 2013;526(2):195-204.

ACCEPTED MANUSCRIPT

[28] Wang L, Zhu JY, Qian C, Fang Q, & Ye GY. Venom of the parasitoid wasp Pteromalus puparum contains an odorant binding protein. Archives of insect biochemistry and physiology 2015;88(2):101-110.

[29] Calvello M, et al. Expression of odorant-binding proteins and chemosensory proteins in some Hymenoptera. Insect biochemistry and molecular biology 2005;35(4):297-307.

[30] Esteves FG, Santos-Pinto JR, Saidemberg DM, Palma MS. Using a proteometabolomic approach to investigate the role of Dufour's gland in pheromone biosynthesis in the social wasp *Polybia paulista*. Journal of proteomics 2017;151:122.

[31] Weinstock GM, et al. Honeybee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee Apis mellifera. Nature 2006;443(7114):931.

[32] Werren JH, et al. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science 2010;327(5963):343-348.

[33] Nadzirin N, Firdaus-Raih M. Proteins of Unknown function in the Protein Data Bank (PDB): an inventory of true uncharacterized proteins and computational tools for their analysis. International journal of molecular sciences 2012;13(10):12761-12772.

[34] Wiesner J, Premsler T, Sickmann A. Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. Proteomics 2008;8:4466-4483.

[35] Johnson H, Eyers CE. Analysis of post-translational modifications by LC-MS/MS. Methods Mol Biol. 2010;658:93-108.

[36] Dutta D, Mandal C, Mandal C. Unusual glycosylation of proteins: Beyond the universal sequon and other amino acids. Biochim Biophys Acta Gen Subj. 2017;1861(12):3096-3108.

[37] Perez-Riverol A, et al. Phospholipase A1-based cross-reactivity among venoms of clinically relevant Hymenoptera from Neotropical and temperate regions. Molecular Immunology 2018;93:87-93.

[38] Hoffman DR. Hymenoptera venom allergens. Clinical reviews in Allergy & Immunology 2006;30(2):109-128.

[39] Santos, L.D., et al. Proteomic characterization of the multiple forms of the PLAs from the venom of the social wasp *Polybia paulista*. Proteomics 2011;11:1403-1412.

[40] Jakob T, Muller U, Helbling A, Spillner E. Component Resolved Diagnostics for Hymenoptera Venom Allergu. Curr Opin Allergy Clin Immunol 2017;17(5):363-372.

[41] Xiao H, Pan H, Liao K, Yang M, Huang C. Snake Venom PLA_2, a Promising Target for Broad-Spectrum Antivenom Drug Development. Biomed Res Int. 2017;6592820.

[42] Lomonte B, Rangel J. Snake venom Lys49 myotoxins: from phospholipases A_2 to nonenzymatic membrane disruptors. Toxicon 2012;60(4):520-53057.

[43] Borja M, et al. Biological and Proteolytic Variation in the Venom of *Crotalus scutulatus scutulatus* from Mexico. Toxins 2018;10(1):35.

[44] Bouzid W, et al. De Novo sequencing and transcriptome analysis for *Tetramorium bicarinatum*: a comprehensive venom gland transcriptome analysis from an ant species. BMC genomics 2014;15(1):987.

[45] Baptista-Saidemberg NB, et al. Protonectin (1-6): a novel chemotactic peptide from the venom of the social wasp *Agelaia pallipes pallipes*. Toxicon. 2010;56(6):880-889.

ACCEPTED MANUSCRIPT

[46] Hancock RE, Diamond G. The role of cationic antimicrobial peptides in innate host defences. Trends in microbiology 2000;8(9):402-410.

[47] Trevisan-Silva D, et al. Astacin-like metalloproteases are a gene family of toxins present in the venom of different species of the brown spider (genus *Loxosceles*). Biochimie 2010;92(1):21-32.

[48] Sanggaard KW, Bechsgaard JS, Fang X, Duan J, Dyrlund TF. Spider genomes provide insight into composition and evolution of venom and silk. Nature Communications 2014;5:3765.

[49] Fry BG, et al. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. Annu. Rev. Genom. Hum. Genet. 2009;10; 483-511.

[50] Pinto JR, Santos LD, Arcuri HA, Dias NB, Palma, MS. Proteomic characterization of the hyaluronidase (E.C. 3.2.1.35) from the venom of the social wasp *Polybia paulista*. Protein Pep. Lett. 2012;19:624-634.

[51] Müller U, Schmid-Grendelmeier P, Hausmann O, Helbling A. IgE to recombinant allergens Api m 1, Ves v 1, and Ves v 5 distinguish double sensitization from crossreaction in venom allergy. Allergy: European Journal of Allergy and Clinical Immunology 2012;67:1069-1073.

[52] Blank S, et al. Vitellogenins are new high molecular weight components and allergens (Api m 12 and Ves v 6) of *Apis mellifera* and *Vespula vulgaris* venom. PLoS One 2013;8(4).

[53] Justo-Jacomini DL, et al. Hyaluronidase from the venom of the social wasp *Polybia paulista* (Hymenoptera, Vespidae): cloning, structural modeling, purification, and immunological analysis. Toxicon 2013;64:70-80.

[54] Hossen MS, Shapla UM, Gan SH, Khalil MI. Impact of Bee Venom Enzymes on Diseases and Immune Responses. Molecules 2016;22(1).

[55] Altmann F.The role of protein glycosylation in allergy. International archives of allergy and immunology 2007;142(2):99-115.

[56] Fotisch, K., Vieths S. N- and O-linked oligosaccharides of allergenic glycoproteins. Glycoconjugate J. 2001;18(5):373-90.

[57] Ollert M, Blank S. Anaphylaxis to insect venom allergens: role of molecular diagnostics. Current allergy and asthma reports 2015;15(5):1-11.

[58] dos Santos-Pinto JRA, et al. B-cell linear epitopes mapping of antigen5 allergen from *Polybia paulista* wasp venom. Journal of Allergy and Clinical Immunology 2015;135(1):264.

[59] Blank S, et al. Identification, recombinant expression, and characterization of the 100 kDa high molecular weight Hymenoptera venom allergens Api m 5 and Ves v 3. J Immunol. 2010;184(9):5403-13.

[60] Tomsitz D, Brockow K. Component resolved diagnosis in Hymenoptera anaphylaxis. Current allergy and asthma reports 2017;17(6):38.

[61] Seismann H, et al. Recombinant phospholipase A1 (Ves v 1) from yellow jacket venom for improved diagnosis of hymenoptera venom hypersensitivity. Clinical and Molecular Allergy 2010;8:7.

[62] Monsalve RI, et al. Component-resolved diagnosis of vespid venom-allergic individuals: Phospholipases and antigen 5s are necessary to identify Vespula or Polistes sensitization. Allergy: European Journal of Allergy and Clinical Immunology 2012;67:528-536.

[63] Perez-Riverol A, et al. Venoms of Neotropical wasps lack cross-reactive carbohydrate determinants enabling reliable protein-based specific IgE determination. J Allergy Clin Immunol. 2018b;141(5):1917-1919.

ACCEPTED MANUSCRIPT

[64] Yu CJ, Lin, YF, Chiang BL, Chow LP. Proteomics and immunological analysis of a novel shrimp allergen, Pen m 2. The Journal of Immunology 2003;170(1):445-453.

[65] Farioli L, et al. Mite-Induced Asthma and IgE Levels to Shrimp, Mite, Tropomyosin, Arginine Kinase, and Der p 10 Are the Most Relevant Risk Factors for Challenge-Proven Shrimp Allergy. International archives of allergy and immunology 2007;174(3-4):133-143.

[66] Kamath DS, et al. IgE reactivity to shrimp allergens in infants and their cross-reactivity to house dust mite. Pediatric, Allergy and Immunologi 2017;174(3-4):133-143

[67] Palma MS, In Handbook of Biologically Active Peptides; Kastin, A., Ed.; Academic Press: San Diego, 2006;409-417.

[68] Pringos E, Vignes M, Martinez J, Rolland V. Peptide neurotoxins that affect voltage-gated calcium channels: a close-up on ω-agatoxins. Toxins 2011;3(1):17-42.

[69] Garb JE, Hayashi CY. Molecular evolution of α-latrotoxin, the exceptionally potent vertebrate neurotoxin in black widow spider venom. Molecular biology and evolution 2013;30(5):999-1014.

[70] Rodríguez de La Vega RC, Possani LD. Current views on scorpion toxins specific for K^+channels. Toxicon 2004;43:865-875.

[71] Song W, et al. Substitutions in the domain III voltage-sensing module enhance the sensitivity of an insect sodium channel to a scorpion beta-toxin. J Biol Chem. 2011;286(18):15781-8.

[72] Zhao F, et al. Proteotranscriptomic Analysis and Discovery of the Profile and Diversity of Toxin-like Proteins in Centipede. Molecular & Cellular Proteomics 2018;mcp-RA117.

[73] Olivera BM. Peptide neurotoxins from fish-hunting cone snails. Science 1985;230:1338-1343.

[74] Bae N, Li L, Lödl M, Lubec' G. Peptide toxin glacontryphan-M is present in the wings of the butterfly *Hebomoia glaucippe* (Linnaeus, 1758) (Lepidoptera: Pieridae). Proc Natl Acad Sci USA 2012;109:17920-17924.

[75] Torres AF, et al. Transcriptome analysis in venom gland of the predatory giant ant *Dinoponera quadriceps*: Insights into the polypeptide toxin arsenal of hymenopterans. PLoS One 2014;9(1):e87556.

[76] Lee SH, Baek JH, Yoon KA. Differential properties of venom peptides and proteins in solitary vs. social hunting wasps. Toxins 2016;8(2):32.

Figure captions

Fig. 1. Gene Ontology (GO) enrichment analysis. The most abundant Gene ontology (GO) categories for **A -** Biological process, **B -** Cellular component, and **C -** Molecular function, in *P. paulista wasp venom*. **D -** Total protein distribution according to their general function.

Fig. 2. Representative mass spectra of phospholipase A1 molecular forms identified in *P. paulista* venom. **A -** CID spectrum of tryptic peptide PFN*EYDILFFVYTR, which was acquired by selecting the m/z 871.0637 $[M + 3H]^{3+}$ as a precursor ion and showing the N-linked glycosylation Hex2HexNac2 (Δ = +730.26) site detected on phospholipase A1, UniProtKB accession number A2VBC4. **B -** CID spectrum of tryptic peptide FN*EYDILFFVYTR, which was acquired by selecting the m/z 871.0599 $[M + 3H]^{3+}$ as a precursor ion and showing the N-linked glycosylation Hex2HexNac2Pent1 (Δ = +862.31) site detected on phospholipase A1, UniProtKB accession number E7D4M5. The glutamic acid (E) residue is presenting an increase of mass (Δ = +21.98), characterizing the presence of a sodium adduct.

ACCEPTED MANUSCRIPT

Fig. 3. Representative mass spectra that present the assignment of N-glycosylation sites for one uncharacterized protein, and for one peptide sequence (without identification and characterization) identified in *P. paulista* venom. **A -** CID spectrum of tryptic peptide GN*LLGTRLDN*PEK, which was acquired by selecting the 875.8884 $[M + 2H]^{2+}$ as a precursor ion and showing the two N-linked glycosylation hexose (Δ = +162.05) sites detected on uncharacterized protein, NCBInr accession number XP_011309886.1. **B -** CID spectrum of tryptic peptide EDSLGVCPN*PR, which was acquired by selecting the m/z 702.8536 $[M + 2H]^{2+}$ as a precursor ion and showing the N-linked glycosylation hexosamine (Δ = +161.07) site detected on peptide sequence (without identification and characterization) by *de novo* sequencing approach. The cysteine (C) residue is presenting an increase of mass (Δ = +57.02), characterizing the carbamidomethyl modification.

Fig. 4. Proteomic analysis of N-linked glycosylated protein fraction from *P. paulista* venom. **A -** Representative 1-DE profile, and **B -** Immunoblotting with the sera of patients sensitive to *P. paulista* venom, revealing the protein bands reactive to human specific-IgE: (1) crude venom fraction, (2) N-linked glycosylated protein fraction and (3) non-N-linked glycosylated protein fraction. The black arrows are indicating the presence of N-linked glycosylated phospholipase A1 molecular forms; and the white arrow is indicating the presence of non-N-linked glycosylated phospholipase A1 molecular form. **C -** Multiple sequence alignment between the phospholipases A1 molecular forms (UniProtKB accession numbers E7D4M5 and A2VBC4), highlighting the position of the different types of N-linked glycosylation (indicated by colored arrows) detected in both molecular forms. * (%) Sequence coverage obtained by trypsin enzyme digestion for each phospholipase A1 is indicated by red letters; conserved regions are colored in green. B-cell linear epitopes prediction along the phospholipase A1 (E7D4M5) primary sequence is indicated by black bars. **D -** Representative three-dimensional molecular model of the phospholipase A1 (E7D4M5). The image was generated using SWISS-MODEL. B-cell linear epitopes prediction is shown in black and the N-linked glycosylations are shown in colored arrows - Hex2HexNac2Pent1 in blue arrow; Hexosamine in red arrow; and Hexose in pink arrow.

Fig. 5. Schematized representation of the general envenoming mechanism proposed for the *P. paulista* venom, involving the proteins identified in the present study. **A -** *P. paulista* wasp, **B -** Stinging: venom injection, and **C -** Venom action in the victim after the stinging.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1. Protein identification by gel-free shotgun proteomic approach - (i) Typical venom proteins described in social Hymenoptera venoms.

Accession number	Protein name	Taxonomy	MW (kDa)
E2AUG7	ADAMTS-like protein 3	*Camponotus floridanus*	172
gi|328780480	97 kDa heat shock protein isoformX1	*Apis mellifera*	98
Q06601	Apidaecin-2	*Apis mellifera*	19
Q6QAN5	70 kDa heat shock protein	*Megachile rotundata*	54
gi|345486559	A disintegrin and metalloproteinase	*Nasonia vitripennis*	101
gi|156554461	A disintegrin and metalloproteinase	*Nasonia vitripennis*	194
gi|156543048	A disintegrin and metalloproteinase	*Nasonia vitripennis*	67
gi|345493675	A disintegrin and metalloproteinase	*Nasonia vitripennis*	90
Q25BT8	Alpha glucosidase 2	*Apis mellifera*	67
K7IXA0	Angiotensin-converting enzyme	*Nasonia vitripennis*	90
O61367	Arginine kinase	*Apis mellifera*	40
R9RSL8	Arginine kinase	*Bombus perezi*	17
K7J676	Carboxypeptidase	*Nasonia vitripennis*	49
gi|48095159	Chymotrypsin-1	*Apis mellifera*	28
P00769	Chymotrypsin-2	*Vespa crabro*	24
gi|345480378	Chymotrypsin-2	*Nasonia vitripennis*	27
gi|238859591	Cysteine-rich/KU venom protein precursor	*Nasonia vitripennis*	87
Q8MMH0	Cysteine-rich venom protein 7	*Pimpla hypocondriaca*	12
I1ZEL1	Defensin 2-3	*Nasonia vitripennis*	11
E2AWB2	Dipeptidase	*Camponotus floridanus*	21
gi|239050560	Endonuclease-like venom protein precursor	*Nasonia vitripennis*	57
gi|156547889	Heat shock 70 kDa protein cognate 4-like	*Nasonia vitripennis*	71
gi|345479692	Heat shock 70 kDa protein-like	*Nasonia vitripennis*	9
gi|345480729	Heat shock protein 83	*Nasonia vitripennis*	83
gi|110757651	Heat shock protein beta-1-like isoformX1	*Apis mellifera*	26
gi|229892210	Heat shock protein cognate 4	*Apis mellifera*	71
gi|345485308	Heat shock protein DDB_G0288861-like isoform X5	*Nasonia vitripennis*	101
gi|229892265	Heat shock protein Hsp70Ab-like	*Apis mellifera*	70
P86687	Hyaluronidase	*Polybia paulista*	34
Q08169	Hyaluronidase - Allergen Api m II	*Apis mellifera*	44
P49371	Hyaluronidase - allergen Dol m 2	*Dolichovespula maculata*	39
Q9U6V9	Hyaluronidase - allergen Pol a 2	*Polistes annularis*	43
R9RVN4	Hymenoptaecin	*Bombus campestris*	8
gi|238859585	Kazal type serine protease inhibitor-like venom protein 1	*Nasonia vitripennis*	7
E2C982	Major antigen	*Harpegnathos saltator*	225
Q17060	Major royal jelly protein 3	*Apis mellifera*	61
O97432	Major royal jelly protein 5	*Apis mellifera*	70
Q4ZJX1	Major royal jelly protein 9	*Apis mellifera*	48
A2VBC4	Phospholipase A1	*Polybia paulista*	36
E7D4M5	Phospholipase A1	*Polybia paulista*	33
P49369	Phospholipase A1	*Vespula vulgaris*	38
P0CH86	Phospholipase A1	*Vespula squamosa*	32
Q06478	Phospholipase A1	*Dolichovespula maculata*	36
Q6Q249	Phospholipase A1	*Polistes dominula*	35
gi|328791555	Phospholipase A2-like isoform X2	*Apis mellifera*	47
gi|156553185	Protein lethal (2) essential for life	*Nasonia vitripennis*	19
P0C1R1	Protonectin	*Protonectarina sylverae*	1
gi|156538621	Similar to A disintegrin and metalloproteinase	*Nasonia vitripennis*	84
gi|156538921	Similar to heat shock protein	*Nasonia vitripennis*	58
gi|156545866	Similar to metalloproteinase-putative	*Nasonia vitripennis*	35
gi|156550073	Similar to venom acid phosphatase	*Nasonia vitripennis*	45
gi|156554475	Similar to venom acid phosphatase	*Nasonia vitripennis*	42
A0A087ZRC7	Venom acid phosphatase Acph-1	*Apis mellifera*	46
Q5BLY5	Venom acid phosphatase Acph-1 - allergen Api m 3	*Apis mellifera*	45
P81656	Venom allergen 5	*Polistes dominula*	25
P86686	Venom allergen 5	*Polybia paulista*	23
Q7Z156	Venom allergen 5	*Polybia scutellaris rioplatensis*	23

ACCEPTED MANUSCRIPT

D4P2Y4	Venom allergen 5 2	*Polybia paulista*	15
gi|156545770	Venom allergen 5-like	*Nasonia vitripennis*	34
K7J718	Venom carboxylesterase-6 isoform X1	*Nasoni vitripennis*	62
B2D0J5	Venom carboxylesterase-6 precursor Allergen Api m 8	*Apis mellifera*	64
B1A4F7	Venom dipeptidyl peptidase 4 - allergen Ves v 3	*Vespula vulgaris*	88
B2D0J4	Venom dipeptidyl peptidase 4 (allergen Api m 5)	*Apis mellifera*	87
gi|239787114	Venom protein F precursor	*Nasonia vitripennis*	28
gi|239787134	Venom protein U precursor	*Nasonia vitripennis*	24
C9WMM5	Venom serine carboxypeptidase (allergen Api m 9)	*Apis mellifera*	53
gi|345483324	Venom serine carboxypeptidase isoform X1	*Nasonia vitripennis*	54
Q8MQS8	Venom serine protease 34 precursor Allergen Api m 7	*Apis mellifera*	46
Q868N5	Vitellogenin	*Apis mellifera*	201
gi|345495674	Zinc metalloproteinase nas-6	*Nasonia vitripennis*	44
gi|345484862	Zinc metalloproteinase nas-7-like	*Nasonia vitripennis*	34

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2. Protein identification by gel-free shotgun proteomic approach - (i) Typical venom proteins described in other animal groups (Arthropoda and NCBInr databases).

Accession number	Protein name	Taxonomy	MW (kDa)
Q25641	Allergen Cr-PI-like	*Periplaneta americana*	81
A9UJY0	Allergen Pen m 2	*Fenneropenaeus chinensis*	6
P17728	Alpha-insect toxin LqhaIT-like	*Leiurus q. hebraeus*	9
P0DJE4	Alpha-latrotoxin-Lm1a-like	*Latrodectus mactans*	22
P0C910	Alpha-toxin Amm3-like	*Androctonus m. mauritanicus*	7
C9D7R3	Astacin-like metalloprotease toxin 3-like	*Loxosceles intermedia*	28
P0C5I3	Beta-insect depressant toxin Lqh-dprIT3a-like	*Leiurus q. hebraeus*	9
P15147	Beta-insect excitatory toxin 2 (AaH IT2)-like	*Androctonus australis*	10
A1KXI9	Blo t Gal d 1 allergen	*Blomia tropicalis*	31
Q3YEF1	Conotoxin MaIr34-like	*Conus marmoreus*	10
Q9U659	Conotoxin PnMKLT1-0122-like	*Conus pennaceus*	9
D2DRC0	Cytotoxin A isoform 1-like	*Malo kingi*	51
O76199	Delta-ctenitoxin-Pn2c-Neurotoxin Pn2-5ª-like	*Phoneutria nigriventer*	9
Q8MVU3	Der f 16 allergen	*Dermatophagoides farinae*	55
A7XZI6	Group 14 allergen Blo t 14	*Blomia tropicalis*	87
Q9U785	High molecular weight allergen M-177	*Euroglyphus maynei*	191
Q9BJC9	Major allergen	*Brugia malayi*	29
B2KSI1	Major allergen protein	*Pieris rapae*	74
G5EF32	Nematode Polyprotein Allergen related	*Caenorhabditis elegans*	165
D9U298	Neurotoxin LmNaTx3-like	*Lychas mucronatus*	9
D2Y2L8	Omega-theraphotoxin-Hhn1e	*Haplopelma hainanum*	9
P59936	Potassium channel toxin alpha-KTx 12.1	*Tityus serrulatus*	5
C5J896	Potassium channel toxin alpha-KTx 6.16	*Opisthacanthus cayaporum*	8
Q0GY41	Potassium channel toxin Hge-beta-KTx	*Hadrurus gertschi*	9
B8XGY7	Putative alpha toxin Tx405-like	*Buthus o. israelis*	9
P0C8D4	Scolopendra 20528.11 Da toxin-like	*Scolopendra angulata*	3
B2KSG8	Single domain major allergen protein	*Eucheira socialis*	24
C6JUN2	Snake venom metalloprotease	*Philodryas olfersii*	69
Q68PG3	Toxin Cex12-like	*Centruroides exilicauda*	9
P56609	Toxin Tb2 (P-beta* NaTx5.3) (Toxin III-8)-like	*Tityus bahiensis*	7
O61379	Tropomyosin - Allergen Pan s I	*Panulirus stimpsoni*	32
Q5Y4Y0	U2-agatoxin-Ao1f	*Agelena orientalis*	8
B1P1D7	U13-theraphotoxin-Cg1a-like	*Chilobrachys guangxiensis*	9
Q23737	Venom chitinase	*Chelonus sp.*	52
gi|193634351	Venom protease-like	*Acyrthosiphon pisum*	45
gi|328700427	Venom protease-like	*Acyrthosiphon pisum*	40
J9KAT6	Venom serine protease-like	*Acyrthosiphon pisum*	50
O44931	Vespid allergen antigen homolog	*Onchocerca volvulus*	25

ACCEPTED MANUSCRIPT

Table 3. N-linked glycosylated protein identification by gel-free shotgun proteomic approach - (i) Typical venom proteins described in Hymenoptera venoms. Asparagine (N) residues marked in red represent the N-linked glycosylated sites. [#] n.d. - not detected; Δ - mass values (Da) addition detected in the sequence fragments. PTMs mass values can be found at: http://www.unimod.org.

Accession number	Protein name	MW (kDa)	Taxonomy	Peptide sequences (MASCOT ions-score)	[#] Modification type of N-linked glycosylation detected in the sequence fragments
UPI00071224FE	60 kDa heat shock protein	60	*Dinoponera quadriceps*	EGFEKISK (22); GIELKDK (35)	n.d.
E2C982	Major antigen	225	*Harpegnathos saltator*	SRINTLAQENDALR (23); QALQSFGVK (32)	n.d.
UPI0005D41243	Maltase 1-like	66	*Pogonomyrmex barbatus*	NSLIYQIYPR (50); HPKTADTK (32)	(N) Hexose (Δ = +162.05)
UPI000718CCC5	Maltase 1-like	65	*Polistes canadensis*	NSLIYQIYPR (50); NVYSFWMNR (25)	(N) Hex1HexNAc1dHex1 (Δ = +511.19)
A0A026W1C9	Multidrug resistance protein-like protein	145	*Cerapachys biroi*	IAAKTMTTAKSPK (31); YFGGGK (21)	n.d.
UPI00046C9D13	Peroxiredoxin-5	26	*Nasonia vitripennis*	VNLAQLSQGK (26); STLFVVTHIR (22)	n.d.
A2VBC4	Phospholipase A1	36	*Polybia paulista*	PFNEYDILFFVYTR (55); DNFLVISVDWK (43)	(N) Hex2HexNac2 (Δ = +730.26)
E7D4M5	Phospholipase A1	33	*Polybia paulista*	FNEYDILFFVYTR (51); STNCPDRLCETDAQYVQVLHTSSK (40); DNFLVMSVDWR (56); YVADFTK (40); AVLYMTECIK (43); DICVCVGLNAK (50)	(N) Hex2HexNac2Pent1 (Δ = +862.31) (N) Hexose (Δ = +162.05) (N) Hexosamine (Δ = +161.07)
Q3ZU95	Phospholipase A1-like	33	*Vespula germanica*	YSEIIGLDPAGPSFSSNKCSDRL (53); LVGQYIATITQK (48)	(N) Hex2HexNac2 (Δ = +730.26)
UPI00073D6DA6	Phospholipase A1 2-like	39	*Polistes dominula*	NTRAVGQYVANFTK (44); IPLSKCTK (38)	n.d.
UPI00063F0DCF	Phospholipase A1-like	38	*Monomorium pharaonis*	VNESCDLFDLSK (39); SITDMCGVPLR (32)	n.d.
UPI000718AE4A	Phospholipase A1-like	37	*Polistes canadensis*	DGIILTER (38); NMTTFDVFTK (41)	n.d.
UPI0005D363BD	Probable chitinase 3	26	*Pogonomyrmex barbatus*	GLLPLVILATLSKAQFR (24); TLLQEPQPSK (30)	n.d.
UPI000258F4B0	Putative serine protease	54	*Megachile rotundata*	MTATSLGSPLQR (26); LAHVITDGNKK (30)	n.d.
UPI000718CB44	Serine-protein kinase	336	*Polistes canadensis*	DMEAAMGVSGVEGIMR (22); QMEESAYK (32)	n.d.
A0A151JR25	Tropomyosin	32	*Trachymyrmex cornetzi*	LEKDNAMDR (20); ALLCEQQARDANAR (23)	(N) Hexose (Δ = +162.05)
UPI000626A6B4	Tropomyosin isoform X9	32	*Athalia rosae*	LATATAK (26); LAMVEADLER (23)	n.d.
UPI000580D67D	Tropomyosin-1	32	*Acromyrmex echinatior*	VLEEVADLTK (26); SLTNAESEVAALNR (57); VQLIEEDLER (18); LLAEDADGKSDEVSR (56); LAFVEDELEVAEDR (48)	(N) Hexose (Δ = +162.05)
UPI000718CA7A	Tropomyosin-1	32	*Polistes canadensis*	SLTNAESEVAALNR (57); SEERLNTATAKLAEASQAADESSR (23) LLAEDADGKSDEVSR (56);	(N) Hex1HexNAc1dHex1 (Δ = +511.19)

ACCEPTED MANUSCRIPT

				LAFVEDELEVAEDR (48); IMELEEELKVVGNSLK (39)	
UPI000771B6A9	Tropomyosin-2 isoform X13	32	*Cephus cinctus*	IVELEEELR (56); DNAMDR (28)	(N) Hexosamine (Δ = +161.07)
UPI0006261515	Tropomyosin-2 isoform X8	32	*Athalia rosae*	AEEEARSLQK (36); IQTIENELDQTQEALMQVNAK (32)	(N) Hex1HexNac2 (Δ = +568.21)
UPI0006254231	Tropomyosin-2 isoform X9	32	*Orussus abietinus*	ALQNAESEVAALNR (40); IQLLEEDLER (56); IVELEEELRVVGNNLK (40); VVGNNLKSLEVSEEK (35); AIEREETFK (22)	(N) Hexose (Δ = +162.05) (N) Hexosamine (Δ = +161.07)
UPI000718D8DF	Venom dipeptidyl peptidase 4	88	*Polistes canadensis*	NLGTVEVEDQITVTR (37); FLLIGYDVQK (22)	(N) Hexose (Δ = +162.05)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Biological significance

The present study significantly expanded the number of proteins identified in *P. paulista* venom, contributing to improvements in our understanding of the envenoming mechanism produced by sting accidents caused by this wasp. For example, novel wasp venom neurotoxins have been identified, but no studies have assessed the presence of this type of toxin in social wasp venoms. In addition, 23 N-linked glycosylated venom proteins were identified in the *P. paulista* venom proteome, and some of these proteins might be relevant allergens that are immunoreactive to human IgE.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Highlights

- Shotgun proteomic approach was used to explore the proteome of *P. paulista* wasp venom.
- More than 100 proteinaceous toxins were identified.
- Twenty-three N-linked glycosylated venom proteins were identified.
- *P. paulista* venom contains neurotoxin-like proteins.
- The described toxin arsenal improves our understanding of wasp-induced envenoming.

ACCEPTED MANUSCRIPT

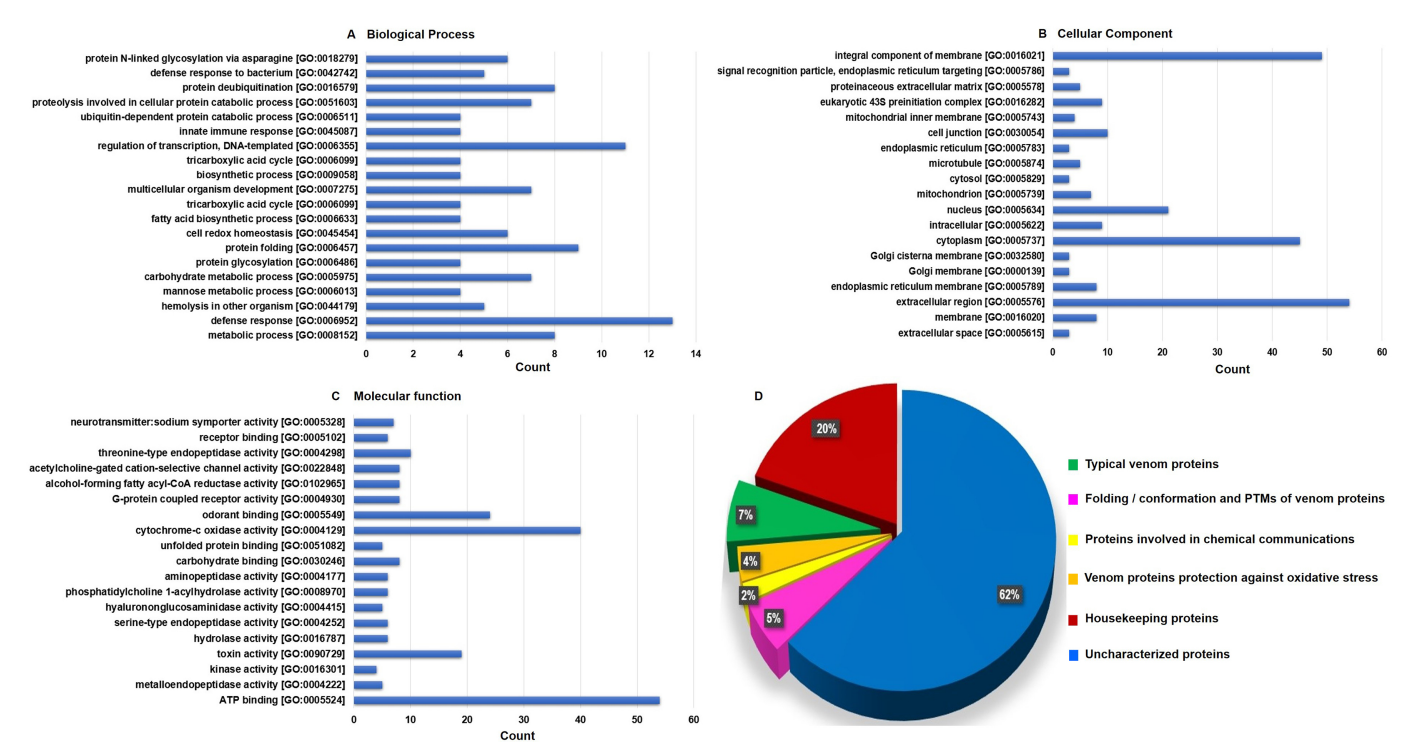

A Biological Process
protein N-linked glycosylation via asparagine [GO:0018279]
defense response to bacterium [GO:0042742]
protein deubiquitination [GO:0016579]
proteolysis involved in cellular protein catabolic process [GO:0051603]
ubiquitin-dependent protein catabolic process [GO:0006511]
innate immune response [GO:0045087]
regulation of transcription, DNA-templated [GO:0006355]
tricarboxylic acid cycle [GO:0006099]
biosynthetic process [GO:0009058]
multicellular organism development [GO:0007275]
tricarboxylic acid cycle [GO:0006099]
fatty acid biosynthetic process [GO:0006633]
cell redox homeostasis [GO:0045454]
protein folding [GO:0006457]
protein glycosylation [GO:0006486]
carbohydrate metabolic process [GO:0005975]
mannose metabolic process [GO:0006013]
hemolysis in other organism [GO:0044179]
defense response [GO:0006952]
metabolic process [GO:0008152]
0 2 4 6 8 10 12 14
Count
B Cellular Component
integral component of membrane [GO:0016021]
signal recognition particle, endoplasmic reticulum targeting [GO:0005786]
proteinaceous extracellular matrix [GO:0005578]
eukaryotic 43S preinitiation complex [GO:0016282]
mitochondrial inner membrane [GO:0005743]
cell junction [GO:0030054]
endoplasmic reticulum [GO:0005783]
microtubule [GO:0005874]
cytosol [GO:0005829]
mitochondrion [GO:0005739]
nucleus [GO:0005634]
intracellular [GO:0005622]
cytoplasm [GO:0005737]
Golgi cisterna membrane [GO:0032580]
Golgi membrane [GO:0000139]
endoplasmic reticulum membrane [GO:0005789]
extracellular region [GO:0005576]
membrane [GO:0016020]
extracellular space [GO:0005615]
0 10 20 30 40 50 60
Count
C Molecular function
neurotransmitter:sodium symporter activity [GO:0005328]
receptor binding [GO:0005102]
threonine-type endopeptidase activity [GO:0004298]
acetylcholine-gated cation-selective channel activity [GO:0022848]
alcohol-forming fatty acyl-CoA reductase activity [GO:0102965]
G-protein coupled receptor activity [GO:0004930]
odorant binding [GO:0005549]
cytochrome-c oxidase activity [GO:0004129]
unfolded protein binding [GO:0051082]
carbohydrate binding [GO:0030246]
aminopeptidase activity [GO:0004177]
phosphatidylcholine 1-acylhydrolase activity [GO:0008970]
hyaluronoglucosaminidase activity [GO:0004415]
serine-type endopeptidase activity [GO:0004252]
hydrolase activity [GO:0016787]
toxin activity [GO:0090729]
kinase activity [GO:0016301]
metalloendopeptidase activity [GO:0004222]
ATP binding [GO:0005524]
0 10 20 30 40 50 60
Count
D
20%
7%
4%
2%
5%
62%
Typical venom proteins
Folding / conformation and PTMs of venom proteins
Proteins involved in chemical communications
Venom proteins protection against oxidative stress
Housekeeping proteins
Uncharacterized proteins

Figure 1

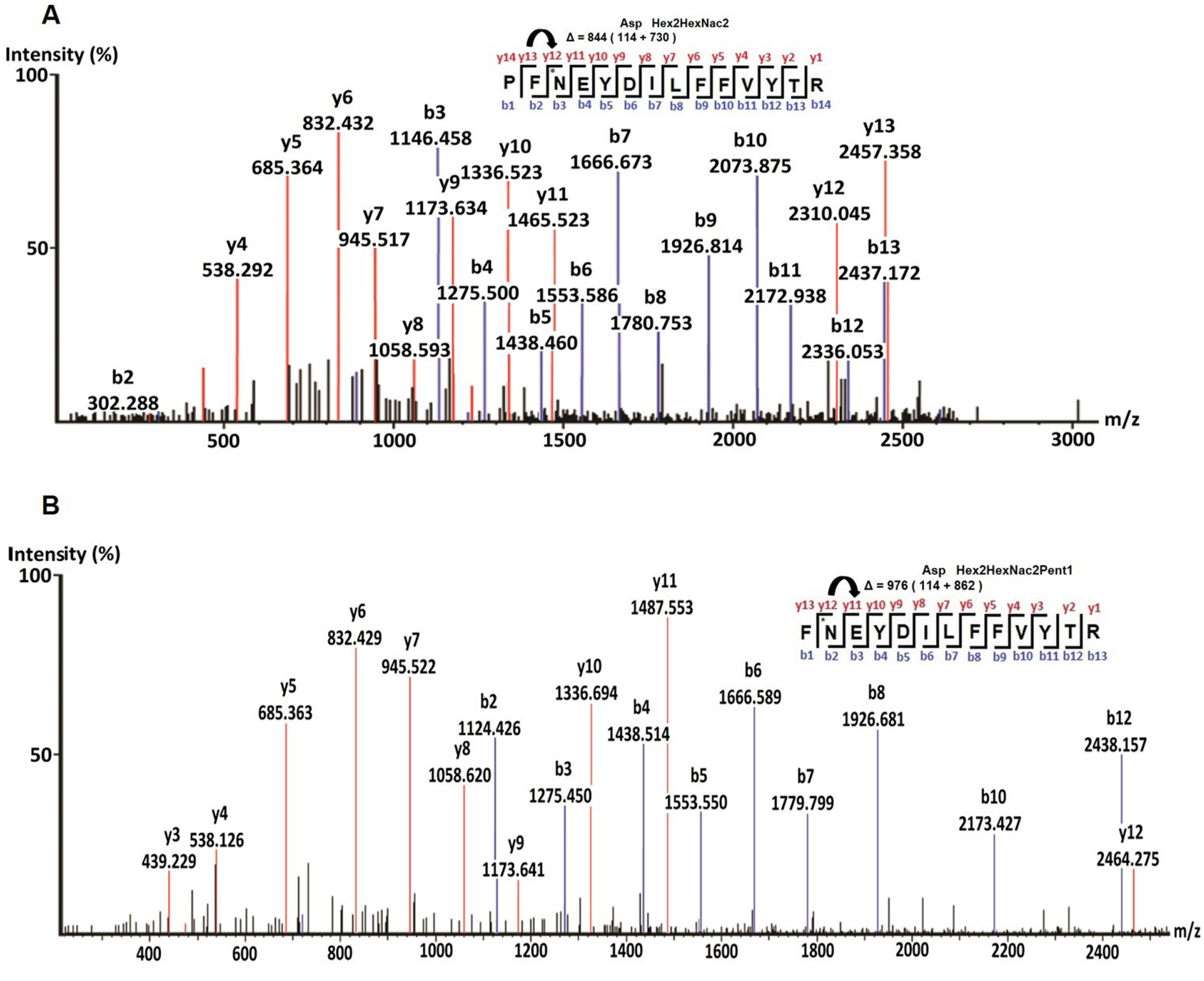

A
Intensity (%)
100
50
Asp Hex2HexNac2
Δ = 844 (114 + 730)
y14 y13 y12 y11 y10 y9 y8 y7 y6 y5 y4 y3 y2 y1
P F N E Y D I L F F V Y T R
b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13 b14
y6 832.432
b3 1146.458
y5 685.364
y10 1336.523
b7 1666.673
b10 2073.875
y13 2457.358
y9 1173.634
y11 1465.523
y12 2310.045
y7 945.517
b9 1926.814
y4 538.292
b4 1275.500
b6 1553.586
b11 2172.938
b13 2437.172
b5 1438.460
b8 1780.753
y8 1058.593
b12 2336.053
b2 302.288
500
1000
1500
2000
2500
3000
m/z
B
Intensity (%)
100
50
Asp Hex2HexNac2Pent1
Δ = 976 (114 + 862)
y13 y12 y11 y10 y9 y8 y7 y6 y5 y4 y3 y2 y1
F N E Y D I L F F V Y T R
b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13
y11 1487.553
y6 832.429
y7 945.522
y10 1336.694
b6 1666.589
b8 1926.681
y5 685.363
b2 1124.426
b4 1438.514
b12 2438.157
y8 1058.620
b3 1275.450
b5 1553.550
b7 1779.799
b10 2173.427
y4 538.126
y3 439.229
y9 1173.641
y12 2464.275
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z

Figure 2

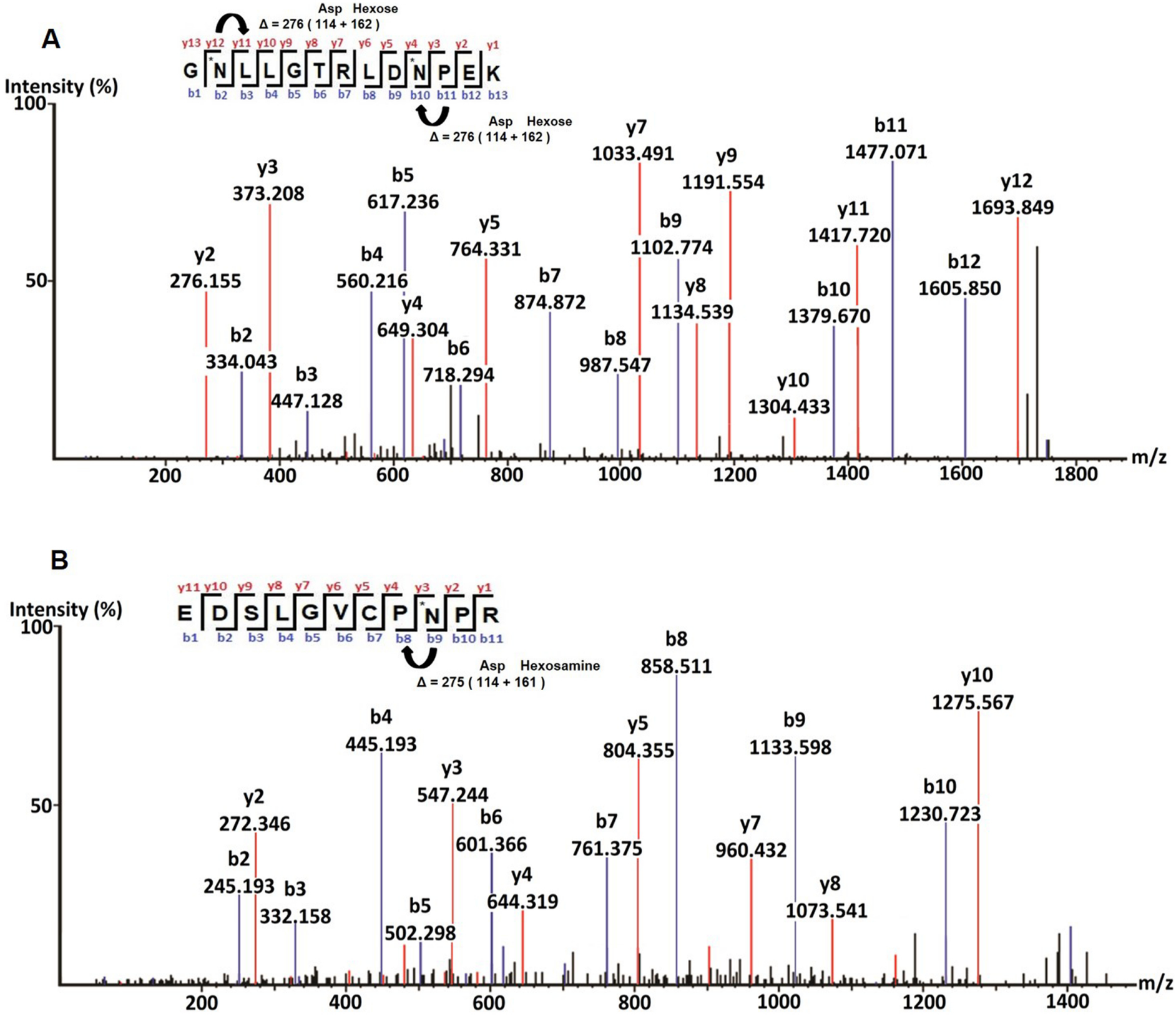
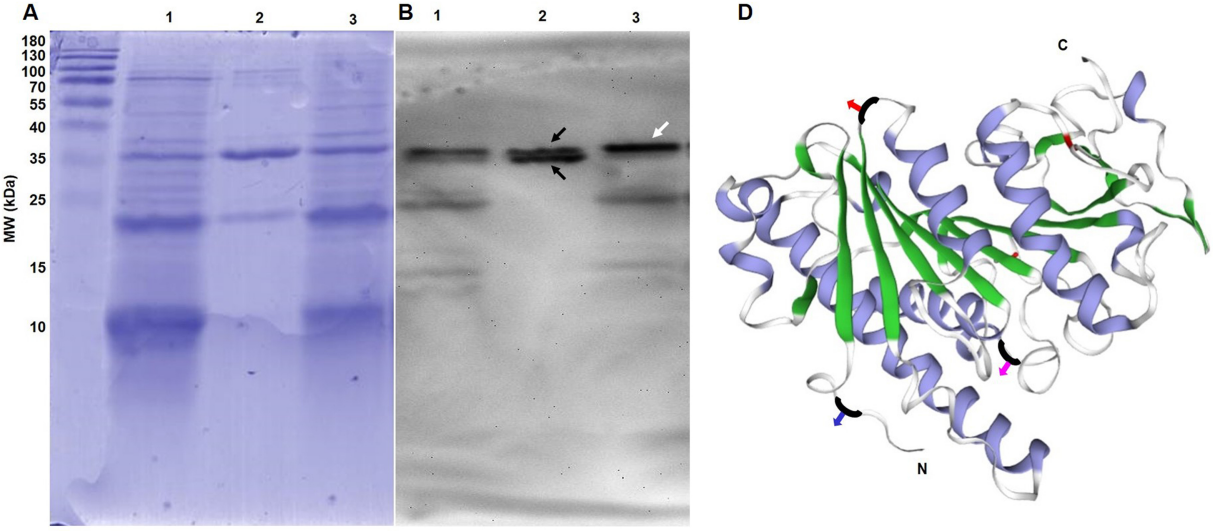

A
Asp Hexose
Δ = 276 (114 + 162)
y13 y12 y11 y10 y9 y8 y7 y6 y5 y4 y3 y2 y1
G N L L G T R L D N P E K
b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13
Asp Hexose
Δ = 276 (114 + 162)
Intensity (%)
100
50
y7 1033.491
b11 1477.071
y9 1191.554
y12 1693.849
y3 373.208
b5 617.236
y11 1417.720
b9 1102.774
y5 764.331
y2 276.155
b4 560.216
b12 1605.850
b7 874.872
y8 1134.539
b10 1379.670
y4 649.304
b2 334.043
b8 987.547
b6 718.294
y10 1304.433
b3 447.128
200 400 600 800 1000 1200 1400 1600 1800 m/z
B
y11 y10 y9 y8 y7 y6 y5 y4 y3 y2 y1
E D S L G V C P N P R
b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11
Asp Hexosamine
Δ = 275 (114 + 161)
Intensity (%)
100
50
b8 858.511
y10 1275.567
b4 445.193
y5 804.355
b9 1133.598
y3 547.244
b10 1230.723
y2 272.346
b6 601.366
b7 761.375
y7 960.432
b2 245.193
y4 644.319
b3 332.158
y8 1073.541
b5 502.298
200 400 600 800 1000 1200 1400 m/z

Figure 3



C

```
          1                                                                                                                  60
                                                  Hex2HexNac2Pent1
E7D4M5                                  GLI PECPFNEYDI LFFVYTRDKQ DGIILKKNNL TSYNLFQQPQ
A2VBC4    MNFKYSILFI CFGTLDRGLI PECPFNEYDI LFFVYTRQQR DGIVLTEETL QNYDLFKKST
                                                  Hex2HexNac2
          61                                                 Hexosamine                                                   120
E7D4M5    ISHRVVFLIH GFISTGNNGN FDAMAKALIA KDNFLVMSVD WRRAACDA-- SVFNVAGYSK
A2VBC4    ISRQVVFIDH GFLSNGNNEN FIAMAKALIE KDNFLVISVD WKKGACNAFA STLDYLGYST

          120                                                                                                            180
E7D4M5    AVANTRRVGK YVADFTKILV DKYKVPMSNI RLIGHSLGAH TSGFAGKQVQ KLKLGKYSEI
A2VBC4    AVGNTRHVGK YVADFTKLLV EQYKVSMSNI RLIGHSLGAH TSGFAGKEVQ ELKLNKYSNI
                                  Hexose
          181                                                                                                            240
E7D4M5    IGLDPAGPSF RSTNCPDRLC ETDAQYVQVL HTSSKLGTYD QIGSVDFYVN YGGSQPGC--
A2VBC4    DGLDPAGPSF DSNDCPERLC ETDAEYVQII HTSNILGVYS KIGTVDFYMN YGSHQPGCGR

          241                                                                                                            300
E7D4M5    -LVRTCSHTK AVLYMTECIK RECCLIGTPW SSTGSSPKPK PISACKRDIC VCVGLNAKSY
A2VBC4    FFSPSCSHTK AVKYLTECIK HECCLIGTPW KKYFSTPKP- -ISQCTKDTC VCVGLNAKSY

          301                                  324
E7D4M5    PAKGSFYVPV EKNSPYCHNE GIKL
A2VBC4    PARGSFYVPV EATAPYCHNE GIKL
```

*Sequence coverage:
E7D4M5 - 40.06%
A2VBC4 - 18.32%

Figure 4

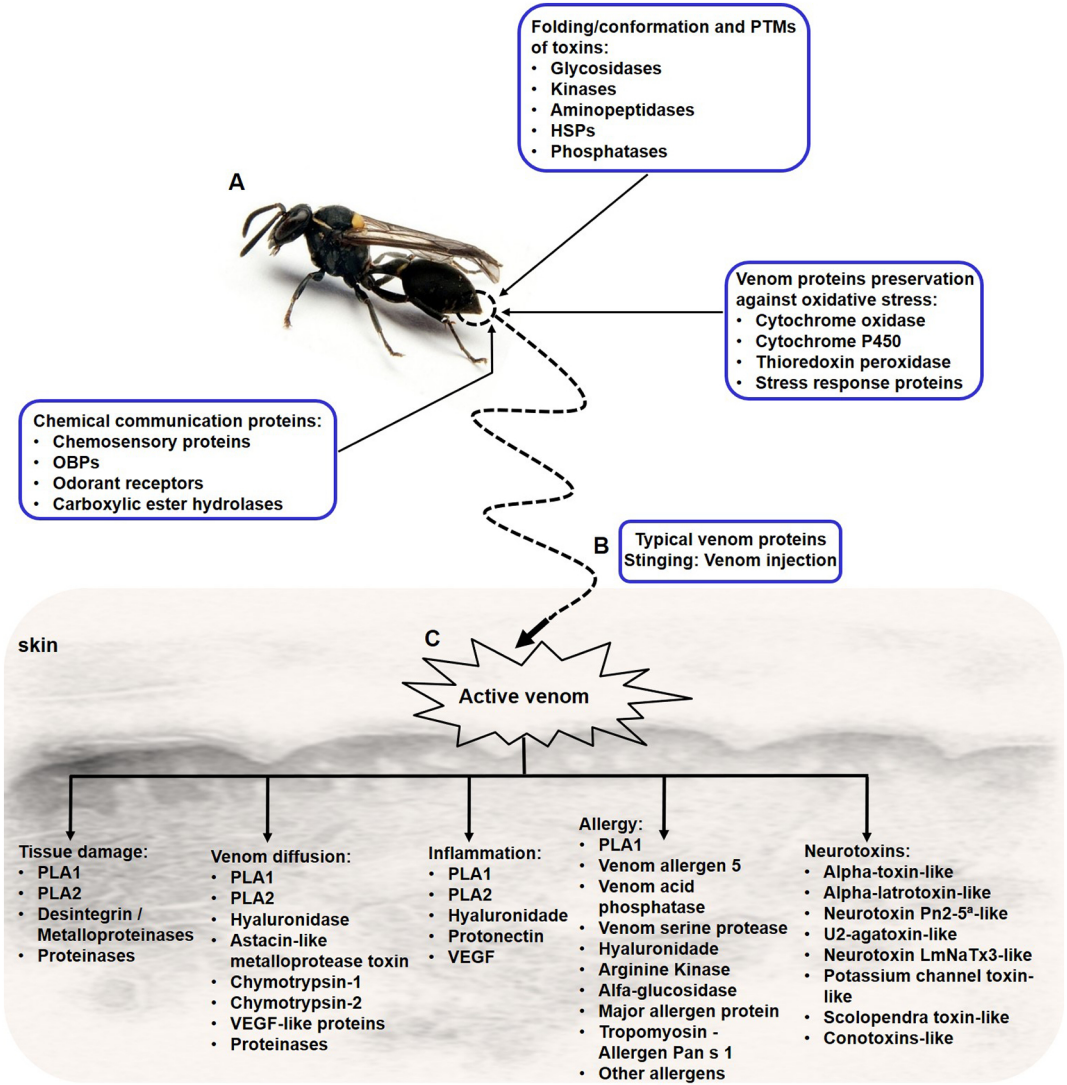

Folding/conformation and PTMs of toxins:
• Glycosidases
• Kinases
• Aminopeptidases
• HSPs
• Phosphatases
A
Venom proteins preservation against oxidative stress:
• Cytochrome oxidase
• Cytochrome P450
• Thioredoxin peroxidase
• Stress response proteins
Chemical communication proteins:
• Chemosensory proteins
• OBPs
• Odorant receptors
• Carboxylic ester hydrolases
B
Typical venom proteins
Stinging: Venom injection
skin
C
Active venom
Tissue damage:
• PLA1
• PLA2
• Desintegrin / Metalloproteinases
• Proteinases
Venom diffusion:
• PLA1
• PLA2
• Hyaluronidase
• Astacin-like metalloprotease toxin
• Chymotrypsin-1
• Chymotrypsin-2
• VEGF-like proteins
• Proteinases
Inflammation:
• PLA1
• PLA2
• Hyaluronidade
• Protonectin
• VEGF
Allergy:
• PLA1
• Venom allergen 5
• Venom acid phosphatase
• Venom serine protease
• Hyaluronidade
• Arginine Kinase
• Alfa-glucosidase
• Major allergen protein
• Tropomyosin - Allergen Pan s 1
• Other allergens
Neurotoxins:
• Alpha-toxin-like
• Alpha-latrotoxin-like
• Neurotoxin Pn2-5ª-like
• U2-agatoxin-like
• Neurotoxin LmNaTx3-like
• Potassium channel toxin-like
• Scolopendra toxin-like
• Conotoxins-like

Figure 5