

# Using Proteomic Strategies for Sequencing and Post-Translational Modifications Assignment of Antigen-5, a Major Allergen from the Venom of the Social Wasp *Polybia paulista*

José Roberto Aparecido dos Santos-Pinto,<sup>†,§</sup> Lucilene Delazari dos Santos,<sup>‡,§</sup> Helen Andrade Arcuri,<sup>§,||</sup> Fábio Morato Castro,<sup>§,||</sup> Jorge Elias Kalil,<sup>§,||</sup> and Mario Sergio Palma<sup>\*,†,§</sup>

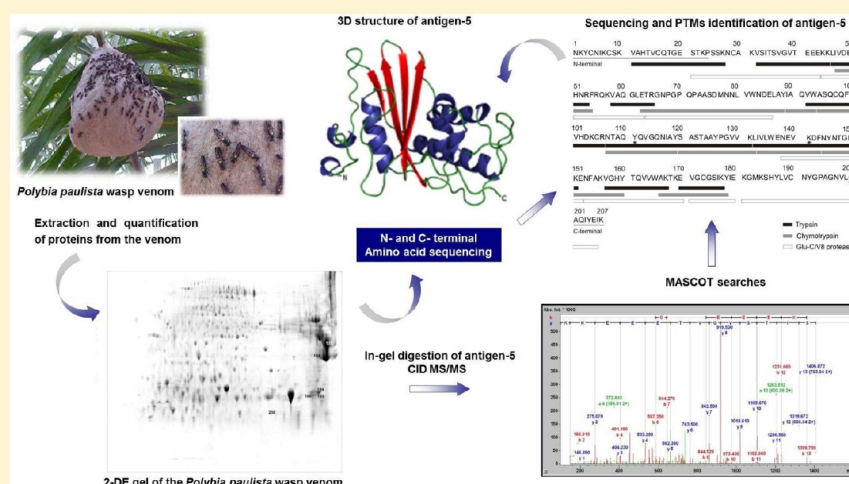
<sup>†</sup>Institute of Biosciences of Rio Claro, Department of Biology, Center of the Study of Social Insects, University of São Paulo State (UNESP), Rio Claro, SP, Brazil

<sup>‡</sup>Center for the Study of Venoms and Venomous Animals (CEVAP), University of São Paulo State (UNESP), Botucatu, SP, Brazil

<sup>§</sup>INCT-iii, São Paulo, Brazil

<sup>||</sup>Discipline of Allergy and Immunology (HC/Incor/FMUSP), SP, Brazil

## **S** Supporting Information



**ABSTRACT:** Antigen-5 is one of the major allergens identified in wasp venoms, and despite the fact that its biological function is still unknown, many studies have demonstrated its allergenicity. In this study, the biochemical and structural characterization of antigen-5 from the venom of the social wasp *Polybia paulista* are reported. A gel-based mass spectrometry strategy with CID fragmentation methods and classical protocols of protein chemistry, which included N- and C-terminal sequencing, were used to assign the complete sequence and determine the presence/location of the post-translational modifications (PTMs) of this protein. Six different isoforms of antigen-5 were identified in the crude venom of *P. paulista*; the most abundant, which corresponds to the intact form of this protein, was recognized by the pool of human specific-IgE. This protein was extensively sequenced through CID mass spectrometry, and a series of PTMs were observed such as hydroxylation, phosphorylation, and glycosylation. Sequence data revealed that this protein has 59.3–93.7% identity with antigen-5 proteins from other known vespid venoms. The molecular model of *P. paulista* antigen-5 shows that this protein has three  $\alpha$ -helices, one  $3_{10}$  helix, and four  $\beta$ -sheets covering 28 and 17.9% of the sequence, respectively. The identification and characterization of allergenic compounds is essential for the development of advanced component-resolved allergy diagnostics and treatment.

**KEYWORDS:** allergen, mass spectrometry, peptide sequence, post-translational modification, molecular modeling

## ■ ■ INTRODUCTION

The insect venoms of the Hymenoptera order contain a variety of allergenic proteins. Among the social Hymenoptera, the venoms of bees and wasps have been extensively studied, and many of their molecular components have been isolated and identified. Phospholipase, hyaluronidase, antigen-5,<sup>1–6</sup> and serine protease<sup>7</sup> are the major antigenic proteins found in

these venoms. The antigen-5 proteins were first reported in fire ant and wasp venoms from a temperate climate. This protein forms a major and distinct clade constituting the superfamily CAP, which is composed of cysteine-rich secretory proteins

**Received:** August 31, 2013

(CRISP), antigen-5, and pathogenesis-related-proteins (Pr 1);<sup>8</sup> they have molecular masses of approximately 23 kDa and unknown biological function. Subsequently, antigen-5-like proteins were also identified in *Drosophila* midgut,<sup>9</sup> in the saliva of ticks,<sup>10</sup> sandflies,<sup>11</sup> and mosquitoes,<sup>12</sup> and recently in the honeybee venom gland.<sup>13</sup> Due to the evolutionary diversity of proteins belonging to the CAP superfamily, several functional relationships have been proposed for these proteins, such as immune system regulation.<sup>8</sup> Within blood-feeding ticks, flies, and mosquitoes,<sup>10,14</sup> the antigen-5 proteins are part of a cocktail of salivary proteins that are believed to function either in the suppression of the host immune system or in the prevention of clotting to prolong feeding.<sup>15</sup> The antigen-5 proteins are the most abundant proteins in the wasp venoms, and they are often associated with allergic responses in humans; this protein was previously reported in the venoms from different species of social wasps of the *Polistes*, *Dolichovespula*, *Vespa*, *Vespula*, and *Polybia* genera.<sup>2,4,16–22</sup> With the exception of *P. paulista*, all the other species are endemic to the northern hemisphere. Although detailed characterizations of the allergenic responses have reported that the reactivity of this protein is specific to IgE and IgG,<sup>4,18–20</sup> the function of antigen-5 in wasp and fire ant venom still remains unknown. The primary structure and immunological aspects of the antigen-5 of venom from wasp species that are endemic to the northern hemisphere have been widely studied.<sup>22–25</sup> These studies reported that the identity between the sequences of antigen-5 in species from the same genus is approximately 98%, while among different genera, such as *Vespula* and *Polistes*, this value is approximately 57%. The high identity between the antigen-5 sequences from several social wasp species could explain the broad cross-reactivity between the proteins from different species.<sup>2,23</sup> Post-translational modifications (PTMs) such as glycosylation could also be a cause of cross-reactivity in patients who are allergic to Hymenoptera venom;<sup>24</sup> however, no study using a proteomic approach for characterizing PTMs in wasp venom allergens has been conducted previously. The three-dimensional structure of antigen-5 from *V. vulgaris* venom was determined by X-ray crystallography, revealing that it is composed of five  $\alpha$ -helices and four  $\beta$ -sheets.<sup>25</sup> There are only two studies characterizing antigen-5 from the venom of the neotropical social wasp species; using Edman degradation chemistry, the sequence of the antigen-5 from *Polybia rioplatense scutellaris* venom was assigned as presenting 207 amino acid residues, including eight cysteine residues forming four disulfide bonds, and to have a molecular mass of approximately 23 kDa and a pI of approximately 9.0.<sup>21,26</sup>

*P. paulista* is a very aggressive social wasp that is endemic to Southeast Brazil, where it causes hundreds of stinging incidents of medical importance for humans every year. Santos et al.<sup>4</sup> identified six different isoforms of antigen-5 in *P. paulista* venom, indicating the difficulty of purifying intact antigen-5 from this venom. Some of these forms were recognized by specific-IgE from the sera of *P. paulista* venom-sensitive patients. Thus, to obtain a better understanding regarding the biochemical and immunological aspects of this allergen, we used a proteomic approach combining two-dimensional electrophoresis (2-DE) with in-gel protein digestion by different proteolytic enzymes, followed by mass spectrometry analysis to determine the complete protein sequence and PTMs.

## EXPERIMENTAL SECTION

### *Polybia paulista* Venom

Workers of *Polybia paulista* were collected at the University Campus, at Rio Claro, SP, southeast Brazil and were immediately frozen and dissected. The venom reservoirs were removed and carefully washed and suspended in a solution containing a cocktail of protease inhibitors (2 mM AEBSE, 0.3  $\mu$ M Aprotinin, 130  $\mu$ M Bestatin, 1 mM EDTA, 14  $\mu$ M E-64, and 1  $\mu$ M Leupeptin, Sigma-Aldrich), thawed, punctured, washed three times with the protease inhibitors solution to extract the venom, and centrifuged at 10 000g for 10 min at 4 °C. The supernatants were collected, lyophilized, and maintained at –80 °C until use.

### Two-Dimensional Gel Electrophoresis

Venom samples containing 700  $\mu$ g of protein were applied by rehydration to 13 cm IPG strips, pH 3–10. Isoelectric focusing (IEF) was performed using an IPGphor system (GE Healthcare) at 3500 V for 17 000 Vh. After isoelectric focusing, the proteins were then reduced with 0.5% (w/v) DTT and alkylated with 4% (w/v) iodoacetamide for 15 min in equilibration buffer [50 mM Tris-HCl, pH 8.8, 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS]. The second dimension was run on casted SDS-PAGE gels [15% (w/v) polyacrylamide and 0.8% (w/v) bis (*N,N'*-methylenebisacrylamide)] at 15 mA/gel for 15 min and 30 mA/gel for 3 h, at 10 °C in a Ruby Red system (GE Healthcare). Gels were stained overnight with Coomassie Brilliant Blue R-250 (CBB) followed by scanning and digitizing (BioImage, GE Healthcare) for documentation. The images were analyzed using Image Master Platinum software v.7 (GE Healthcare).

### In-Gel Digestion

Protein spots were excised from the gel and destained twice for 30 min at 25 °C with 50 mM ammonium bicarbonate/50% (v/v) acetonitrile, dehydrated in acetonitrile, dried, and treated with three proteolytic enzymes: 20  $\mu$ g/mL trypsin (Promega, Madison, WI) in 25 mM ammonium bicarbonate, pH 7.8 at 37 °C for 18 h; 20  $\mu$ g/mL Chymotrypsin (Sigma, St. Louis, MO) in 25 mM ammonium bicarbonate, pH 7.8 at 30 °C for 3 h; and 20  $\mu$ g/mL Glu-C/V8 protease (Sigma, St. Louis, MO) in 50 mM ammonium bicarbonate, pH 7.8 at 37 °C for 18 h. Peptide extraction was performed with 0.5% (v/v) formic acid and 0.5% (v/v) formic acid in 30% (v/v) acetonitrile. The extracted peptides were pooled, desalted, and cleaned with PerfectPure C18 pipet tips (Eppendorf, Hamburg, Germany).

### NanoLC-ESI-CID

The HPLC used was a Nano-Advance UHPLC system (Bruker, Daltonics, Bremen, Germany) equipped with a PepMap100 C-18 trap column (300 mm  $\times$  5 mm) and a PepMap100 C-18 analytical column (75 mm  $\times$  150 mm). The gradient was (A - 0.1% FA in water, B - 0.08% FA in ACN) 4–30% B from 0 to 105 min, 80% B from 105 to 110 min, and 4% B from 110 to 125 min. An Amazon ETD mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with a CaptiveSpray source (Bruker, Daltonics, Bremen, Germany) was used to record peptide spectra over the mass range of *m/z* 350–3500 and MS/MS spectra in the information-dependent data acquisition mode over the mass range of *m/z* 100–3500. MS spectra were recorded followed by five data-dependent CID MS/MS spectra generated from the five highest intensity precursor ions. An active exclusion of 0.4 min after two spectra was used to detect low abundant peptides. The voltage between

the ion spray tip and the spray shield was set to 1500 V. The collision energy was set automatically according to the mass and charge state of the peptides chosen for fragmentation. Multiply charged peptides were chosen for the MS/MS experiments due to their good fragmentation characteristics. MS/MS spectra were interpreted, and peak lists were generated by Data Analysis 4.1 (Bruker Daltonics).

### Protein Identification

MASCOT searches for protein identification were performed using MASCOT v. 2.2 (<http://www.matrixscience.com>) against the most recent NCBI database. Database search parameters were set as follows: taxonomy - limited to other Metazoa; enzyme - selected as trypsin (or corresponding enzymes); maximum missed cleavage sites allowed -2; peptide mass tolerance - 0.2 Da for MS and 0.2 Da for MS/MS spectra; carbamidomethylation (C), methionine oxidation, and phosphorylation (Y, T, S) were specified in MASCOT as fixed and variable modifications, respectively. Positive protein identifications were based on a significant MOWSE score, and after protein identification, an error-tolerant search was conducted to detect unspecific cleavages and unassigned modifications. The resulting protein identification were manually inspected and filtered to obtain confirmed protein identification.

### Identification of Post-Translational Modifications

Phosphopeptide enrichment was performed using the following protocol. Acidified proteolytic digests from antigen-5 spots were enriched for phosphopeptides using the  $\text{TiO}_2$  suspension protocol previously reported by Zielinska et al.<sup>27</sup> with some modifications. Briefly, 25 mg of titanspheres  $\text{TiO}_2$  (10  $\mu\text{m}$ ; GL Sciences, Inc., Japan) was suspended in 50  $\mu\text{L}$  of a solution constituted of 0.8% (m/v) DHB, 20% (v/v) acetonitrile, and 0.05% (v/v) TFA in Milli-Q water. Ten microliters of this  $\text{TiO}_2$  suspension was added to the samples and incubated under shaking for 20 min. The titanspheres were then sedimented by centrifugation at 4500g for 90 s, and the supernatants were collected and used to make another suspension of titanspheres as described above. The pellets were resuspended in 150  $\mu\text{L}$  of 30% (v/v) MeCN containing 3% (v/v) TFA and transferred to a 200  $\mu\text{L}$  pipet tip, plugged with one layer of C8-bonded silica sorbents filter (3M). The  $\text{TiO}_2$  beads were washed four times with a solution of 25% (v/v) acetonitrile containing 2% (v/v) TFA, and the peptides were eluted from the beads with 100  $\mu\text{L}$  of 50% acetonitrile (v/v).

Enrichment of glycopeptides was performed using hydrophilic interaction chromatography (HILIC) in microcolumns, modified from a previously described protocol.<sup>28</sup> Briefly, 50  $\mu\text{L}$  of the proteolytic digest spots were mixed with 80  $\mu\text{L}$  of MeCN containing 0.5% (v/v) TFA and loaded onto a HILIC Uptitip (Interchim, France), equilibrated with 200  $\mu\text{L}$  of 80% MeCN containing 0.1% TFA. The bound peptides were washed twice with the same buffer and eluted with 50  $\mu\text{L}$  of 0.1% (v/v) TFA.

The PTMs identification were initially analyzed with the help of the software BioTools 3.2 (Bruker Daltonics) and then manually inspected and filtered to obtain confirmed protein modification.

### Immunoblotting

Sera from five *P. paulista*-sensitized patients were obtained in the Division of Clinical Immunology and Allergy from the Clinics Hospital, University of São Paulo Medical School, São Paulo, Brazil, with the approval of the University Ethical Board. Sera were collected before starting the specific immunotherapy

treatment of each patient and stored at  $-20^\circ\text{C}$  until used. The protein profiles obtained as described above were electrotransferred to nitrocellulose membranes (Hybond-C Extra, Amersham) for 1 h at room temperature. The binding of IgE antibodies to the membrane-immobilized allergens was analyzed by Western Blot using individual sera from five *P. paulista*-allergic patients. The membranes were blocked with PBS containing 0.5% Tween 20 and incubated with sera from each patient, diluted 1:10 in blocking buffer overnight at  $8^\circ\text{C}$ . After the membranes were washed with PBS containing 0.1% (v/v) Tween 20, the membranes were incubated with a secondary antibody (mouse IgG antihuman IgE) conjugated to peroxidase (1:10000, Zymed) for 1 h at room temperature. Chemiluminescence detection reagents (ECL Chemiluminescence Reagent Plus Western, GE Healthcare) were added to the membrane according to the manufacturer's instructions. The membrane was incubated with Hyperfilm film (GE Healthcare) in a X-ray cassette, and the film was submitted to conventional photographic film development.

### N-Terminal Amino Acid Sequencing

N-terminal sequence analysis of intact antigen-5 and its different forms were carried out with a pulse liquid-phase microsequencer (Shimadzu, mod, PPSQ-21) operated using standard programs. The samples used in this protocol were the antigen-5 spots that were electrotransferred to PVDF membranes. The protein profiles obtained as described above were electrotransferred to PVDF (Fluoreto de Polivinilideno) membrane in transfer buffer (20 mM tris, 192 mM glycine, 0.1% SDS pH 8.3 in 10% methanol). It was performed using a Semy-Dry TE 70 system (GE Healthcare) at 250 V, 200 mA for 1 h at room temperature. After that, the membranes were stained with Coomassie Brilliant Blue R-250 (CBB) [0.025% (m/v) in 40% (v/v) methanol] for 5 min.

### C-Terminal Amino Acid Sequencing

The C-terminal sequencing was performed based on derivatization with acetylthiocyanate to yield amino acid thiohydantoins (TH-AAs). This simple chemistry was automated using an ABI 473A N-terminal sequencer (all reagents, R1: trimethylsilylthiocyanate; R3: alkaline thiocyanate for cleavage), and solvents required for sequencing were accommodated on the sequencer, which was modified to deliver liquid R2 (acetyl chloride) to the reaction vessel. The conversion flask was used for preparing the TH-AAs for analysis by online HPLC using a graphitized carbon (Hypercarb) column. The samples used in this protocol were the antigen-5 spots that were electrotransferred to PVDF membranes. This setting permitted the determination of at least four residues from the C-terminus.

### Secondary Structure Analysis and Molecular Modeling

The secondary structure prediction for antigen-5 was performed by the JNET Secondary Structure Prediction program<sup>29</sup> and edited by the JALVIEW program.<sup>30</sup> Antigen-5 molecular modeling used the restrained-based modeling approach as implemented in the program MODELER 9v4. MODELER is an automated approach for comparative modeling by the fulfillment of spatial restraints. A total of 1000 models were generated, and the final model was selected based on stereochemical quality and MODELER objective function. The images of three-dimensional structures of the models were generated from the program PyMOL.<sup>31</sup> The PROCHECK<sup>32</sup> program was used to check bond lengths, bond

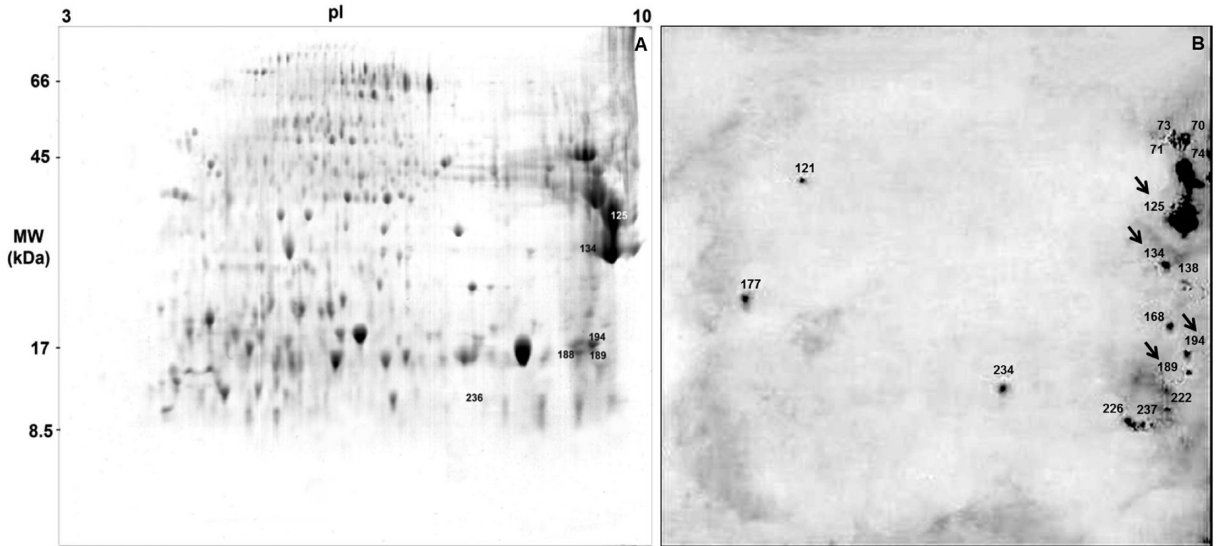


**Figure 1.** (A) Representative 2-DE profile of the *P. paulista* wasp venom, showing the six different isoforms of antigen-5; (B) submitted to immunoblotting with the sera of patients sensitive to *P. paulista* venom, revealing the protein spots recognized by human specific-IgE. The four antigen-5 forms (125, 134, 189, and 194) recognized by human specific-IgE are marked with a black arrow.

**Table 1. Identification on the 2-DE Gels of Antigen-5 from the *Polybia paulista* Wasp Venom**

| spot | protein   | access code | Mascot score <sup>a</sup> | coverage % | MW (Da) | pI   | peptide sequences (ion score)                                                                               |
|------|-----------|-------------|---------------------------|------------|---------|------|-------------------------------------------------------------------------------------------------------------|
| 125  | antigen-5 | Q7Z156      | 221                       | 33         | 26.479  | 9.29 | VSITSVGVTEEEKK (53), LIVDEHNRFR (31), QKVAQGLETR (35), DFNYNTGITK (29), VGHYTVVWAKTK (42), EVGCGSIKYEK (36) |
| 134  | antigen-5 | Q7Z156      | 206                       | 20         | 22.102  | 9.39 | LIVDEHNRFR (32), QKVAQGLETR (45), DFNYNTGITK (38), VGHYTVVWAKTK (40)                                        |
| 188  | antigen-5 | Q7Z156      | 143                       | 13         | 16.188  | 9.65 | VAHTVCQTGESTKPKSSK (36), LIVDEHNRFR (48)                                                                    |
| 189  | antigen-5 | Q7Z156      | 132                       | 18         | 16.188  | 9.40 | VAHTVCQTGESTKPKSSK (42), LIVDEHNRFR (30), DFNYNTGITK (51)                                                   |
| 194  | antigen-5 | Q7Z156      | 189                       | 16         | 18.557  | 9.50 | VSITSVGVTEEEKK (49), LIVDEHNRFR (37), DFNYNTGITK (63)                                                       |
| 236  | antigen-5 | Q7Z156      | 118                       | 17         | 12.013  | 8.13 | DFNYNTGITK (56), VGHYTVVWAKTK (34), EVGCGSIKYEK (28)                                                        |

<sup>a</sup>Mascot score: protein scores greater than 80 are significant ( $p < 0.05$ ).

**Table 2. N-Terminal and C-Terminal Sequences of the Proteins Spots Corresponding to the Different Isoforms of Antigen-5**

| spot | N-terminal sequence | C-terminal sequence | start–end | experimental MW (Da) | calculated MW (Da) |
|------|---------------------|---------------------|-----------|----------------------|--------------------|
| 125  | NKYCNIKCSKVAHT...   | ...GAQIYEIK         | 1–207     | 26.479               | 22.994             |
| 134  | NKYCNIKCSKVAHT...   | ...GPAGNVL          | 1–199     | 22.102               | 22.091             |
| 194  | NKYCNIKCSKVAHT...   | ...IKYIEKG          | 1–182     | 18.557               | 20.245             |
| 189  | NKYCNIKCSKVAHT...   | ...FAKVGHYT         | 1–161     | 16.188               | 17.940             |
| 188  | NKYCNIKCSKVAHT...   | ...NFAKVGH          | 1–159     | 16.188               | 17.676             |
| 236  | ASQCQFFVHDKCRN...   | ...GAQIYEIK         | 94–207    | 12.013               | 12.727             |

angles, peptide bonds, and side-chain ring planarities, chirality, main-chain and side-chain torsion angles. The results of this analysis are shown in Ramachandran diagrams and the values of the complete G-factor. VERIFY 3D program, which measures the compatibility of a protein model with its sequence using a 3D profile, was also used for model validation.<sup>33</sup> The RMSD differences from ideal geometries for bond lengths and bond angles were calculated with XPLOR.<sup>34</sup>

### Phylogenetic Analysis

Phylogenetic analysis was performed to show the evolution of the structural relationship between the sequences of the antigen-5 from the venoms of social wasps using the bioinformatics tool “Phylogeny Analysis” described by Dereeper et al.<sup>35</sup>

### RESULTS

Considering the complexity of *P. paulista* venom composition and the difficulty of purifying intact antigen-5 in this venom, a proteomic approach was used to detect and characterize the structure of the allergenic protein. The 2-DE profile of *P. paulista* venom exhibits six different isoforms of antigen-5 which were assigned from spots 125, 134, 188, 189, 194, and 236, with molecular weights that ranged from 12.013 to 26.479 Da, and pI values that ranged from 8.13 to 9.65 (Figure 1A), and GenBank ID: Q7Z156 in Protein Database of National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov>). The proteomic data identifying these six isoforms are shown in Table 1. The different isoforms of antigen-5 were identified with protein scores from 118 to 221 and sequence coverage from 13 to 33%. To verify the integrity of the six isoforms of antigen-5, the proteins from the 2-DE

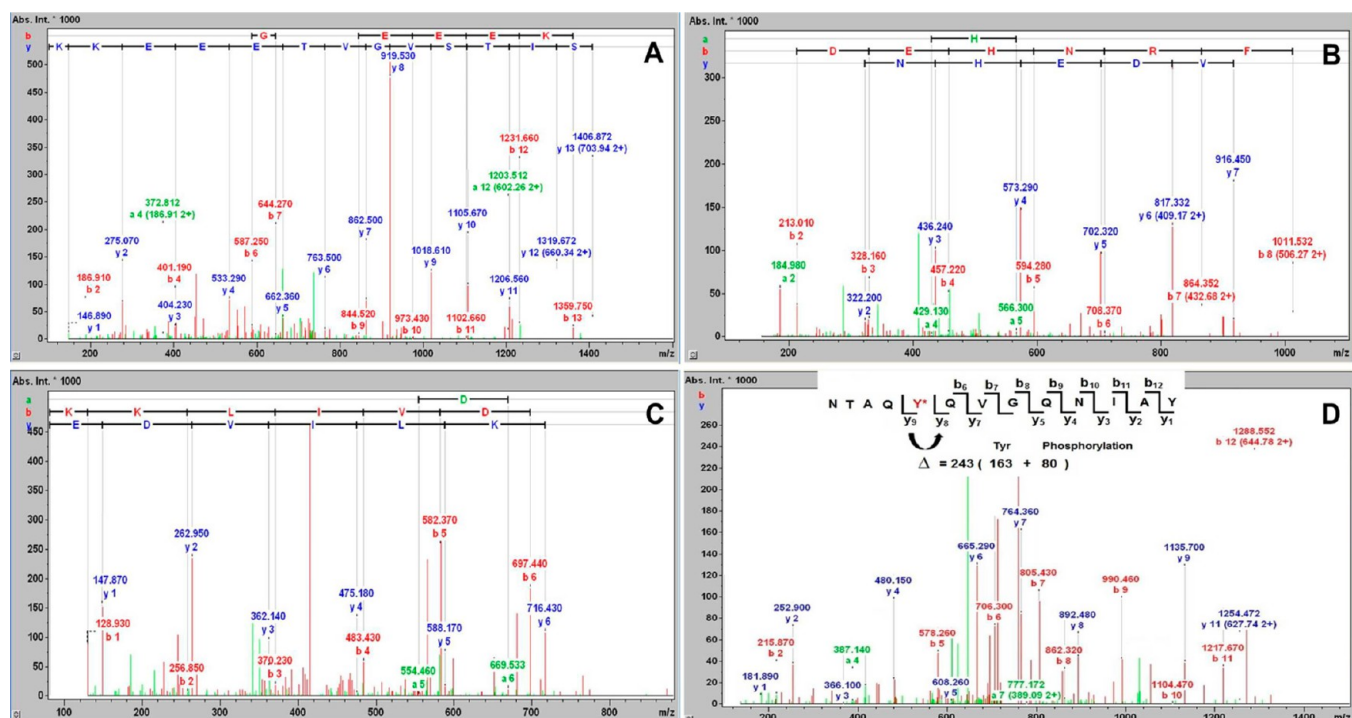


**Figure 2.** Representative CID spectra of antigen-5. (A) CID spectrum of tryptic peptide K.VSITSVGVTEEEKK.L (32–45), obtained by selecting the ion of  $m/z$  753.449  $[M + 2H]^{2+}$  as precursor ion; (B) CID spectrum of chymotryptic peptide L.IVDEHNRF.R (47–57), obtained by selecting the ion of  $m/z$  515.254  $[M + 2H]^{2+}$  as precursor ion; (C) CID spectrum of proteolytic peptide from Glu-C/V8 protease E.KKLIVDE.H (44–50), obtained by selecting the ion of  $m/z$  422.790  $[M + 2H]^{2+}$  as precursor ion; and (D) CID MS/MS spectrum of chymotryptic peptide R.NTAQY\*QVGQNIAY.S (107–119), obtained by selecting the ion of 775.3702  $[M + 2H]^{2+}$  as precursor ion; this spectrum is also indicating the site phosphorylation at the Y111, assigned as Y\*.

were electro-transferred to PVDF membranes, and the spots corresponding to the six isoforms of antigen-5 (spots 125, 134, 188, 189, 194, and 236) were excised and submitted to N- and C-terminal sequencing. The results are shown in the Table 2. The N- and C-terminal sequences of all antigen-5 forms were compared to the sequence of the same allergen isolated from *P. scutellaris rioplatensis* venom (accession code: Q7Z156 in Protein DB of NCBI) (see Figure S1 in the Supporting Information). This wasp species is taxonomically close related to *P. paulista*,<sup>36,37</sup> as well the primary sequence of both antigen-5 molecules form a close related group in the phylogenetic analysis (see Figure S2 in the Supporting Information). The comparison of sequences revealed that the N- and C-termini of the protein present in spot 125 matched exactly to those reported for the mature antigen-5 from *P. scutellaris rioplatensis*. Additionally, the protein present in spot 125 corresponded to the form that had the highest MW among the six detected isoforms. The comparison between the experimental MW values obtained in the 2-DE and those corresponding to the hypothetical molecular masses that were calculated for each different isoform as shown in Table 2 reveals a good agreement of values.

The results shown in a previous publication<sup>4</sup> revealed that the antigen-5 from spots 125 and 134 were detected in the glycosylated form and that spots 125, 134, 189, and 194 were recognized by the pool of human specific-IgE revealing the allergenic potential of these protein isoforms (Figure 1B); in fact, sixteen allergenic proteins were detected on the immunoblotting; however, those corresponding to the antigen-5 are indicated by black arrows in Figure 1B. Considering the well-known allergenicity of *P. paulista* venom

to humans, the reactivity of spot 125 to human specific-IgE as previously reported,<sup>4</sup> and the apparent intact isoform of the protein present in this spot, we decided to sequence and to assign the PTMs of the antigen-5 that was present in spot 125 combining in-gel digestion using different proteolytic enzymes with tandem mass spectrometry analysis under CID conditions.

The amino acid sequence of *P. paulista* venom antigen-5 was determined using mass spectrometry analysis of the proteolytic peptides generated by digestion of spot 125 with trypsin, chymotrypsin, and Glu-C/V8 protease. The complete assignment of the protein sequence and a series of PTMs were determined.

Representative CID spectra of some proteolytic peptides are shown in Figure 2. Figure 2A shows the CID spectrum of the tryptic peptide K.VSITSVGVTEEEKK.L (sequence position from 32 to 45), selecting the  $m/z$  753.449  $[M + 2H]^{2+}$  as a precursor ion. Figure 2B shows the CID spectrum of the chymotryptic peptide L.IVDEHNRF.R (sequence position 47 to 54), selecting the  $m/z$  515.254  $[M + 2H]^{2+}$  as a precursor ion. Figure 2C shows the CID spectrum of the proteolytic peptide E.KKLIVDE.H (sequence position 44 to 50) obtained by cleavage with Glu-C/V8 protease, selecting the  $m/z$  422.790  $[M + 2H]^{2+}$  as a precursor ion. Figure 2D shows the CID spectrum of the chymotryptic peptide R.NTAQY\*QVGQNIAY.S (sequence position 107 to 119), selecting the  $m/z$  775.3702  $[M + 2H]^{2+}$  as a precursor ion; this spectrum also shows the phosphorylation site observed at position Y111 of the antigen-5 protein. At this time, it is important to emphasize that the sample relative to the chymotryptic digest of spot 125 was enriched for phosphopeptides as described in the Experimental Section, but similar results were observed for

**Table 3. Amino Acid Sequence of Antigen-5, Spot 125, from the Venom of the Social Wasp *P. paulista* by In-Gel Protein Digestion Using Different Proteolytic Enzymes. Enzyme Used, Amino Acid Position, *m/z* Value Observed (Z), Delta between Experimental Mass and Theoretical Mass, Number of Missed Cleavage Sites, Peptide Sequences, PTMs, and Mascot Ion Score Were Listed for All Identified Peptides**

| enzyme            | amino acid position | <i>m/z</i> value observed (Z) | delta   | miss | peptide sequences                | PTMs                                    | ion score |
|-------------------|---------------------|-------------------------------|---------|------|----------------------------------|-----------------------------------------|-----------|
| trypsin           | 11–27               | 626.9700 (+3)                 | 0.0091  | 1    | K.VAHTVCQTGESTKPSSK.N            |                                         | 59        |
| trypsin           | 32–44               | 689.3504 (+2)                 | −0.0172 | 0    | K.VSITSVGVTEEEK.K                |                                         | 62        |
| trypsin           | 32–45               | 753.4493 (+2)                 | 0.0856  | 1    | K.VSITSVGVTEEEK.L                |                                         | 79        |
| trypsin           | 46–53               | 498.2506 (+2)                 | −0.0329 | 0    | K.LIVDEHNRF                      |                                         | 50        |
| trypsin           | 58–65               | 437.2318 (+2)                 | −0.0225 | 0    | K.VAQGLETR.G                     |                                         | 48        |
| trypsin           | 92–106              | 826.4501 (+2)                 | 0.0621  | 1    | Q.VWASQC*QFFVHDKCR.N             | carbamidomethyl (C97)                   | 56        |
| trypsin           | 107–131             | 778.8001 (+4)                 | 0.0211  | 0    | R.NTAQYQVGQNIAYASTAAYPGVVK.L     |                                         | 47        |
| trypsin           | 132–141             | 621.8439 (+2)                 | −0.0287 | 0    | K.LIVLWENEVK.D                   |                                         | 50        |
| trypsin           | 132–151             | 853.4721 (+3)                 | 0.0993  | 1    | K.LIVLWENEVK*DFNYNTGITKE         | hexose (K141)                           | 49        |
| trypsin           | 142–151             | 586.7868 (+2)                 | 0.0081  | 0    | K.DFNYNTGITKE                    |                                         | 35        |
| trypsin           | 157–167             | 644.3464 (+2)                 | 0.0011  | 0    | K.VGHYTQVWVAK.T                  |                                         | 44        |
| trypsin           | 170–177             | 425.1986 (+2)                 | −0.0236 | 0    | K.EVGCGSIKY                      |                                         | 27        |
| chymotrypsin      | 47–54               | 515.2547 (+2)                 | −0.0091 | 0    | L.IVDEHNRF.R                     |                                         | 45        |
| chymotrypsin      | 63–80               | 934.8203 (+2)                 | 0.0032  | 0    | L.ETRGNPGPQPAASDMNNLV            |                                         | 36        |
| chymotrypsin      | 63–82               | 1077.4332 (+2)                | −0.1441 | 1    | L.ETRGNPGPQPAASDMNNLVW.N         |                                         | 55        |
| chymotrypsin      | 63–88               | 959.0901 (+3)                 | −0.0132 | 3    | L.ETRGNPGPQPAASDMNNLVWNDELAY.I   |                                         | 99        |
| chymotrypsin      | 63–88               | 969.6701 (+3)                 | −0.0965 | 3    | L.ETRGNPGPQPAASDM*NNLVWND*ELAY.I | oxidation (M77);<br>hydroxylation (D84) | 67        |
| chymotrypsin      | 82–87               | 724.2705 (+1)                 | −0.0443 | 1    | W.NDELAY.I                       |                                         | 38        |
| chymotrypsin      | 89–93               | 616.3501 (+1)                 | 0.0047  | 0    | Y.IAQVW.A                        |                                         | 66        |
| chymotrypsin      | 94–100              | 444.1801 (+2)                 | 0.0061  | 1    | W.ASQC*QFF.V                     | carbamidomethyl (C97)                   | 56        |
| chymotrypsin      | 107–119             | 775.3702 (+2)                 | 0.0154  | 1    | R.NTAQY*QVGQNIAY.S               | phosphorylation (Y111)                  | 72        |
| chymotrypsin      | 107–121             | 814.3810 (+2)                 | −0.0092 | 2    | R.NTAQYQVGQNIAYSA.S              |                                         | 54        |
| chymotrypsin      | 111–124             | 736.8401 (+2)                 | −0.0144 | 1    | Q.YQVGQNIAYASTA.A                |                                         | 48        |
| chymotrypsin      | 112–119             | 446.7012 (+2)                 | 0.0298  | 0    | Y.QVGQNIAY.S                     |                                         | 71        |
| chymotrypsin      | 120–132             | 618.7801 (+2)                 | −0.0472 | 1    | Y.SASTAAYPGVVKLI                 |                                         | 42        |
| chymotrypsin      | 133–143             | 696.3718 (+2)                 | 0.0158  | 2    | L.IVLWENEVKDF.N                  |                                         | 66        |
| chymotrypsin      | 137–143             | 440.7000 (+2)                 | −0.0120 | 0    | W.ENEVKDF.N                      |                                         | 30        |
| chymotrypsin      | 144–160             | 657.9800 (+3)                 | −0.0102 | 2    | F.NYNTGITKENFAK*VGHY.T           | hydroxylation (K156)                    | 76        |
| chymotrypsin      | 146–154             | 1023.4601 (+1)                | −0.0012 | 0    | Y.NTGITKENF.A                    |                                         | 46        |
| chymotrypsin      | 166–178             | 720.8277 (+2)                 | −0.1034 | 0    | W.AKTKEVGCGSIKY.I                |                                         | 52        |
| Glu-C/V8 protease | 21–43               | 789.7265 (+2)                 | 0.0305  | 2    | E.STKPSSKNC AKVSITSVGVTEEEK      |                                         | 41        |
| Glu-C/V8 protease | 44–50               | 422.7906 (+2)                 | 0.0601  | 0    | E.KKLIVDE.H                      |                                         | 28        |
| Glu-C/V8 protease | 51–63               | 791.9190 (+2)                 | −0.0253 | 0    | E.HNRFQKVAQGLE.T                 |                                         | 30        |
| Glu-C/V8 protease | 64–85               | 1200.0242 (+2)                | −0.0269 | 0    | E.TRGNPGPQPAASDM*NNLVWNDEL       | oxidation (M77)                         | 79        |
| Glu-C/V8 protease | 138–152             | 886.4422 (+2)                 | 0.0274  | 1    | E.NEVKDFNYNTGITKE.N              |                                         | 70        |
| Glu-C/V8 protease | 140–152             | 764.8753 (+2)                 | −0.0209 | 0    | E.VKDFNYNTGITKE.N                |                                         | 60        |
| Glu-C/V8 protease | 153–170             | 531.2801 (+4)                 | 0.0021  | 0    | E.NFAK*VGHYTQVWVAKTKE.V          | hydroxylation (K156)                    | 62        |
| Glu-C/V8 protease | 171–180             | 563.2602 (+2)                 | −0.0477 | 0    | E.VGCGSIKYIE.K                   |                                         | 55        |
| Glu-C/V8 protease | 181–207             | 781.1433 (+4)                 | −0.0084 | 1    | E.KGMKSHYLCVNYGPAGNVLGAQIYEIK    |                                         | 33        |

this PTM without the treatment for enrichment. The cleavage of protein from spot 125 with proteases followed by tandem mass spectrometry analysis permitted the detection and sequence assignment of 60.86% coverage of the protein digested with trypsin, 57.97% coverage of the protein digested with chymotrypsin, and 65.21% coverage of the protein digested with Glu-C/V8 protease. All the proteolytic peptides used to assign the complete sequence of *P. paulista* antigen-5 are shown in Table 3, and the corresponding CID spectra are

shown in Figures S3 to S41 (Supporting Information). Taking into account the partial overlapping of common sequences and the complementarity of the proteolytic fragments analyzed by tandem mass spectrometry with those obtained from N- and C-terminal sequencing, the sequence of the antigen-5 from *P. paulista* venom was completely assigned (Figure 3). The protein had 207 amino acid residues in its sequence, which is consistent with a theoretical molecular mass of 22.994 Da (without considering the occurrence of PTMs).



**Figure 3.** Representative complete amino acid sequence of antigen-5. Sequence results were obtained from N- and C-terminal sequences and protein digestion protocols by using trypsin, chymotrypsin, and Glu-C/V8 protease enzymes.

The sequence alignment of *P. paulista* antigen-5 with the sequences of other wasps venom antigen-5 is shown in Figure 4, which included *P. scutellaris rioplatensis* (VA5\_POLSCR), *Polistes gallicus* (VA5\_POLGA), *Polistes annularis* (VA5\_POLAN), *Polistes dominula* (VA5\_POLDO), *Dolichovespula maculata* (VA5\_DOLMA), and *Vespa vulgaris* (VA5\_VESVU). Identity and similarity values among the primary sequences of these allergens were 93.7 and 96.1%, respectively, in relation to VA5\_POLSCR; 78.3 and 89.9%, respectively, in relation to VA5\_POLGA; 78.0 and 88.8%, respectively, in relation to VA5\_POLAN; 77.8 and 90.3%, respectively, in relation to VA5\_POLDO; 61.2 and 82.3%, respectively, in

relation to VA5\_DOLMA; and 59.3 and 83.3%, respectively, in relation to VA5\_VESVU.

With regard to the importance of antigen-5 as an allergen in Vespidae venom, we decided to perform molecular modeling of the protein from *P. paulista* venom because it is necessary to complete its structural characterization for the future understanding of its interaction with IgE. *P. paulista* antigen-5 was modeled using the allergen from *V. vulgaris* (VA5\_VESVU) as a template protein (PDB ID: 1QNX), which was resolved by X-ray crystallography with a resolution of 1.9 Å.<sup>25</sup> The secondary structure analysis for *P. paulista* venom antigen-5 was performed using the JNET Secondary Structure Prediction and PROCHECK, which estimated the secondary structure content as 28%  $\alpha$ -helix, 17.9%  $\beta$ -sheets, and 54.1% turns/random structures (Figures 5 and 6). The Ramachandran plots for the

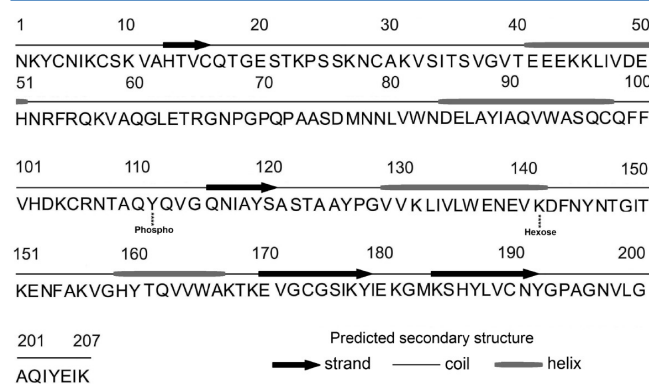


**Figure 5.** Comprehensive characterization of sequence, secondary structure, and the PTMs assignment in *P. paulista* antigen-5.

template and model indicate that over 90% of the residues are in the most favorable regions (data not shown). The analysis of the structural quality of the homology model was performed

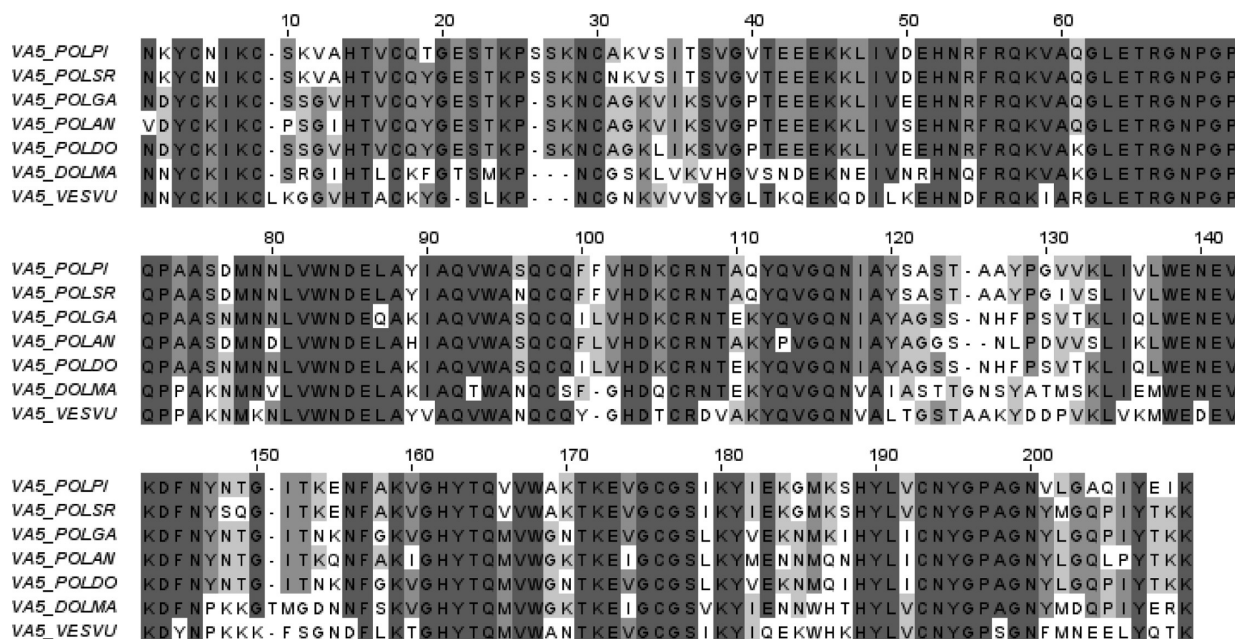


**Figure 4.** Multiple alignment among the primary sequence of the antigen-5 from the social wasp *P. paulista* (VA5\_POLPI) with the primary sequence of the wasps *Polybia scutellaris rioplatensis* (VA5\_POLSCR), *Polistes gallicus* (VA5\_POLGA), *Polistes annularis* (VA5\_POLAN), *Polistes dominula* (VA5\_POLDO), *Dolichovespula maculata* (VA5\_DOLMA), and *Vespa vulgaris* (VA5\_VESVU). The most important regions of identity are marked in dark gray, highlighting the highly conserved residues in this type of alignment.



**Figure 6.** Three-dimensional molecular model of the allergen antigen-5 from the venom of the social wasp *P. paulista*.

using PROCHECK (for Ramachandran plots and G-factor [torsion angles =  $-0.03$ ; covalent geometry =  $0.27$ ; global G-factor =  $0.09$ ]), XPLOR (for RMSD from ideal geometry [bond lengths =  $0.023$  Å; bond angles =  $1.8$  Å]) and VERIFY-3D (quality =  $94.7\%$ ), which strongly indicated that the model was adequate for structural studies.

The structure of *P. paulista* antigen-5 has the classical  $\alpha$ - $\beta$ - $\alpha$  fold of the antigens 5 that are reported in the literature, which seems to be unique among the allergens in general. The  $\beta$ -sheet is composed of four-stranded antiparallel elements of the secondary structure (residues 13–16, 115–120, 170–178, 184–191) that are sandwiched between one long  $\alpha$ -helix (residues 84–97) and a group of helices composed of two  $\alpha$ -helices (residues 41–51 and 129–141) and one helix  $3_{10}$  (residues 159–166) (Figure 5). Figure 6 shows some regions of the protein that present helix-like secondary structures, which were not represented in Figure 5 because they are very short (less than four amino acid residues). The complete sequence has eight cysteine residues that apparently form four disulfide bridges that are responsible for the rigidity of the  $\alpha$ - $\beta$ - $\alpha$  fold. This structure is similar to that previously reported for the allergen-5 from *V. vulgaris*<sup>25</sup> but presents a slightly smaller number of secondary structure elements than the protein that was used as its structural template.

With respect to the importance of the PTMs for the functionality of the proteins, specifically of an allergen, the careful examination and interpretation of the CID spectra permitted the assignment of some PTMs in the sequence of *P. paulista* venom antigen-5. Thus, the following PTMs were observed: hydroxylation at D84 (Figures S19) and K156 (Figures S30 and S39); oxidation at M77 (Figures S19 and S36); carbamidomethylation at C97 and C105 (Figures S8 and S22, respectively); phosphorylation at Y111 (Figure S23); and glycosylation at K141 (Figure S11). As already mentioned for the detection of phosphopeptides, a sample of the chymotryptic digest of the spot 125 was also treated for enrichment of glycosylated peptides, but a similar result was obtained without this treatment. The potential importance of these PTMs is discussed later in this manuscript.

## DISCUSSION

The composition of Vespid venoms has been extensively studied, and the major allergenic components have been identified.<sup>4,38</sup> In general, there is from 2 to 17  $\mu$ g of protein in a vespid venom sac, with 3.3% being phospholipase A<sub>1</sub>, 1.5%

hyaluronidase, and 8.1% antigen-5, which can cause allergic responses in humans.<sup>39,40</sup> Antigen-5 is one of the major allergens found in the venoms of social Hymenoptera. Accordingly, the biochemical and structural characterization of *P. paulista* venom antigen-5 is very important for the future development of epitope mapping of this allergen.

Proteomic analysis confirmed the identification of six isoforms of antigen-5, which were identified with the same accession code (Q7Z156). This prompted us to postulate whether these different isoforms were created by an experimental artifact during sample manipulation or if they were naturally created within the venom. However, since all the venom samples were obtained and manipulated in the presence of a cocktail of proteinase inhibitors, it could be suggested that the different isoforms of antigen-5 in *P. paulista* venom were naturally produced prior to venom collection and extraction. Apparently, these isoforms could have originated from the controlled proteolysis of the intact protein or by gene splicing regulation. The differences between the molecular weights are shown in Table 1 for the different isoforms of antigen-5; therefore, the occurrence of PTMs will be discussed later in this manuscript.

In the present investigation, the analysis of the N- and C-terminal sequences of the six natural isoforms of antigen-5 revealed that one of these isoforms corresponds to the mature and intact antigen-5 (spot 125 in Figure 1A), while the other five isoforms apparently correspond to truncated forms of this protein. These isoforms likely originated from a common protein sequence (i.e., the intact one). The ability of the venomous organisms to produce different isoforms of the same proteins in their venoms increases the diversity of their proteome; however, it is difficult to know the exact biological function of this mechanism in animal venoms. The occurrence of different isoforms of a venom protein could be related to their different molecular targets in the victims of envenoming and/or to the regulation of different levels of affinity by IgE.<sup>4,3</sup>

The experimental approach of combining different proteomic protocols such as 2-DE with multiple proteolytic in-gel protein digestion enzymes followed by MS analysis under CID conditions and classical protocols of protein chemistry, which included N- and C-terminal sequencing, allowed the generation of high sequence coverage of antigen-5 from spot 125. The sequence of the antigen-5 from *P. paulista* venom was revealed to be highly conserved in relation to the Vespidae proteins in general.

The high identity between the sequences of antigen-5 from the venoms of different species of wasps could explain the broad cross-reactivity between the allergens from different species.<sup>25</sup> Cross-reactivity has also been observed between antigen-5 from wasp venom with proteins belonging to the CAP family and from other animals.<sup>2,8,23</sup> Because the complete assignment of antigen-5 sequence from the wasp venom *P. paulista* was obtained in the present study, it was possible to perform the molecular modeling of the 3-D structure of this allergen.

The molecular model revealed that this protein has seven elements of  $\alpha$ -helix (28%) and four antiparallel  $\beta$ -strand structures (17.9%), which constitute a  $\beta$ -sheet. Furthermore, 54.1% of the sequence comprise turns and/or random structures. The following estimated percentage of the elements of secondary structure for this protein are similar to those obtained in studies by Pirpignani et al.<sup>21</sup> for the antigen-5 from *P. scutellaris rioplatensis* venom: 22%  $\alpha$ -helix structures, 22%  $\beta$ -

sheets structures, and 56% turns and random structures, which corroborates the phylogenetic proximity between these species. Apparently, the protein contains four disulfide bridges, which are characteristic of the CAP family and are responsible for the rigidity of the structural  $\alpha$ - $\beta$ - $\alpha$  fold that is typical of the representatives of the CRISP family. This group of proteins also includes the Pr-1 proteins, such as the human glioma pathogenesis-related-1 protein and the Golgi-associated pathogenesis-related-1 proteins. The biological activity/function of wasp venom allergen-5 is currently unknown; however, the members of the CAP family are secreted proteins that present extracellular endocrine/paracrine functions and sometimes act as tumor suppressors agents.<sup>41</sup>

Previously, the known activity for antigen-5 was related to the reactivity with a specific-IgE; thus, the presence of this protein in the venom of social wasps certainly contributes to the known allergy caused by these venoms to humans. There are no previous reports about the contribution of the PTMs of antigen-5 for its immunoreactivity to specific antibodies, especially in allergies. The oxidation of the M77 is apparently related to an artifact of sample manipulation in the presence of light and/or heat from the environment. The carbamidomethylation of the cysteine residues is due to the reaction of reduction and alkylation of the Cys residues during sample preparation for 2-D SDS-PAGE. The alkylations were also detected in the eight Cys residues; however, only the spectra showing the carbamidomethylation at C97 and C105 (Figures S8 and S22, respectively) were reliable enough to be included in the Supporting Information.

Hydroxylation was observed at D84 and K156 of *P. paulista* antigen-5. Studies conducted by Petersen et al.<sup>42</sup> and Fenaile et al.<sup>43</sup> reported the occurrence of the hydroxylation of proline residues in the Ph1 p1, a major allergen from grass pollen, and demonstrated the importance of these modified residues to enhance the reactivity to specific-IgE. Hydroxyprolines were also reported in the major peanut allergen Ara h 2, but apparently the PTMs were not located in the sequences of IgE-epitopes.<sup>44</sup> The results obtained for *P. paulista* antigen-5 will open the possibility of novel investigations to determine the role of hydroxylation of the side chain of the amino acid residues in the allergen stability and/or immunoreactivity.

Another type of PTM observed in *P. paulista* antigen-5 was the phosphorylation site at Y111. The observation that the phosphorylation and glycosylation were reliably identified both in enriched and non-enriched samples suggest that these PTMs occur in relatively high abundance under natural conditions. The examination of the spectral data suggest that the abundance ratio between phosphorylated and the non-phosphorylated forms was about 1:1. Recently, the phosphoproteome from European and Africanized honeybee venoms was profiled without need of enrichment.<sup>45</sup> If this is generally a characteristic of Hymenopteran venom proteins, or constitutes particular situations, still must be observed in future investigations. Predictions of phosphorylation positions for the complete sequence of *P. paulista* antigen-5 with the tool NetPhos 2.0 server ([http://expasy.org/proteomics/post-translational\\_modification](http://expasy.org/proteomics/post-translational_modification)) indicated potential phosphorylation at the sites Y3, S21, S25, S26, S36, S75, Y88, T168, and Y178. Thus, the phosphorylation at the position Y111 seems to be uncommon in the database used as reference by NetPhos 2.0 server.

In general, the introduction of one or more phosphate groups onto amino acids is known to induce significant protein

conformational changes and consequently, profound effects on protein activity and protein-protein interactions.<sup>46</sup> There are many reports in the literature describing the importance of phosphorylation in metabolic pathways and in cellular developments; however, only a few reports describe the role of this PTM in the reactivity of the allergens to IgE. Allergy to cow milk is relatively common in the first months of life. Cow milk potentially contains different allergens, such as  $\beta$ -lactoglobulin, serum albumin, and caseins.<sup>47</sup> The caseins attract special attention due to the importance of the hyperphosphorylation of its sequence, which plays an important role in the allergenicity of this protein.<sup>48</sup> Apparently, the phosphorylation of several sites in the sequence of caseins seems to reduce the affinity of specific-IgE by the epitope sequence in the caseins.<sup>49</sup> The effect of phosphorylation at Y111 of *P. paulista* antigen-5 is currently unknown, but the knowledge about the role of this PTM certainly will be investigated in the near future.

Glycosylation is one of the most common types of post-translational modifications. The intact isoform of *P. paulista* antigen-5 (spot 125) was previously reported as a glycoprotein;<sup>4</sup> however, neither the type nor the position of this PTM was identified. In the present investigation, a glycosylation site at K141 of this protein was observed. A careful examination of the spectral data suggest that the abundance ratio between glycosylated and the nonglycosylated forms was about 3:1. A theoretical prediction of glycosylation for the complete sequence of *P. paulista* antigen-5 was performed using the tool NetGlycate 1.0 server ([http://expasy.org/proteomics/post-translational\\_modification](http://expasy.org/proteomics/post-translational_modification)), used for virtual prediction of glycation at the epsilon amino groups of lysine residues,<sup>50</sup> which revealed a series of potential glycosylation positions such as K2, K7, K10, K45, K104, K131, K169, K184, and K207. Thus, apparently the position K141 in the sequence of *P. paulista* antigen-5 seems to constitute a novel glycation motif.

Glycosylation has been described in wasp venom allergens, such as in the hyaluronidase from the venom of the wasps *P. paulista*, *V. vulgaris*, *D. maculata*, *P. anularis*<sup>1,6,51</sup> and in the PLA<sub>1</sub> and PLA<sub>2</sub> from the *P. paulista* venom.<sup>4,5</sup> The specific role of the carbohydrates attached to the wasp venom proteins is not clear, but it has been proposed that they could constitute secondary immunogenic determinants for IgE and contribute to biological activity, solubility, stability, and resistance to proteases.<sup>51,52</sup> Glucosylation was previously reported in the pollen allergenic protein 1,3-glucanase, which is also known as Ole 9, a pathogenesis-related protein of the family Pr-2;<sup>53</sup> however, the role of the carbohydrate moiety in this protein is unknown.

Whether the carbohydrate epitopes are important for immunogenicity of the antigen-5 can be determined in future investigations that employ epitope mapping of this allergenic protein. The existence of different molecular isoforms for some venom proteins would benefit the wasp, permitting the insects to adapt the same original molecules within different cellular/molecular targets.<sup>4</sup> Each one of the PTMs described above could contribute to stability, resistance to proteases, signaling, and modulating the antigenicity of *P. paulista* antigen-5. Apparently, the PTMs which seem to be biologically relevant are the phosphorylation at Y111 and the glycosylation at K141, while the other ones discussed above seem to have occurred as artifacts of sample manipulation; thus, the sequence of *P. paulista* antigen-5 with the naturally occurring PTMs is represented in Figure 5.

Given the mass contribution of the PTMs to the molecular mass of the intact isoform of antigen-5, this value would be 23.448 Da, which compared to the value obtained in the 2-D SDS-PAGE protocol (26.479 Da), indicates that the mass difference between these values (3031 Da) is higher than that expected by the classical error of the electrophoretic techniques (approximately 5%). This suggests that other PTMs that were not detected in this study may exist in the intact isoform of *P. paulista* antigen-5.

## CONCLUDING REMARKS

The present investigation used a combination of 2-D SDS-PAGE, N- and C-terminal sequencing by degradation chemistry, in-gel proteolytic digestion with three different proteinases, followed by LC-ESI-IT-MS and MS<sup>2</sup> protocols, to obtain the proteomic characterization of different antigen-5 isoforms and to assign the complete sequence of the intact protein. Additionally, the 3-D structure was modeled and compared to antigen-5 from other social wasp species. The allergenicity of some venom proteins is most likely the most well-studied action of the stinging accidents caused by social wasps. Thus, the sequencing, PTM assignment, and structure determination of antigen-5 could contribute to improving our knowledge about one of the major wasp venom allergens. The identification and characterization of allergenic compounds is essential for the development of advanced component-resolved allergy diagnostics and treatment.

## ASSOCIATED CONTENT

### Supporting Information

(Figure S1) Sequence alignment between antigen-5 from the venom of *P. scutellaris rioplatensis* (VA5\_POLSR) and the N- and C-terminal sequences of the six isoforms of antigen-5 from the venom of *P. paulista* (spots 125, 134, 188, 189, 194, and 236). (Figure S2) Representative phylogenetic tree among the primary sequence of the antigen-5 from the social wasp *P. paulista* (VA5\_POLPI) with the primary sequence of the wasps *Polybia scutellaris rioplatensis* (VA5\_POLSCR), *Polistes gallicus* (VA5\_POLGA), *Polistes annularis* (VA5\_POLAN), *Polistes dominula* (VA5\_POLDO), *Dolichovespula maculata* (VA5\_DOLMA), *Vespula vulgaris* (VA5\_VESVU) (<http://www.phylogeny.fr>). (Figures S3 to S41) Representative mass spectra of antigen-5 (spot 125) from the venom of the social wasp *P. paulista* by in-gel protein digestion using trypsin, chymotrypsin, and Glu-C/V8 protease. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: mspalma@rc.unesp.br. Fax: 55-(19)-35348523. Tel.: 55-(19)-35264163.

### Author Contributions

All authors contributed to the writing of the manuscript.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by grants from the BIOprospecTA/FAPESP program (Proc. 2011/51684-1), CNPq, and CAPES. M.S.P. is a researcher from the National Research Council of

Brazil-CNPq. L.D.S. and H.A.A. were Postdoctoral fellows from FAPESP, and J.R.A.S.P. is a Ph.D. student fellow from FAPESP.

## REFERENCES

- (1) Skov, L. K.; Seppl, U.; Coen, J. J. F.; Crickmore, N.; King, T. P.; Monsalve, R. Structure of recombinant Ves v 2 at 2.0 Å resolution: structural analysis of an allergenic hyaluronidase from wasp venom. *Acta Crystallogr. D. Biol. Crystallogr.* **2006**, *62*, 595–604.
- (2) Hoffman, D. R. Hymenoptera venom allergens. *Clin. Rev. Allergy Immunol.* **2006**, *30*, 109–128.
- (3) Santos, L. D.; Santos, K. S.; Souza, B. M.; Arcuri, H. A.; Castro, F. M.; Kalil, E. J.; Palma, M. S. Purification, sequencing and structural characterization of the phospholipase A1 from the venom of the social wasp *Polybia paulista* (Hymenoptera, Vespidae). *Toxicon* **2007**, *50*, 923–937.
- (4) Santos, L. D.; Santos, K. S.; Pinto, J. R. A. S.; Dias, N. B.; Souza, B. M.; Santos, M. F.; Perales, J.; Domont, G. B.; Castro, F. M.; Kalil, E. J.; Palma, M. S. Profiling the proteome of the venom from the social wasp *Polybia paulista*: a clue to understand the envenoming mechanism. *J. Proteome Res.* **2010**, *9*, 3867–3877.
- (5) Santos, L. D.; Menegasso, A. R. S.; Santos-Pinto, J. R. A.; Santos, K. S.; Castro, F. M.; Kalil, E. J.; Palma, M. S. Proteomic characterization of the multiple forms of the PLAs from the venom of the social wasp *Polybia paulista*. *Proteomics* **2011**, *11*, 1403–1412.
- (6) Santos-Pinto, J. R. A.; Santos, L. D.; Arcuri, H. A.; Dias, N. B.; Palma, M. S. Proteomic characterization of the hyaluronidase (E.C. 3.2.1.35) from the venom of the social wasp *Polybia paulista*. *Protein Pept. Lett.* **2012**, *19*, 625–635.
- (7) Winningham, K. M.; Fitch, C. D.; Schmidt, M.; Hoffman, D. R. Hymenoptera venom protease allergens. *J. Allergy Clin. Immunol.* **2004**, *114*, 928–933.
- (8) Gibbs, G. M.; Roelants, K.; O'Bryan, M. K. The CAP superfamily: cysteine-rich secretory proteins, antigen-5 and pathogenesis-related 1 proteins—Roles in reproduction, cancer, and immune defense. *Endocrine Rev.* **2008**, *29*, 865–897.
- (9) Kovalick, G. E.; Schreiber, M. C.; Dickson, A. K.; Cunningham, R. A. Structure and expression of the antigen-5 related gene of *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.* **1998**, *28*, 491–500.
- (10) Mans, B. J.; Andersen, J. F.; Francischetti, I. M.; Valenzuela, J. G.; Schwan, T. G.; Ribeiro, J. M. Comparative sialomics between hard and soft ticks: implications for the evolution of blood-feeding behavior. *Insect Biochem. Mol. Biol.* **2008**, *38*, 42–58.
- (11) Charlab, R.; Valenzuela, J. G.; Rowton, E. D.; Ribeiro, J. M. Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis*. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 15155–15160.
- (12) Calvo, E.; Dao, A.; Pham, V. M.; Ribeiro, J. M. An insight into the sialome of *Anopheles funestus* reveals an emerging pattern in anopheline salivary protein families. *Insect Biochem. Mol. Biol.* **2007**, *37*, 164–175.
- (13) Van Vaerenbergh, M.; Cardoen, D.; Formesyn, E. M.; Brunain, M.; Van Driessche, G.; Blank, S.; Spillner, E.; Verleyen, P.; Wenseleers, T.; Schoofs, L.; Devreese, B.; De Graaf, D. C. Extending the honey bee venom with the antimicrobial peptide apidaecin and a protein resembling wasp antigen 5. *Insect Mol. Biol.* **2013**, *22*, 199–210.
- (14) Valenzuela, J. G.; Pham, V. M.; Garfield, M. K.; Francischetti, I. M.; Ribeiro, J. M. Toward a description of the sialome of the adult female mosquito *Aedes aegypti*. *Insect Biochem. Mol. Biol.* **2002**, *32*, 1101–1122.
- (15) Ribeiro, J. M.; Francischetti, I. M. Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives. *Annu. Rev. Entomol.* **2003**, *48*, 73–88.
- (16) King, T. P.; Sobotka, A. K.; Alagon, A.; Kochoumian, L.; Lichtenstein, L. M. Protein allergens 15 of white-faced hornet, yellow hornet, and yellow jacket venoms. *Biochemistry* **1978**, *17*, 5165–5174.
- (17) Hoffman, D. R. Allergens in Hymenoptera venom XIV: IgE binding activities of venom proteins from three species of vespids. *J. Allergy Clin. Immunol.* **1985**, *75*, 606–610.

- (18) Bohle, B.; Zwölfer, B.; Fischer, G. F.; Seppälä, U.; Ebner, C. Characterization of the human T cell response to antigen-5 from *Vespa vulgaris* (Ves v 5). *Clin. Exp. Allergy* **2005**, *35*, 367–373.
- (19) Lu, G.; Villalba, M.; Coscia, M. R.; Hoffman, D. R.; King, T. P. Sequence analysis and antigenic cross-reactivity of a venom allergen, antigen 5, from hornets, wasps, and yellow jackets. *J. Immunol.* **1993**, *150*, 2823–2830.
- (20) King, T. P.; Lu, G. Hornet venom allergen antigen 5, Dol m 5: its T-cell epitopes in mice and its antigenic cross-reactivity with a mammalian testis protein. *J. Allergy Clin. Immunol.* **1997**, *99*, 630–639.
- (21) Pirpignani, M.; Rivera, E.; Helman, U.; Biscoglio, M.; Bonino, J. Structural and immunological aspects of *Polybia scutellaris* antigen 5. *Arch. Biochem. Biophys.* **2002**, *407*, 224–230.
- (22) Pantera, B.; Hoffman, D. R.; Caresi, L.; Cappugi, G.; Turillazzi, S.; Manao, G.; Severino, M.; Spadolini, I.; Orsomando, G.; Moneti, G.; Pazzagli, L. Characterization of the major allergens purified from the venom of the paper wasp *Polistes gallicus*. *Biochim. Biophys. Acta* **2003**, *1623*, 72–81.
- (23) King, T. P.; Lu, G.; Gozalez, M.; Qian, N. F.; Soldatova, L. Yellow jacket venom allergens, hyaluronidase and phospholipase: sequence similarity and antigenic-cross reactivity with their hornet and wasp homologs and possible implications for clinical allergy. *J. Allergy Clin. Immunol.* **1996**, *98*, 588–600.
- (24) 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.
- (25) Henriksen, A.; King, T. P.; Mirza, O.; Monsalve, R. I.; Meno, K.; Ipsen, H.; Larsen, J. N.; Gajhede, M.; Spangfort, M. D. Major venom allergen of yellow jackets, Ves v 5: Structural characterization of a pathogenesis-related protein superfamily. *Proteins: Struct., Funct., Bioinf.* **2001**, *45*, 438–448.
- (26) Cascone, O.; Amaral, V.; Ferrara, P.; Vita, N.; Guillemot, J. C.; Diaz, L. E. Purification and characterization of two forms of antigen-5 from the *Polybia scutellaris* venom. *Toxicon* **1995**, *38*, 1367–1379.
- (27) Zielinska, D. F.; Gnad, F.; Jedrusik-Bode, M.; Wisniewski, J. R.; Mann, M. *Caenorhabditis elegans* has a phosphoproteome atypical for metazoans that is enriched in developmental and sex determination proteins. *J. Proteome Res.* **2009**, *8*, 4039–4049.
- (28) Thaysen-Andersen, M.; Thøgersen, I. B.; Nielsen, H. J.; Lademann, U.; Brunner, N.; Enghild, J. J.; Hojrup, P. Rapid and individual glycoprofiling of the low abundance N-glycosylated tissue inhibitor of metalloproteinases-1. *Mol. Cell. Proteomics* **2007**, *6*, 638–647.
- (29) Cuff, J. A.; Barton, G. J. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins: Struct., Funct., Bioinf.* **2000**, *40*, 502–511.
- (30) Clamp, M.; Cuff, J.; Searle, S. M.; Barton, G. J. The Jalview Java Alignment Editor. *Bioinformatics* **2004**, *20*, 426–427.
- (31) Delano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: Palo Alto, CA, 2002.
- (32) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK—A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
- (33) Luthy, R.; Bowie, J.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **1992**, *356*, 83–85.
- (34) Schwieters, C. D. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **2003**, *160*, 66–74.
- (35) Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J. F.; Guindon, S.; Lefort, V.; Lescot, M.; Claverie, J. M.; Gascuel, O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **2008**, *1*, 36 (Web Server issue):W465–W469.
- (36) Richards, O. W.; Richards, M. S. Observations on the social wasps of South America (Hymenoptera Vespidae). *Trans. Royal Entomol. Soc. London* **1951**, *102*, 1–69.
- (37) Gelin, L. F. F. *Análise filogenética de Polybia lepeletier, 1836 (Hymenoptera, Vespidae, Polistinae)*; Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas: São José do Rio Preto, 2009; p 63.
- (38) De Graaf, D. C.; Aerts, M.; Danneels, E.; Devreese, B. Bee, wasp and ant venomics pave the way for a component-resolved diagnosis of sting allergy. *J. Proteomics* **2009**, *72*, 145–154.
- (39) Hoffman, D. R.; Jacobson, R. S. Allergens in Hymenoptera venom XII: how much protein is in a sting? *Ann. Allergy* **1984**, *52*, 276–278.
- (40) King, T. P.; Alagon, A. C.; Kuan, J.; Sobotka, A. K.; Lichtenstein, L. M. Immunochemical studies of yellow jacket venom proteins. *Mol. Immunol.* **1983**, *20*, 297–308.
- (41) Milne, T. J.; Abbenante, G.; Tyndall, J. D.; Halliday, J.; Lewis, R. J. Isolation and Characterization of a Cone Snail Protease with Homology to CRISP Proteins of the Pathogenesis-related Protein Superfamily. *J. Biol. Chem.* **2003**, *278*, 31105–31110.
- (42) Petersen, A.; Schramm, G.; Schlaak, M.; Becker, W.-M. Post-translational modifications influence IgE reactivity to the major allergen Phl p 1 of timothy grass pollen. *Clin. Exp. Allergy* **1998**, *28*, 315–321.
- (43) Fenaille, F.; Nony, E.; Chabre, H.; Lautrette, A.; Couret, M.; Batard, T.; Moingeon, P.; Ezan, E. Mass spectrometric investigation of molecular variability of grass pollen group 1 allergens. *J. Proteome Res.* **2009**, *8*, 4014–4027.
- (44) Li, J.; Shefcheck, K.; Callahan, J.; Fenselau, C. Primary sequence and site-selective hydroxylation of prolines in isoforms of a major peanut allergen protein Ara h 2. *Protein Sci.* **2010**, *19*, 174–182.
- (45) Resende, V. M. F.; Vasilj, A.; Santos, K. S.; Palma, M. S.; Shevchenko, A. Proteome and phosphoproteome of Africanized and European honeybee venoms. *Proteomics* **2013**, *13*, 2638–2648.
- (46) Eyers, C. E.; Gaskell, S. J. Mass spectrometry to identify post-translational modifications. *Wiley Encycl. Chem. Biol.* **2008**, DOI: 10.1002/9780470048672.webc469.
- (47) Poth, A. G.; Deeth, H. C.; Alewood, P. F.; Holland, J. W. Analysis of the human casein phosphoproteome by 2-D electrophoresis and MALDI-TOF/TOF MS reveals new phosphoforms. *J. Proteome Res.* **2008**, *7*, 5017–5027.
- (48) Bernard, H.; Meisel, H.; Creminon, C.; Wal, J. M. Post-translational phosphorylation affects the IgE binding capacity of caseins. *FEBS Lett.* **2000**, *467*, 239–244.
- (49) Cases, B.; Boyano, M. T.; Pérez-Gordo, M.; Pedrosa, M.; Vivanco, F.; Quirce, S.; Pastor-Vargas, C. Phosphorylation reduces the allergenicity of cow casein in children with selective allergy to goat and sheep milk. *J. Investig. Allergol. Clin. Immunol.* **2011**, *21*, 398–400.
- (50) Johansen, M. B.; Kierner, L.; Brunak, S. Analysis and prediction of mammalian protein glycation. *Glycobiology* **2006**, *16*, 844–853.
- (51) 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 *Vespa vulgaris*. *FEBS J.* **2005**, *272*, 5182–5190.
- (52) Altmann, F. The role of protein glycosylation in allergy. *Int. Arch. Allergy Immunol.* **2007**, *142*, 99–115.
- (53) Huecas, S.; Villalba, M.; Rodriguez, R. Ole e 9, a major olive pollen allergen is a 1,3-beta-glucanase-Isolation, characterization, amino acid sequence, and tissue specificity. *J. Biol. Chem.* **2001**, *276*, 27959–27966.

## Supporting information

### Using proteomic strategies for sequencing and post-translational modifications assignment of antigen-5, a major allergen from the venom of the social wasp *Polybia paulista*

José Roberto Aparecido dos Santos-Pinto,<sup>1,3</sup> Lucilene Delazari dos Santos,<sup>2,3</sup>

Helen Andrade Arcuri,<sup>3,4</sup> Fábio Morato Castro,<sup>3,4</sup> Jorge Elias Kalil<sup>3,4</sup> and Mario Sergio Palma<sup>1,3 \*</sup>

<sup>1</sup>Institute of Biosciences of Rio Claro, Department of Biology, Center of the Study of Social Insects, University of São Paulo State (UNESP), Rio Claro, SP, Brazil; <sup>2</sup>Center for the Study of Venoms and Venomous Animals (CEVAP), University of São Paulo State (UNESP), Botucatu, SP; Brazil; <sup>3</sup>INCT-iii, São Paulo, Brazil; <sup>4</sup>Discipline of Allergy and Immunology (HC/Incor/FMUSP), SP, Brazil.

\* Correspondence: Prof. Dr. Mario Sergio Palma. CEIS-IBRC- UNESP, Av. 24A nº 1515, Bela Vista - Rio Claro, SP, Brazil, CEP 13506-900, Phone: 55-(19)-35264163; FAX: 55-(19)-35348523. E-mail: [mspalma@rc.unesp.br](mailto:mspalma@rc.unesp.br)

|           |                       |                   |            |           |          |           |
|-----------|-----------------------|-------------------|------------|-----------|----------|-----------|
|           | 1                     | 10                | 20         | 30        | 40       | 50        |
| VA5_POLSR | NKYCNIKC              | SK VAHTVCQYGE     | STKPSSKN   | CN KVSITS | VGVT E E | EKKLI VDE |
| Spot 125  | NKYCNIKC              | SK VAHT           | .....      |           |          |           |
| Spot 134  | NKYCNIKC              | SK VAHT           | .....      |           |          |           |
| Spot 188  | NKYCNIKC              | SK VAHT           | .....      |           |          |           |
| Spot 189  | NKYCNIKC              | SK VAHT           | .....      |           |          |           |
| Spot 194  | NKYCNIKC              | SK VAHT           | .....      |           |          |           |
| Spot 236  | .....                 |                   |            |           |          |           |
|           | 60                    | 70                | 80         | 90        | 100      |           |
| VA5_POLSR | HNRFRQKVAQGLETRGNP    | GPQPAASDMNNLVW    | NDELAYIAQV | WANQCQFF  |          |           |
| Spot 125  | .....                 |                   |            |           |          |           |
| Spot 134  | .....                 |                   |            |           |          |           |
| Spot 188  | .....                 |                   |            |           |          |           |
| Spot 189  | .....                 |                   |            |           |          |           |
| Spot 194  | .....                 |                   |            |           |          |           |
| Spot 236  | .....                 |                   |            |           |          | ASQCQFF   |
|           | 110                   | 120               | 130        | 140       | 150      |           |
| VA5_POLSR | VHDKCRNTAQ YQVGQNIAYS | ASTAAYPGIV        | SLIVLWENEV | K D F     | NYSQGIT  |           |
| Spot 125  | .....                 |                   |            |           |          |           |
| Spot 134  | .....                 |                   |            |           |          |           |
| Spot 188  | .....                 |                   |            |           |          |           |
| Spot 189  | .....                 |                   |            |           |          |           |
| Spot 194  | .....                 |                   |            |           |          |           |
| Spot 236  | VHDKCRN               | .....             |            |           |          |           |
|           | 160                   | 170               | 180        | 190       | 200      |           |
| VA5_POLSR | KENFAKVGHY TQVWWAKTKE | VGCGSIKYIEKGMKSHY | LVCNYGPAGN | YMG       |          |           |
| Spot 125  | .....                 |                   |            |           | G        |           |
| Spot 134  | .....                 |                   |            |           | GPAGNVL  |           |
| Spot 188  | .. NFAKVGH            | .....             |            |           |          |           |
| Spot 189  | .... FAKVGHY T        | .....             |            |           |          |           |
| Spot 194  | .....                 |                   |            |           | IKYIEKG  |           |
| Spot 236  | .....                 |                   |            |           | G        |           |
|           | 207                   |                   |            |           |          |           |
| VA5_POLSR | QPIYTKK               |                   |            |           |          |           |
| Spot 125  | AQIYEIK               |                   |            |           |          |           |
| Spot 134  | .....                 |                   |            |           |          |           |
| Spot 188  | .....                 |                   |            |           |          |           |
| Spot 189  | .....                 |                   |            |           |          |           |
| Spot 194  | .....                 |                   |            |           |          |           |
| Spot 236  | AQIYEIK               |                   |            |           |          |           |

**Figure S1-** Alignment sequences between the antigen-5 from the venom of *P. scutellaris rioplatensis* (VA5\_POLSR) and the N- and C-terminal sequences of the six forms of antigen-5 from the venom of *P. paulista* (spots 125, 134, 188, 189, 194, and 236).

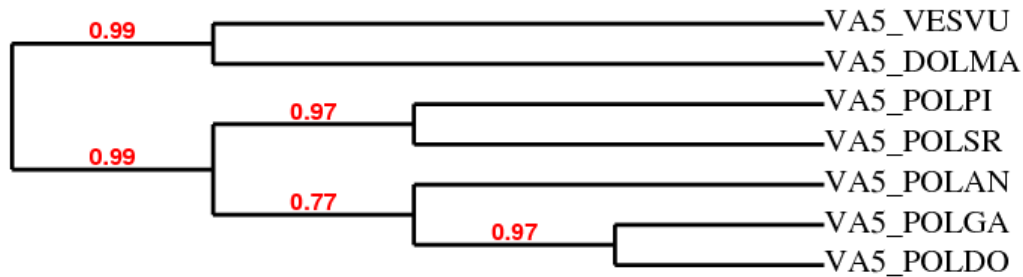


**Figure S2-** Representative phylogenetic tree based on alignment of the primary sequences of the antigen-5 from the venoms of social wasps *P. paulista* (VA5\_POLPI), *Polybia scutellaris rioplatensis* (VA5\_POLSCR), *Polistes gallicus* (VA5\_POLGA), *Polistes annularis* (VA5\_POLAN), *Polistes dominula* (VA5\_POLDO), *Dolichovespula maculata* (VA5\_DOLMA), *Vespula vulgaris* (VA5\_VESVU). (<http://www.phylogeny.fr>).<sup>35</sup>

**Representative mass spectra of antigen-5 (spot 125) from the venom of the social wasp *P. paulista* by *in-gel* protein digestion using trypsin.**

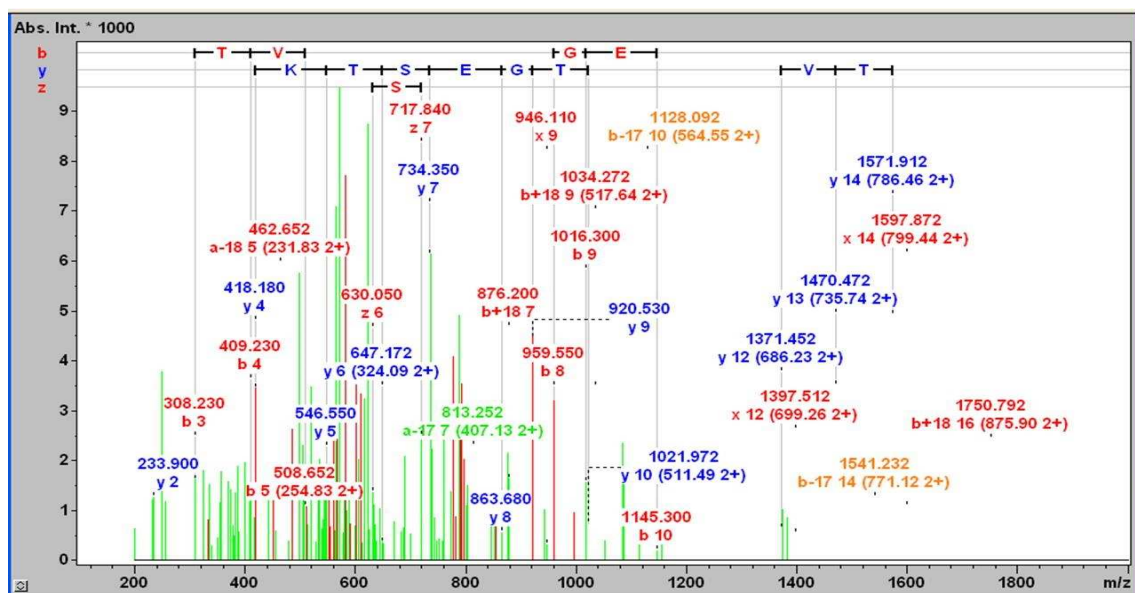


**Figure S3.** Representative CID MS/MS spectrum of tryptic peptide K.VAHTVCQTGESTKPSSK.N (11-27), selecting the  $m/z$  626.970  $[M + 3H]^{3+}$  as precursor ion.

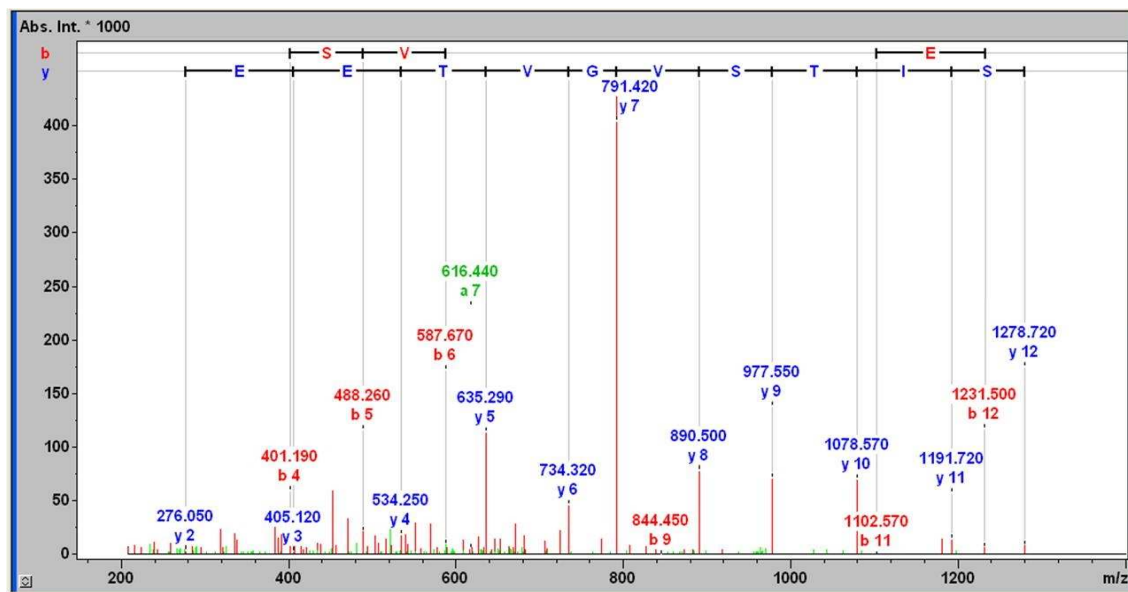


**Figure S4.** Representative CID MS/MS spectrum of tryptic peptide K.VSITSVGVTEEEK.K (32-44), selecting the m/z 689.350  $[M + 2H]^{2+}$  as precursor ion.

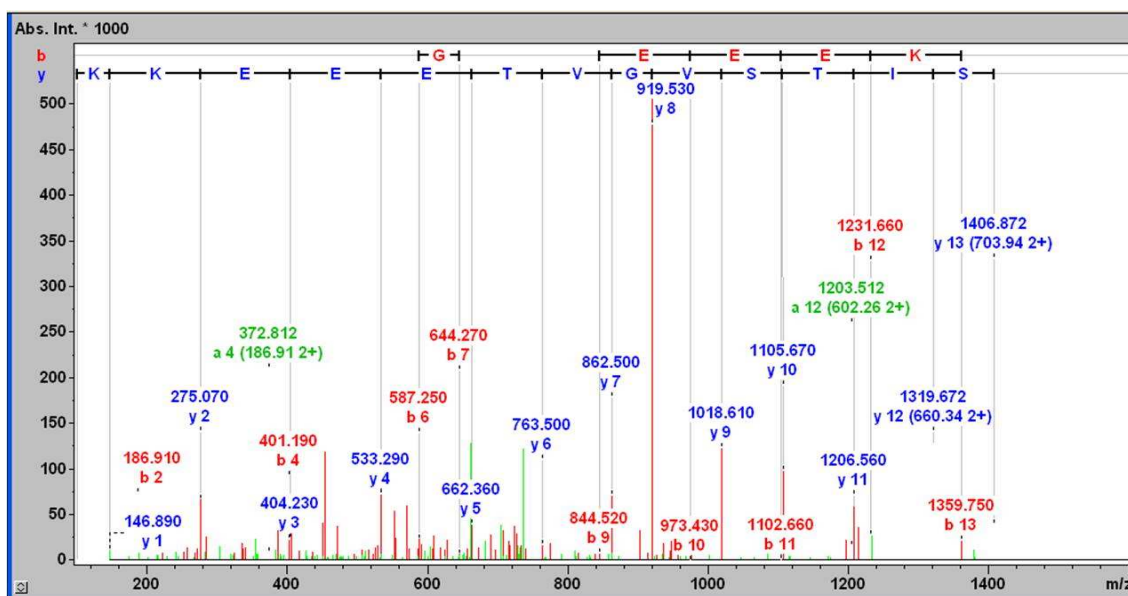


**Figure S5.** Representative CID MS/MS spectrum of tryptic peptide K.VSITSVGVTEEEKK.L (32-45), selecting the m/z 753.449  $[M + 2H]^{2+}$  as precursor ion.

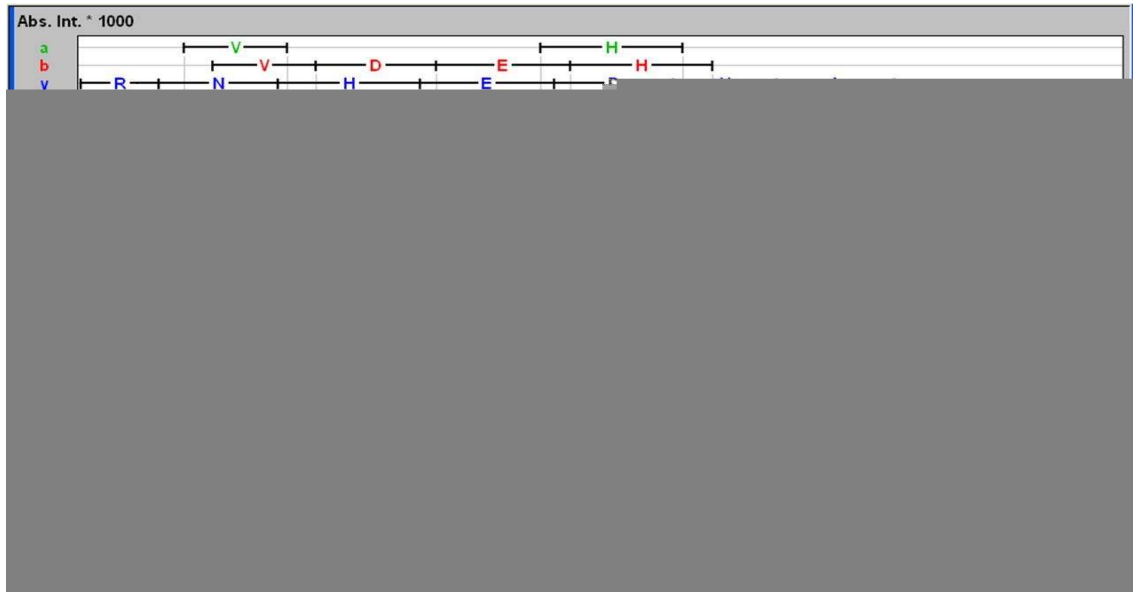


**Figure S6.** Representative CID MS/MS spectrum of tryptic peptide K.LIVDEHNR.F (46-53), selecting the  $m/z$  498.250  $[M + 2H]^{2+}$  as precursor ion.

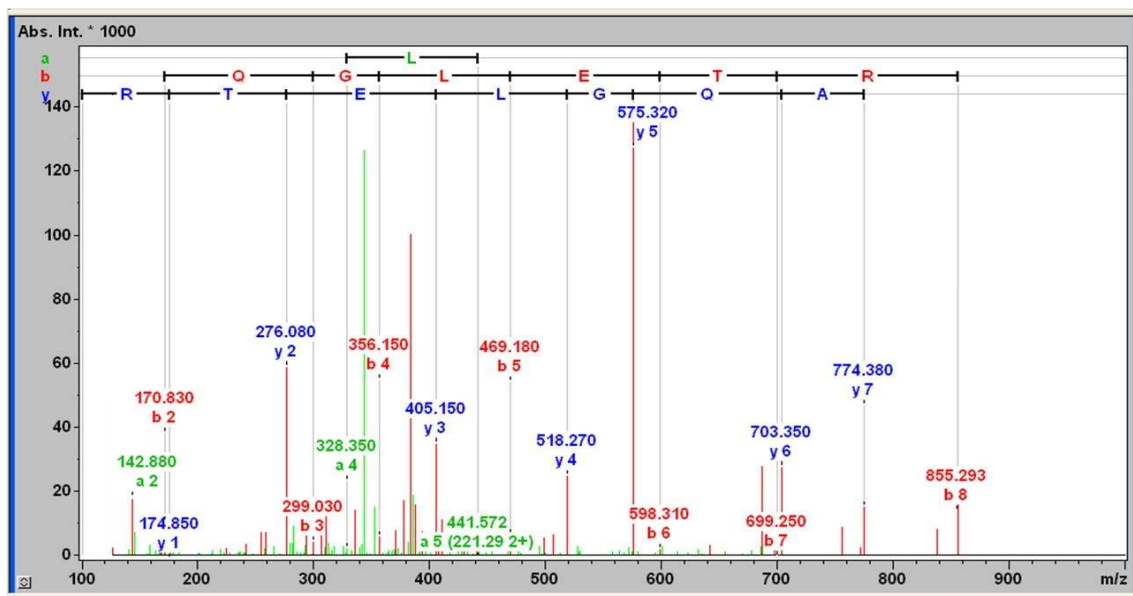


**Figure S7.** Representative CID MS/MS spectrum of tryptic peptide K.VAQGLETR.G (58-65), selecting the  $m/z$  437.231  $[M + 2H]^{2+}$  as precursor ion.

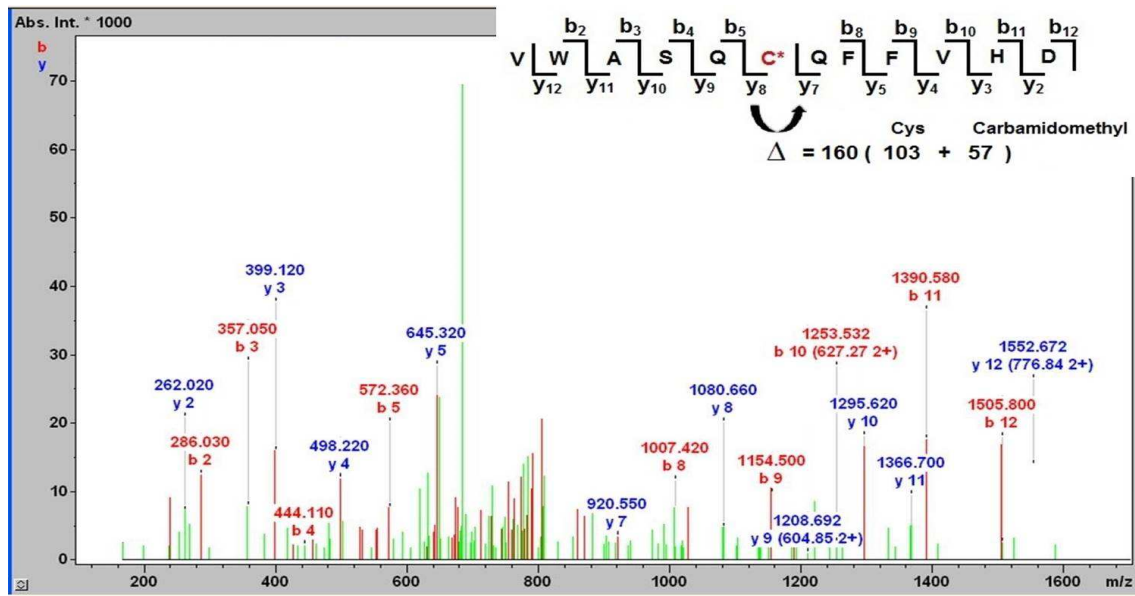


**Figure S8.** Representative CID MS/MS spectrum of tryptic peptide Q.VWASQ\*QFFVHDKCR.N (92-106), selecting the m/z 826.450 [M + 2H]<sup>2+</sup> as precursor ion; and showing the C97 carbamidomethyl site observed on the antigen-5 protein.

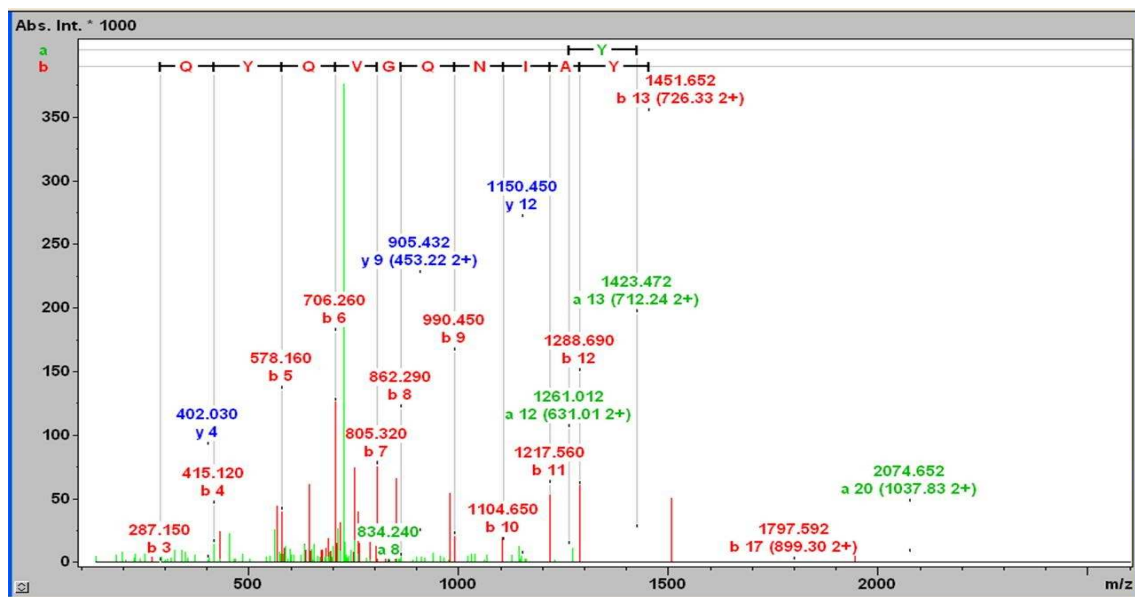


**Figure S9.** Representative CID MS/MS spectrum of tryptic peptide R.NTAQYQVGQNIAYSASTAAYPGVVK.L (107-131), selecting the m/z 778.800 [M + 4H]<sup>4+</sup> as precursor ion.

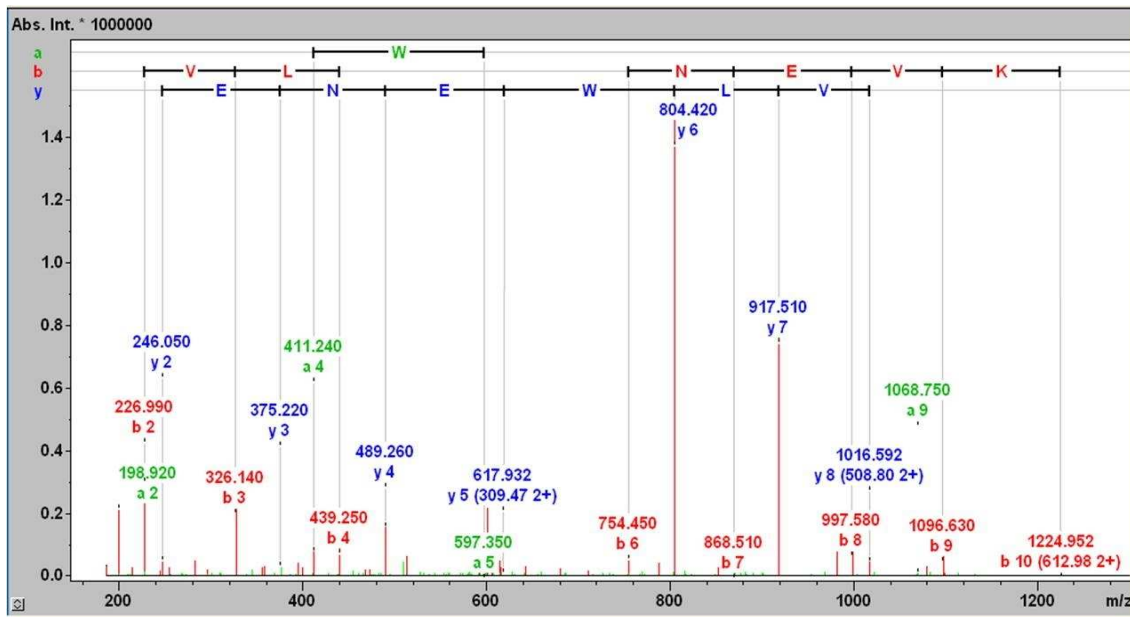


**Figure S10.** Representative CID MS/MS spectrum of tryptic peptide K.LIVLWENEVK.D (132-141), selecting the  $m/z$  621.843  $[M + 2H]^{2+}$  as precursor ion.

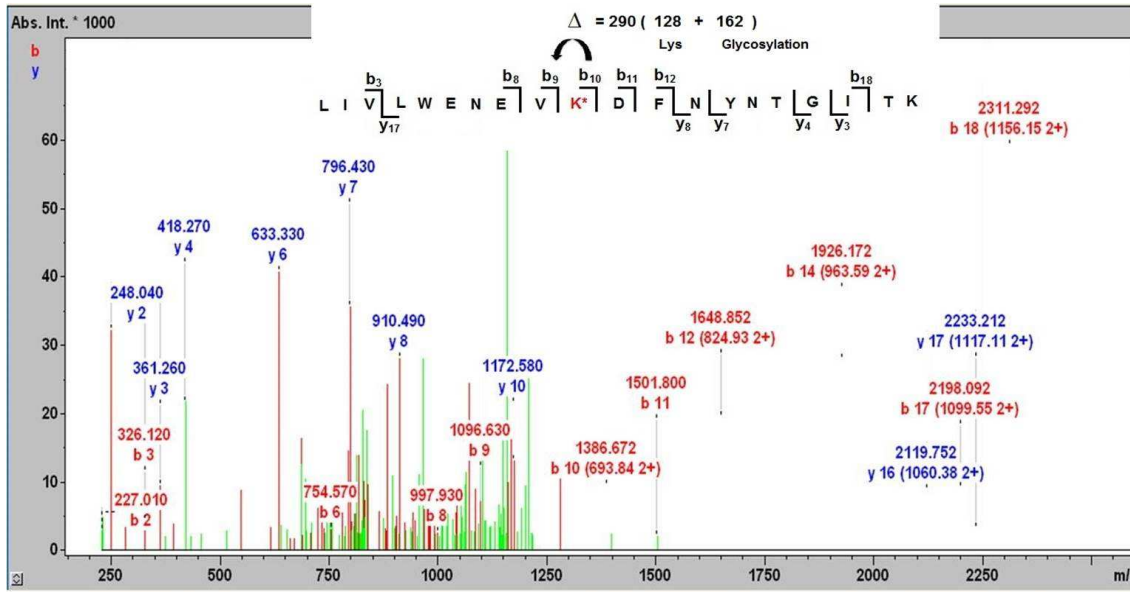


**Figure S11.** Representative CID MS/MS spectrum of tryptic peptide K.LIVLWENEVK\*DFNYNTGITK.E (132-151), selecting the  $m/z$  853.472  $[M + 3H]^{3+}$  as precursor ion; and showing the K141 glycosylation site observed on the antigen-5 protein.

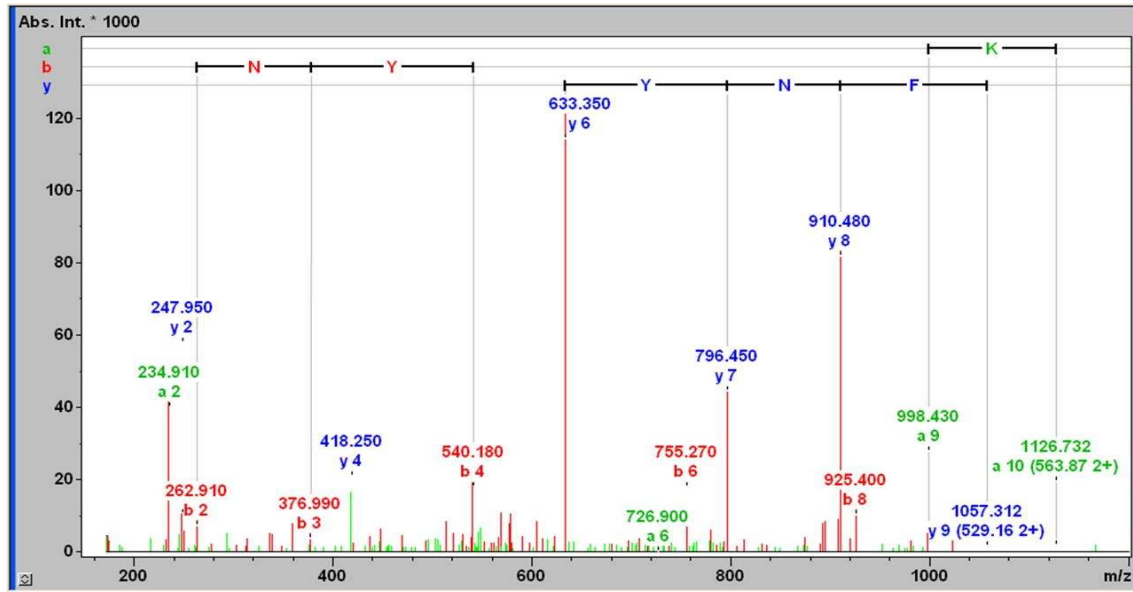


**Figure S12.** Representative CID MS/MS spectrum of tryptic peptide K.DFNNTGITE (142-151), selecting the m/z 586.786 [M + 2H]<sup>2+</sup> as precursor ion.

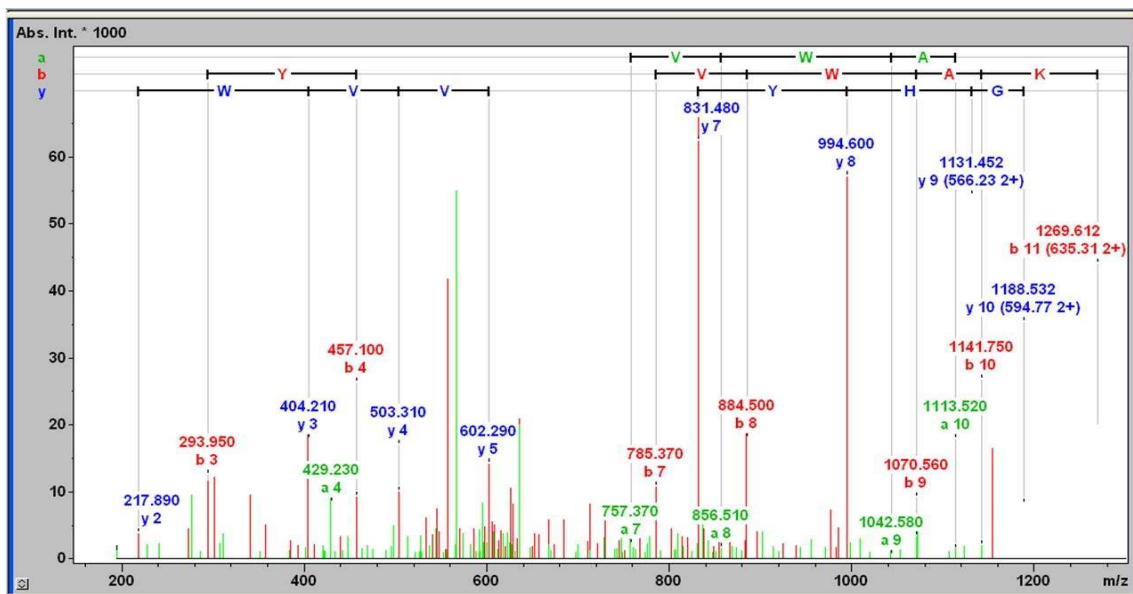


**Figure S13.** Representative CID MS/MS spectrum of tryptic peptide K.VGHYTVVWAK.T (157-167), selecting the m/z 644.346 [M + 2H]<sup>2+</sup> as precursor ion.

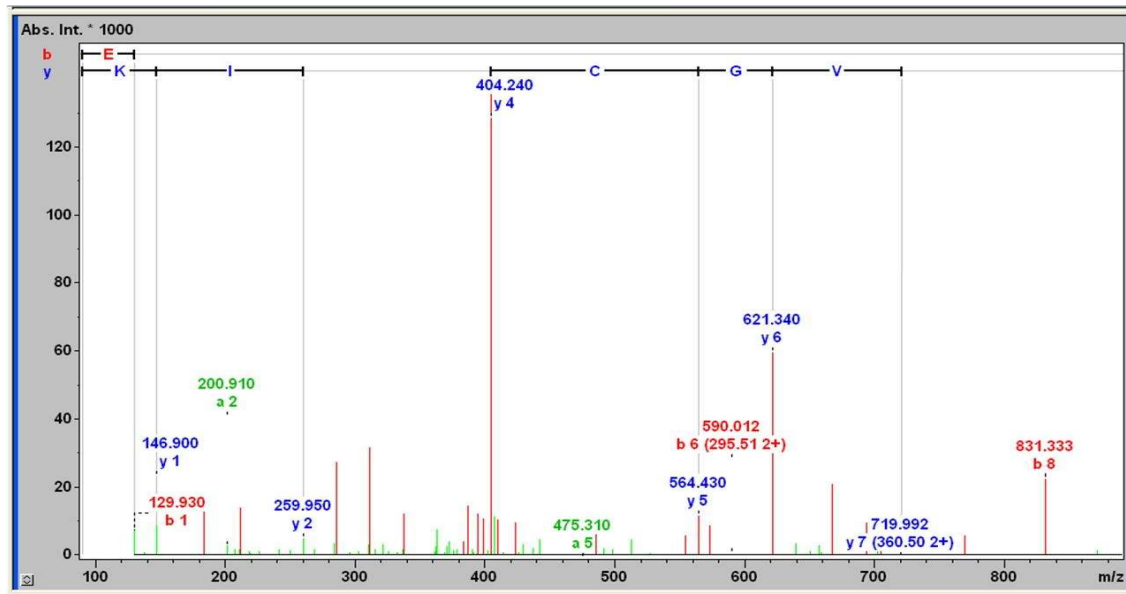


**Figure S14.** Representative CID MS/MS spectrum of tryptic peptide K.EVGCGSIK.Y (170-177), selecting the m/z 425.198 [M + 2H]<sup>2+</sup> as precursor ion.

Representative mass spectra of antigen-5 (spot 125) from the venom of the social wasp *P. paulista* by *in-gel* protein digestion using chymotrypsin.

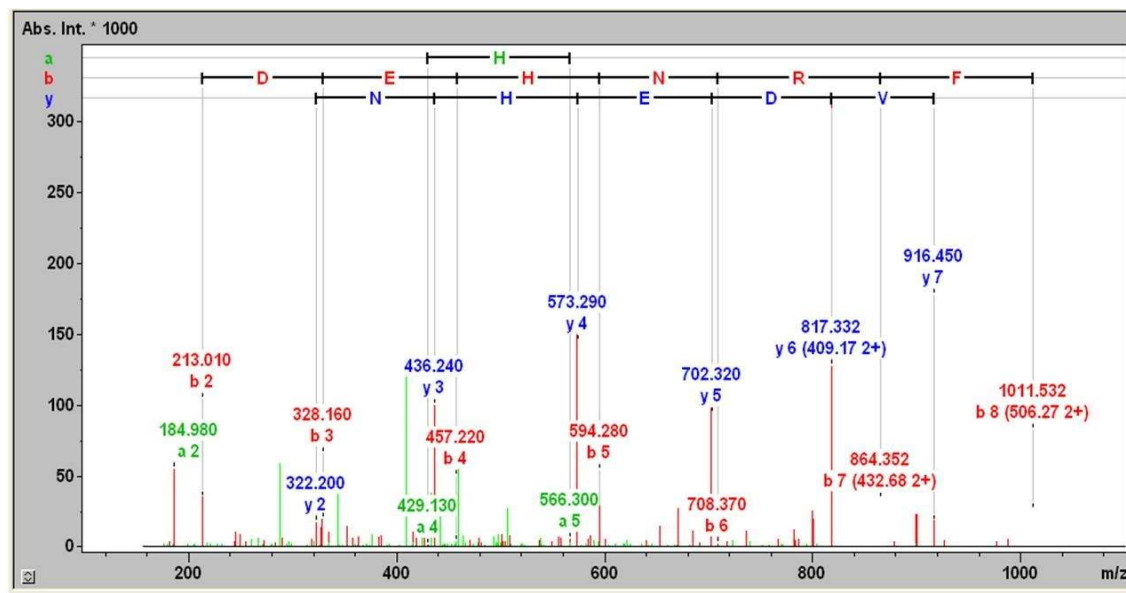


**Figure S15.** Representative CID MS/MS spectrum of chymotryptic peptide L.IVDEHNRF.R (47-57), selecting the m/z 515.254 [M + 2H]<sup>2+</sup> as precursor ion.

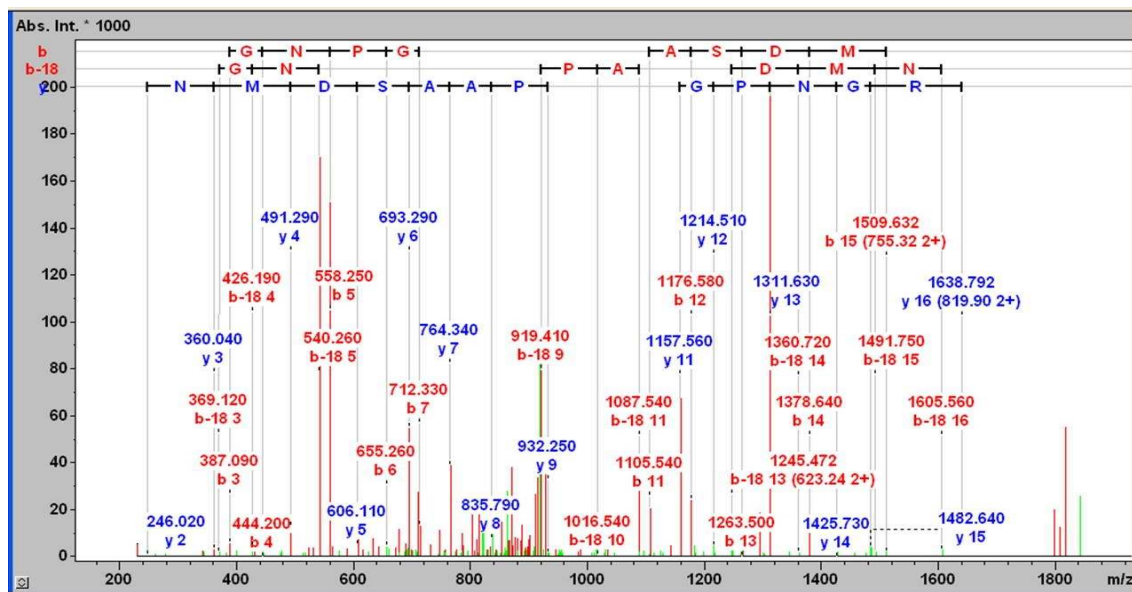


**Figure S16.** Representative CID MS/MS spectrum of chymotryptic peptide L.ETRGNPGPQPAASDMNVL.V (63-80), selecting the  $m/z$  934.820  $[M + 2H]^{2+}$  as precursor ion.

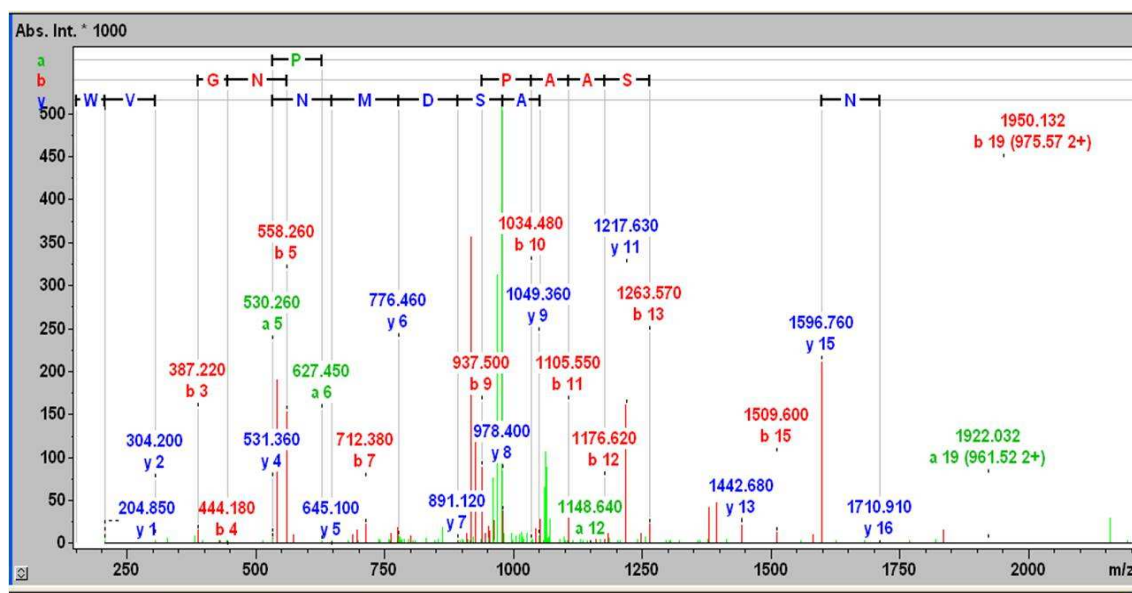


**Figure S17.** Representative CID MS/MS spectrum of chymotryptic peptide L.ETRGNPGPQPAASDMNVLVW.N (63-82), selecting the  $m/z$  1077.433  $[M + 2H]^{2+}$  as precursor ion.

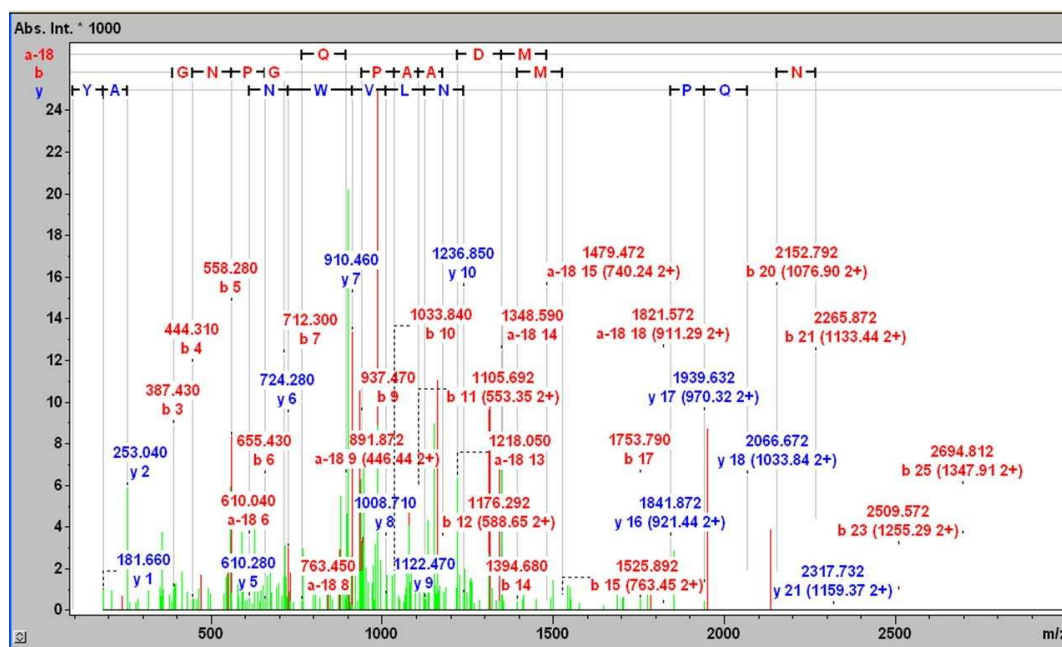


**Figure S18.** Representative CID MS/MS spectrum of chymotryptic peptide L.ETRGNPGPQPAASDMNNLVWNEDELAY.I (63-88), selecting the  $m/z$  959.090  $[M + 3H]^{3+}$  as precursor ion; and showing the D76 hydroxylation site observed on the antigen-5 protein.

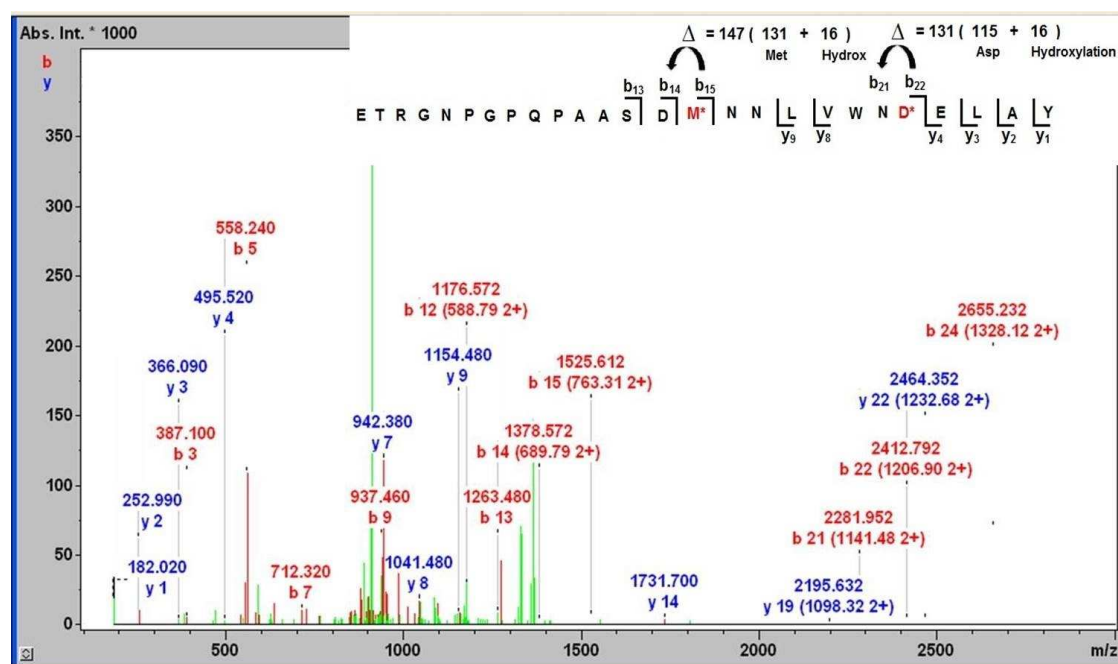


**Figure S19.** Representative CID MS/MS spectrum of chymotryptic peptide L.ETRGNPGPQPAASDM\*NNLVWND\*ELAY.I (63-88), selecting the  $m/z$  969.670  $[M + 3H]^{3+}$  as precursor ion; and showing the M77 oxidation site, and D84 hydroxylation site observed on the antigen-5 protein.

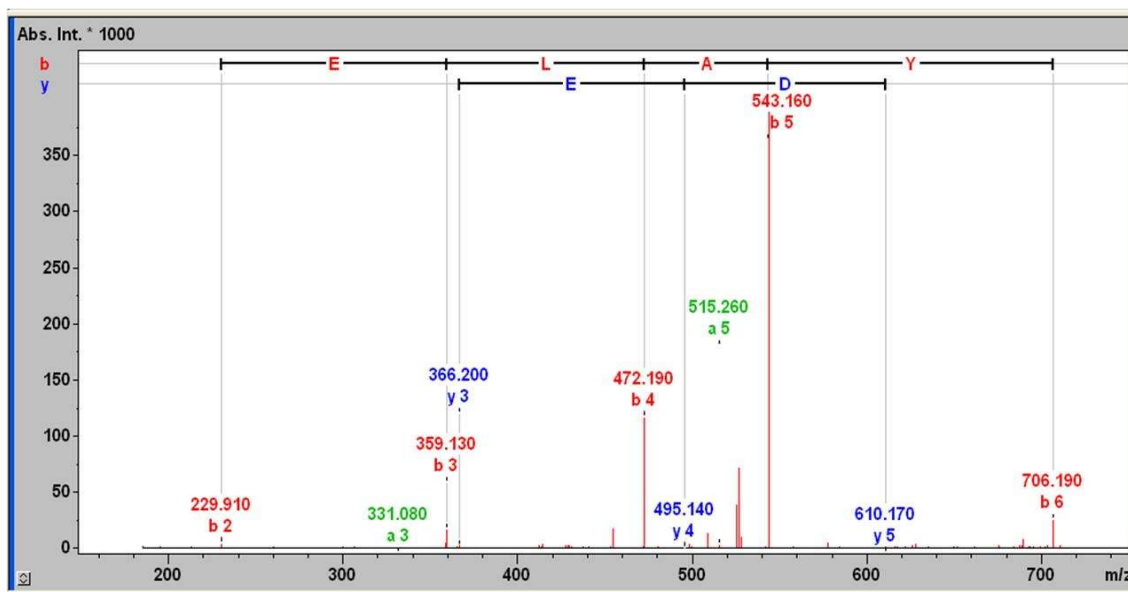


**Figure S20.** Representative CID MS/MS spectrum of chymotryptic peptide W.NDELAY.I (82-87), selecting the m/z 724.270  $[M + H]^+$  as precursor ion.

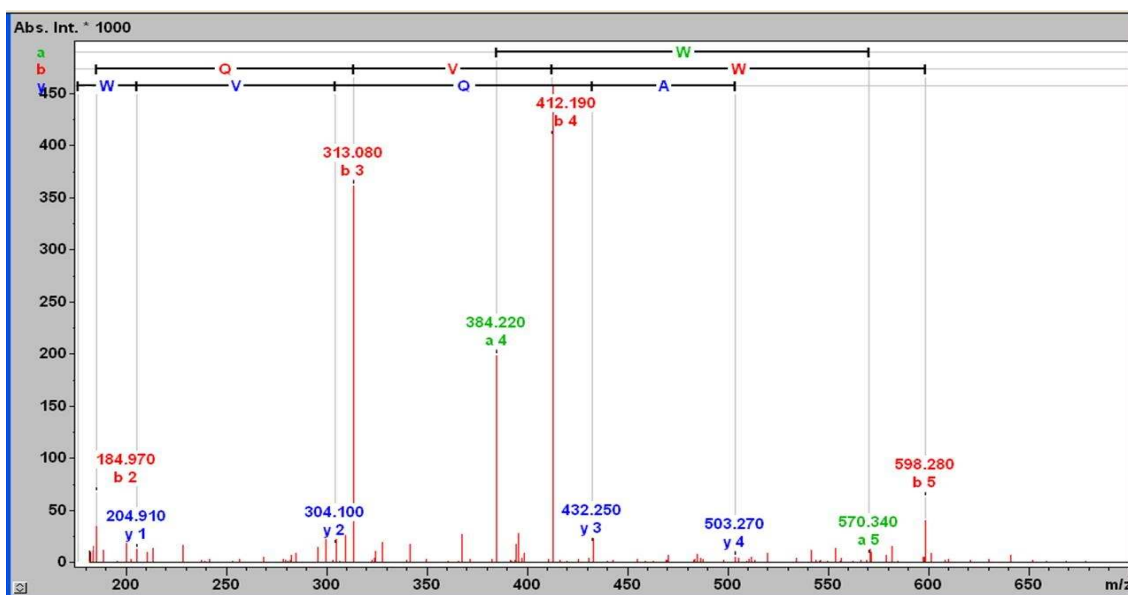


**Figure S21.** Representative CID MS/MS spectrum of chymotryptic peptide Y.IAQVW.A (89-93), selecting the m/z 616.350  $[M + H]^+$  as precursor ion.

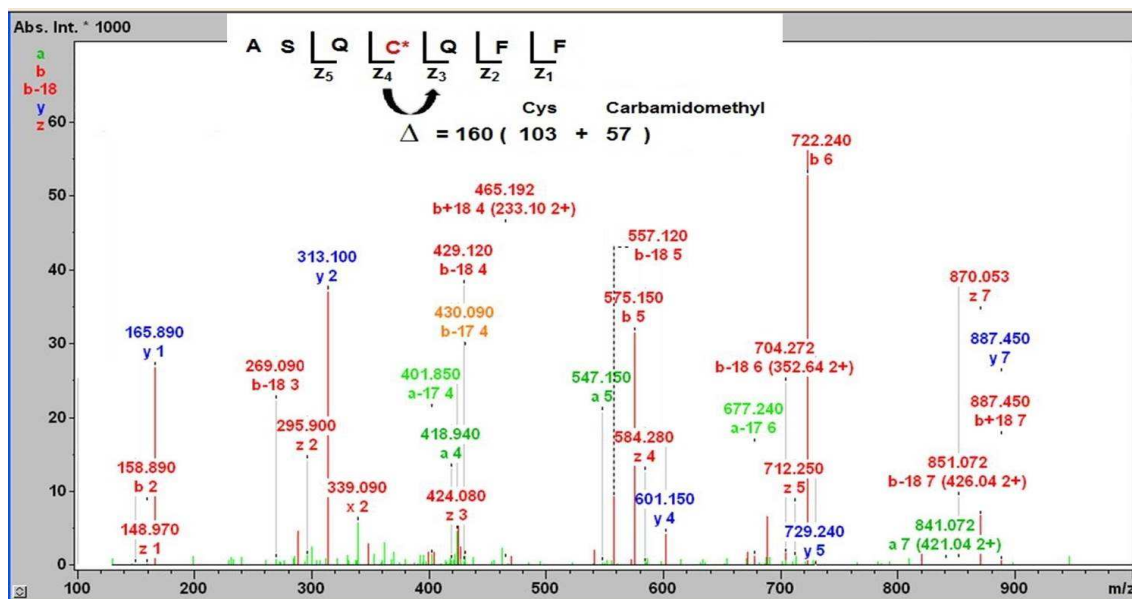


**Figure S22.** Representative CID MS/MS spectrum of chymotryptic peptide W.ASQC\*QFF.V (94-100), selecting the m/z 444.180 [M + 2H]<sup>2+</sup> as precursor ion; and showing the C97 carbamidomethyl site observed on the antigen-5 protein

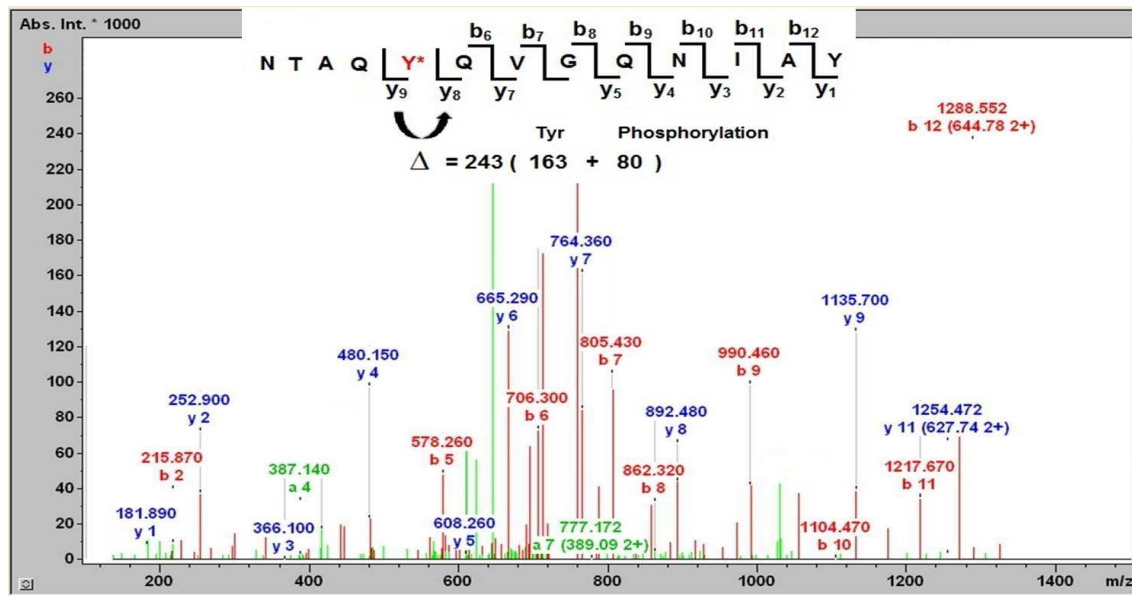


**Figure S23.** Representative CID MS/MS spectrum of chymotryptic peptide R.NTAQY\*QVGQNIAY.S (107-119), selecting the m/z 775.3702 [M + 2H]<sup>2+</sup> as precursor ion; and showing the Y111 phosphorylation site observed on the antigen-5 protein.

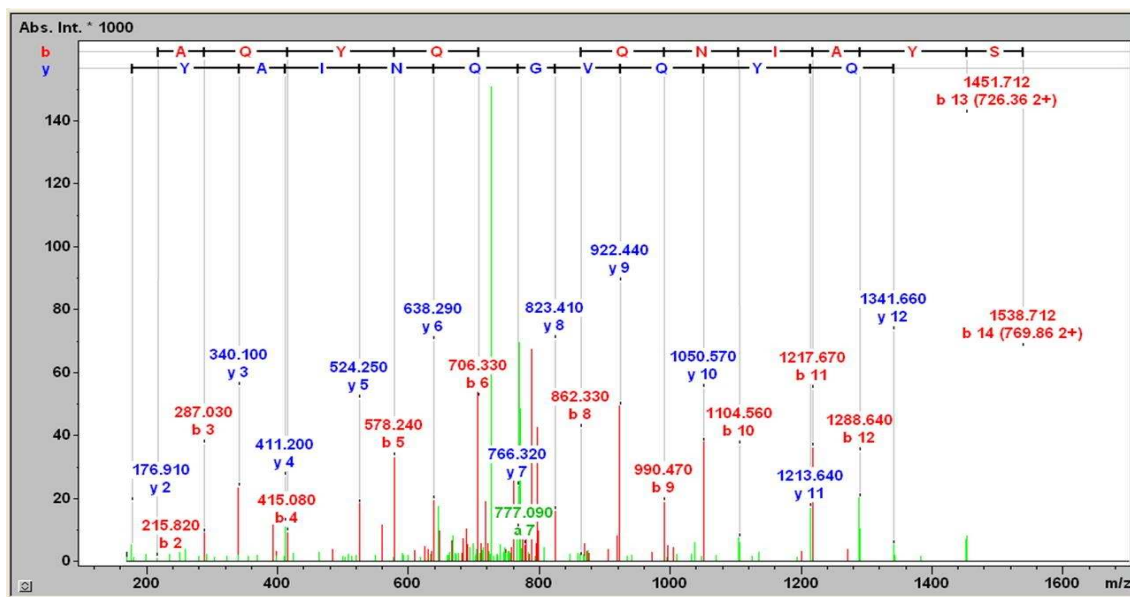


**Figure S24.** Representative CID MS/MS spectrum of chymotryptic peptide R.NTAQYQVGQNIAYSA.S (107-121), selecting the  $m/z$  814.381  $[M + 2H]^{2+}$  as precursor ion.

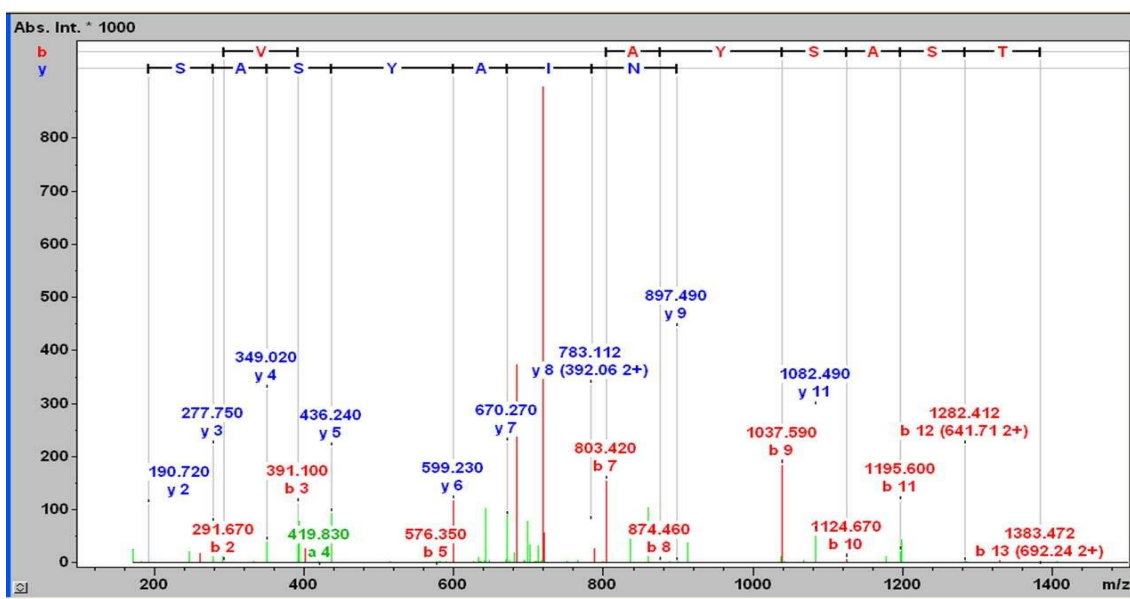


**Figure S25.** Representative CID MS/MS spectrum of chymotryptic peptide Q.YQVGQNIAYSASTA.A (111-124), selecting the  $m/z$  736.840  $[M + 2H]^{2+}$  as precursor ion.

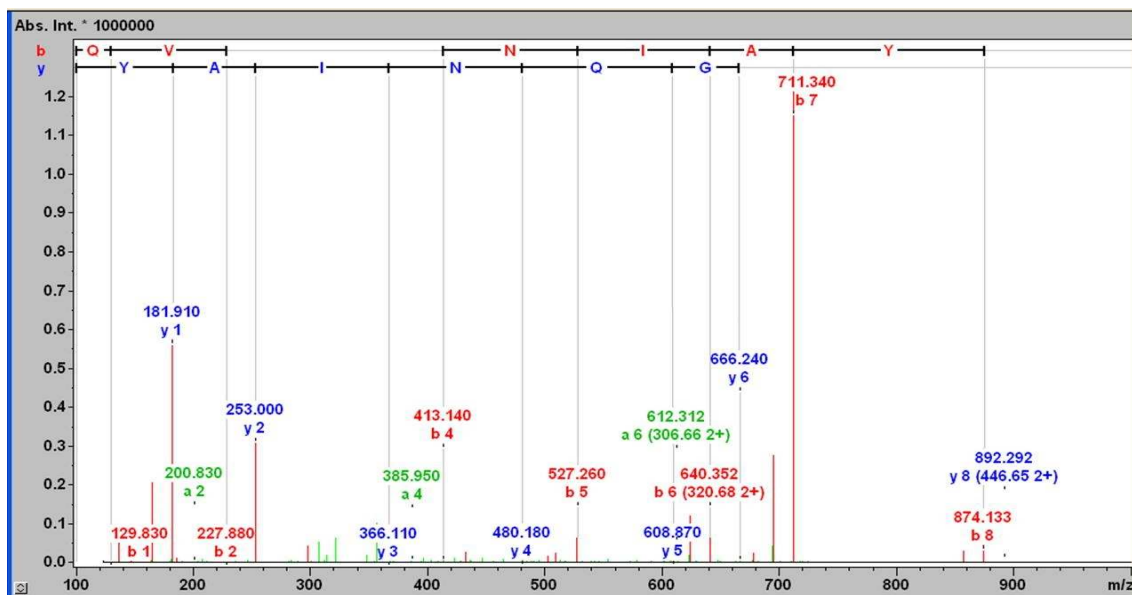


**Figure S26.** Representative CID MS/MS spectrum of chymotryptic peptide Y.QVGQNIAY.S (112-119), selecting the m/z 446.701  $[M + 2H]^{2+}$  as precursor ion.

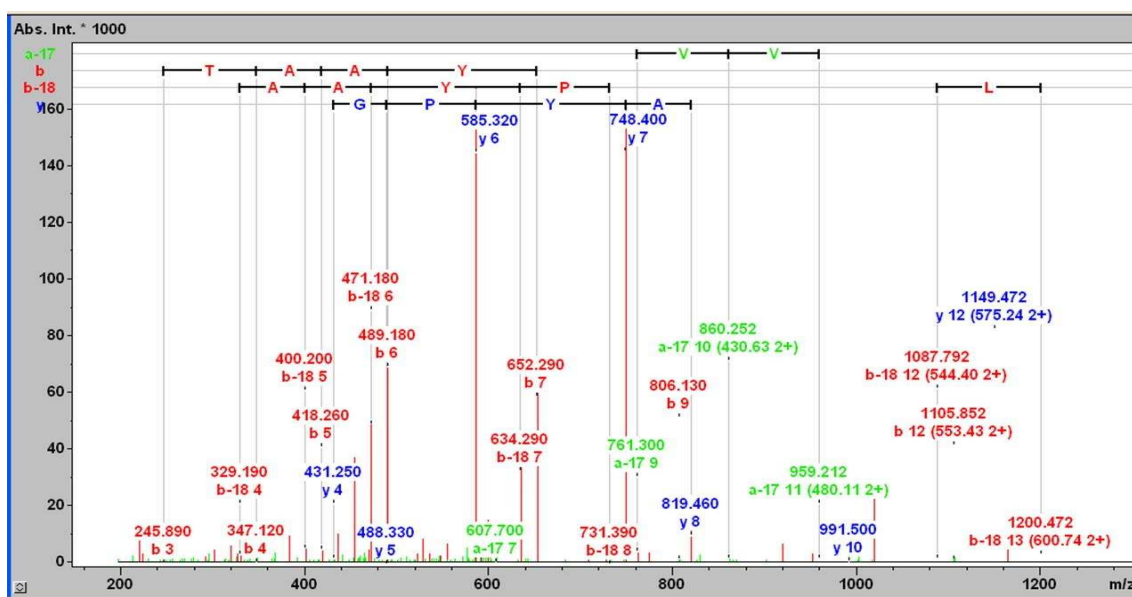


**Figure S27.** Representative CID MS/MS spectrum of chymotryptic peptide Y.SASTAAYPGVVKL.I (120-132), selecting the m/z 618.780  $[M + 2H]^{2+}$  as precursor ion.

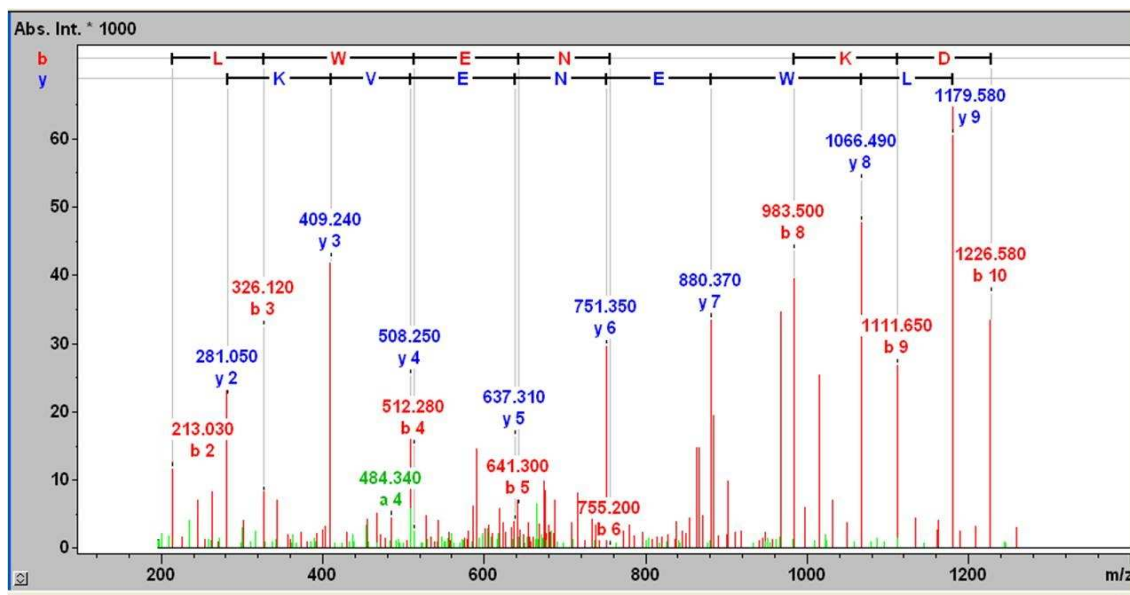


**Figure S28.** Representative CID MS/MS spectrum of chymotryptic peptide L.IVLWENEVKDF.N (133-143), selecting the  $m/z$  696.371  $[M + 2H]^{2+}$  as precursor ion.

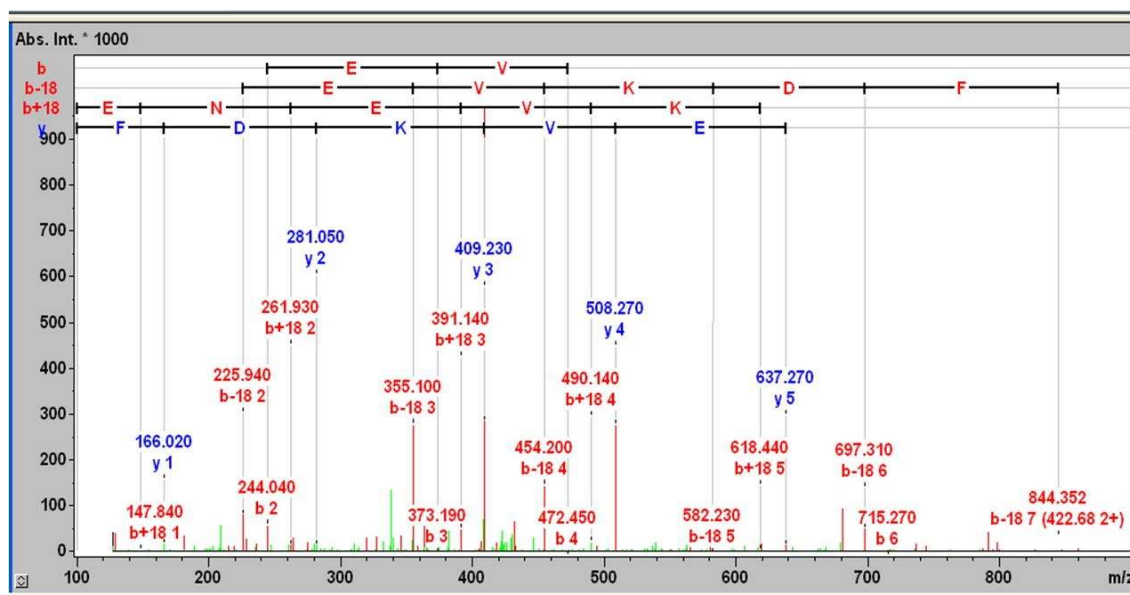


**Figure S29.** Representative CID MS/MS spectrum of chymotryptic peptide W.ENEVKDF.N (137-143), selecting the  $m/z$  440.700  $[M + 2H]^{2+}$  as precursor ion.

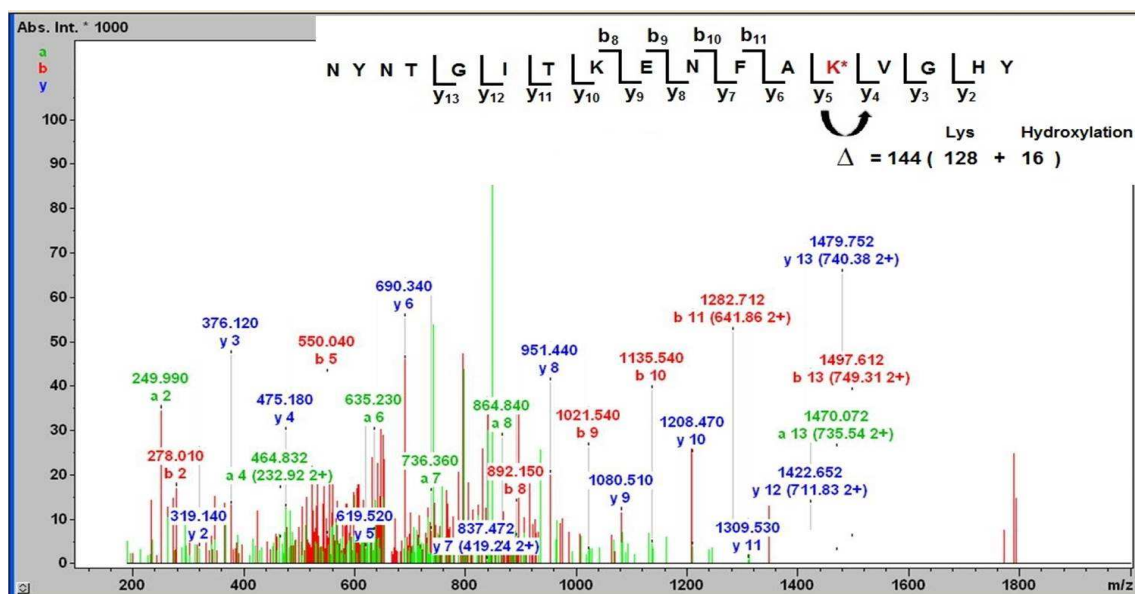


**Figure S30.** Representative CID MS/MS spectrum of chymotryptic peptide F.NYNTGITKENFAK\*VGHY.T (144-160), selecting the m/z 657.980 [M + 3H]<sup>3+</sup> as precursor ion; and showing the K156 hydroxylation site observed on the antigen-5 protein.

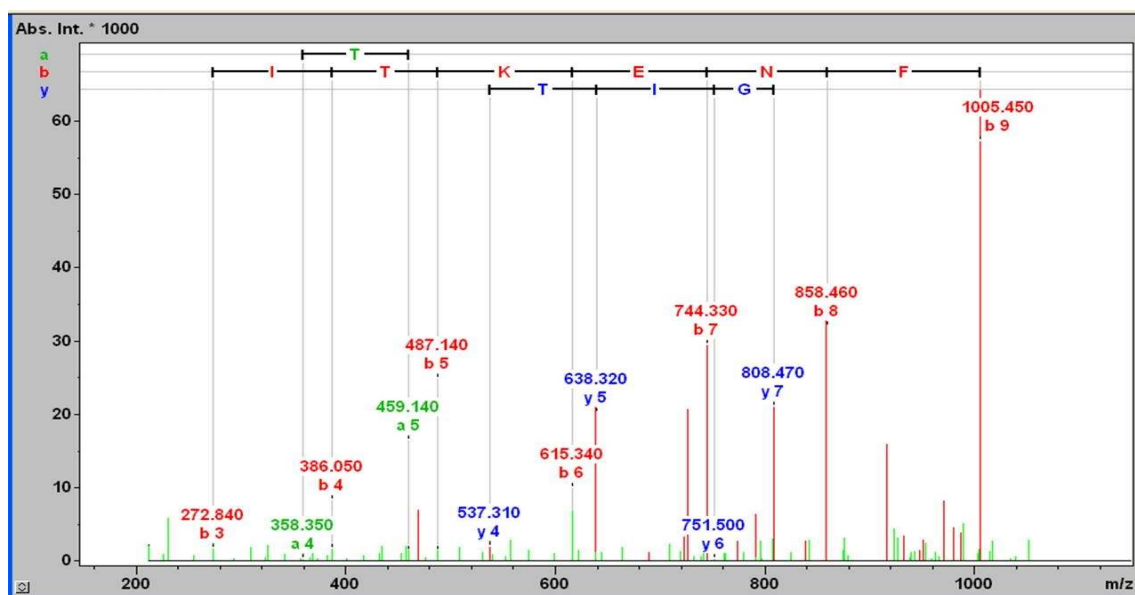


**Figure S31.** Representative CID MS/MS spectrum of chymotryptic peptide Y.NTGITKENF.A (146-154), selecting the m/z 1023.460 [M + H]<sup>+</sup> as precursor ion.

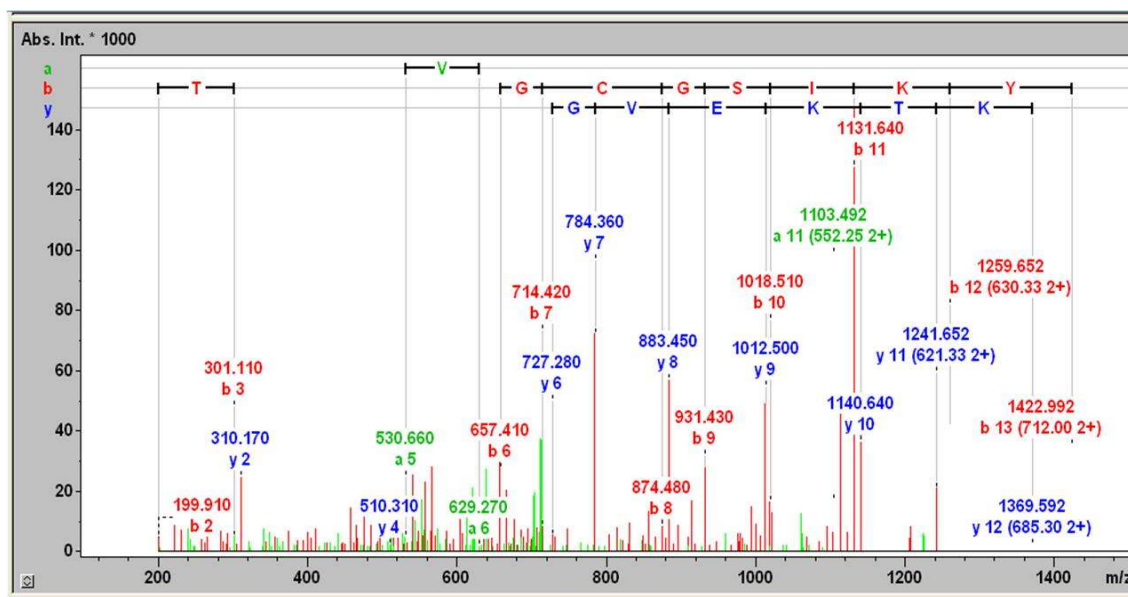


**Figure S32.** Representative CID MS/MS spectrum of chymotryptic peptide W.AKTKEVGCSTIKY.I (166-178), selecting the  $m/z$  720.827  $[M + 2H]^{2+}$  as precursor ion.

Representative mass spectra of antigen-5 (spot 125) from the venom of the social wasp *P. paulista* by *in-gel* protein digestion using Glu-C/V8 protease.

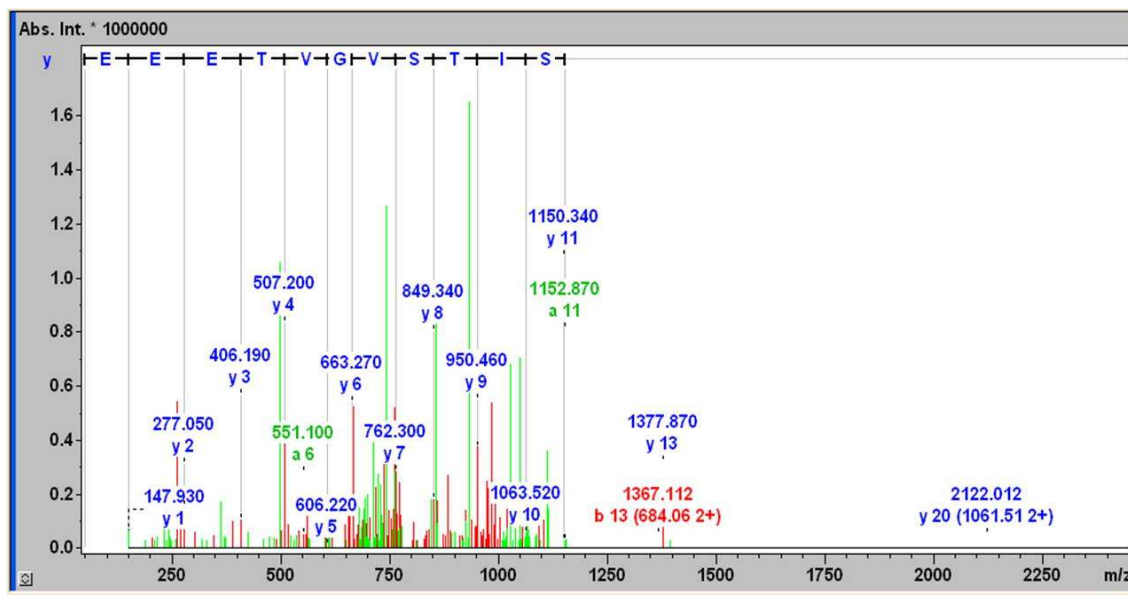


**Figure S33.** Representative CID MS/MS spectrum of proteolytic peptide from Glu-C/V8 protease E.STKPSSKNCAKVSITSGVTEEE.K (21-43), selecting the  $m/z$  789.726  $[M + 2H]^{2+}$  as precursor ion.

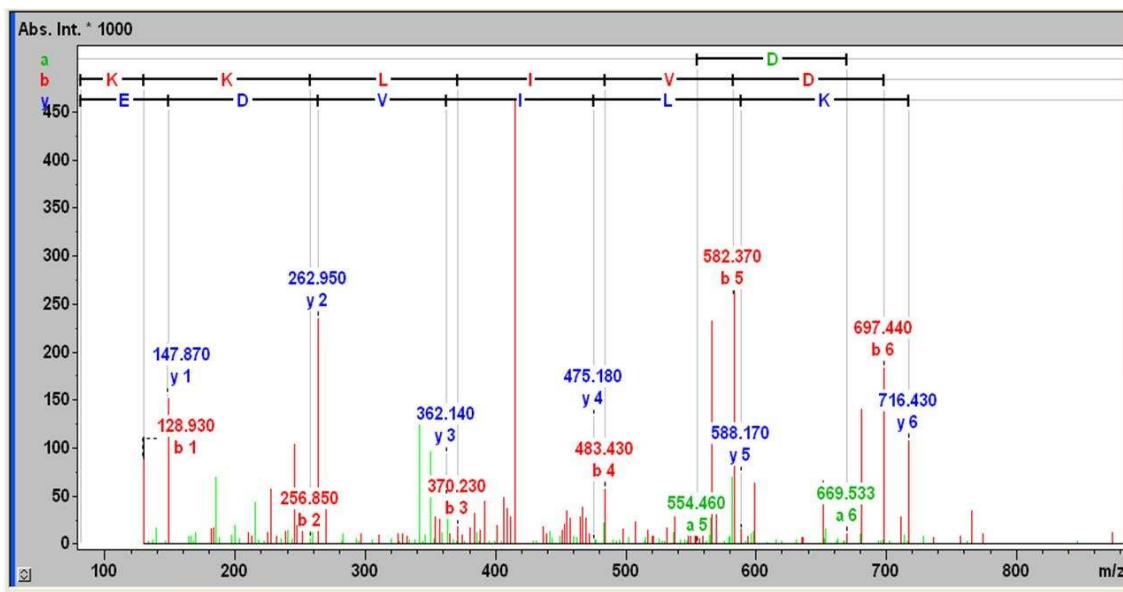


**Figure S34.** Representative CID MS/MS spectrum of proteolytic peptide from Glu-C/V8 protease E.KKLIVDE.H (44-50), selecting the  $m/z$  422.790  $[M + 2H]^{2+}$  as precursor ion.

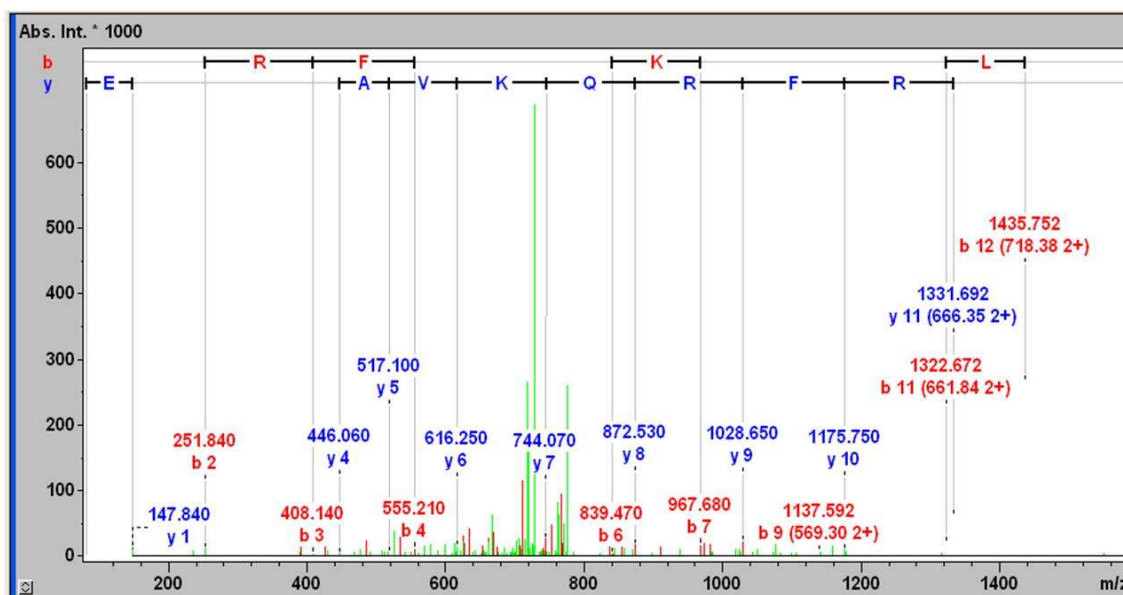


**Figure S35.** Representative CID MS/MS spectrum of proteolytic peptide from Glu-C/V8 protease E.HNRFRQKVAQG.L (51-63), selecting the  $m/z$  791.919  $[M + 2H]^{2+}$  as precursor ion.

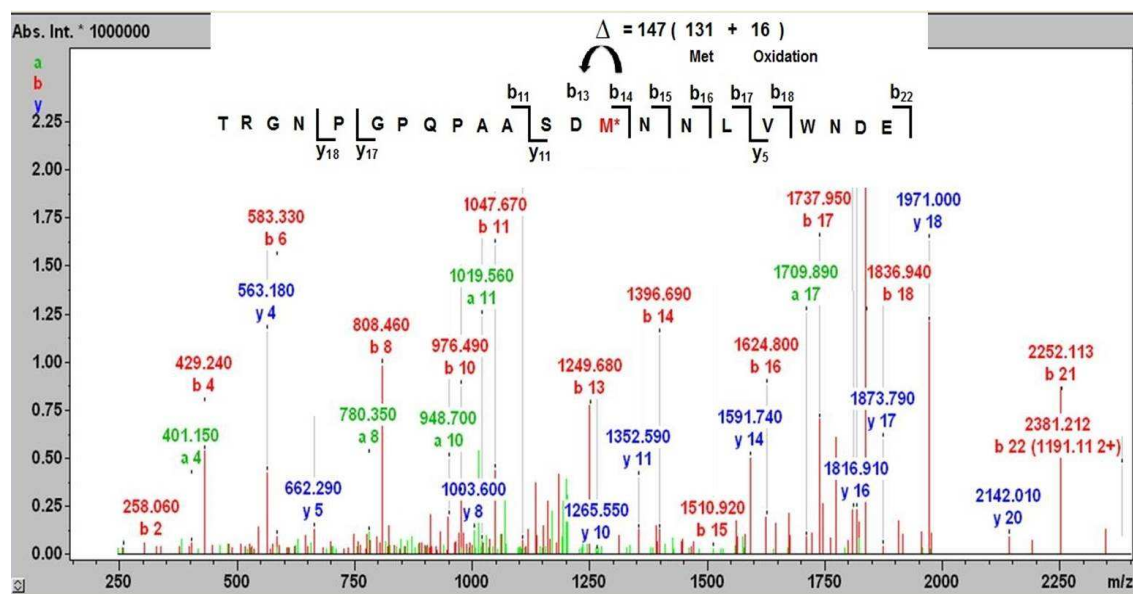


**Figure S36.** Representative CID MS/MS spectrum of proteolytic peptide from Glu-C/V8 protease E. TRGNPGPQPAASD M\* NNLVWNDE.L (64-85), selecting the m/z 1200.024 [M + 2H]<sup>2+</sup> as precursor ion; and showing the M77 oxidation site observed on the antigen-5 protein.

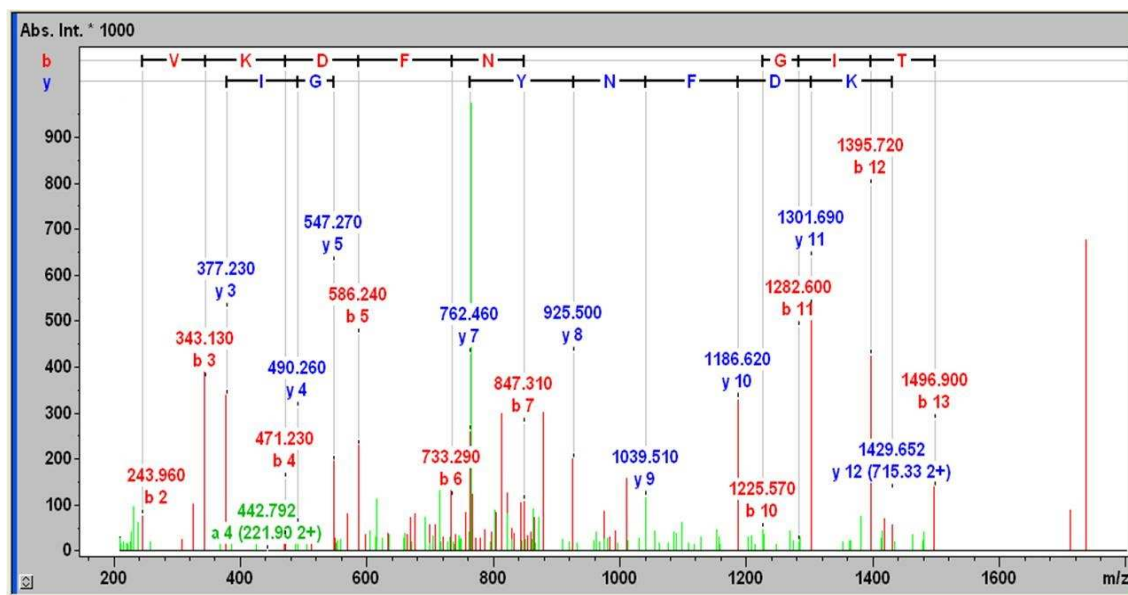


**Figure S37.** Representative CID MS/MS spectrum of proteolytic peptide from Glu-C/V8 protease E. NEVKDFNYNTGITKE.N (138-152), selecting the m/z 886.442 [M + 2H]<sup>2+</sup> as precursor ion.

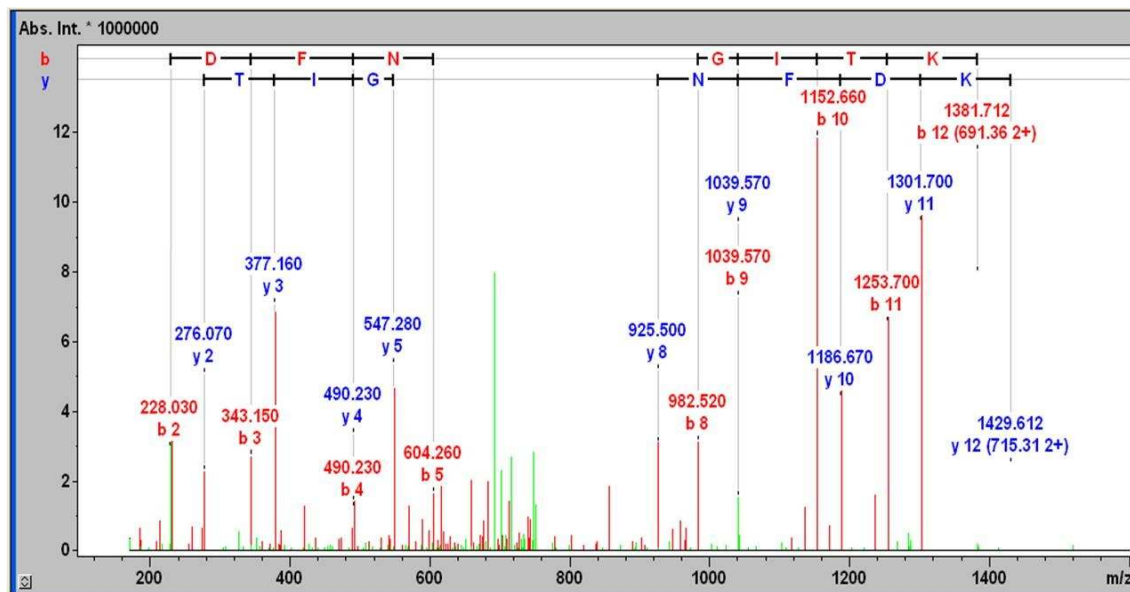


**Figure S38.** Representative CID MS/MS spectrum of proteolytic peptide from Glu-C/V8 protease E.VKDFNYNTGITKE.N (140-152), selecting the  $m/z$  764.875  $[M + 2H]^{2+}$  as precursor ion.

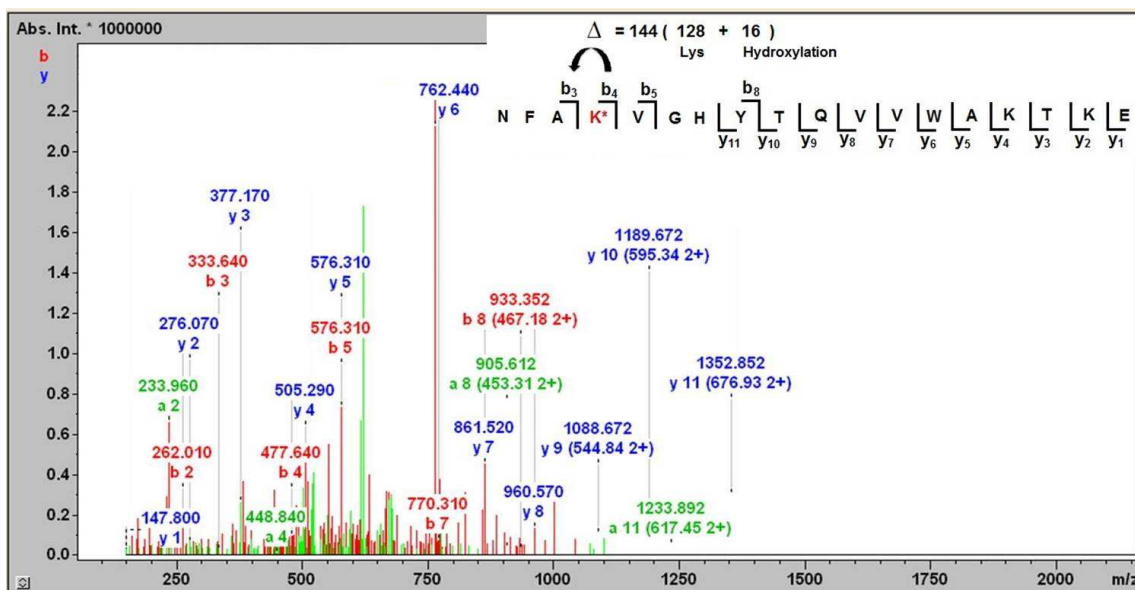


**Figure S39.** Representative CID MS/MS spectrum of proteolytic peptide from Glu-C/V8 protease E.NFAK\*VGHYTQVWAKTKE.V (153-170), selecting the  $m/z$  531.280  $[M + 4H]^{4+}$  as precursor ion; and showing the K156 hydroxylation site observed on the antigen-5 protein.

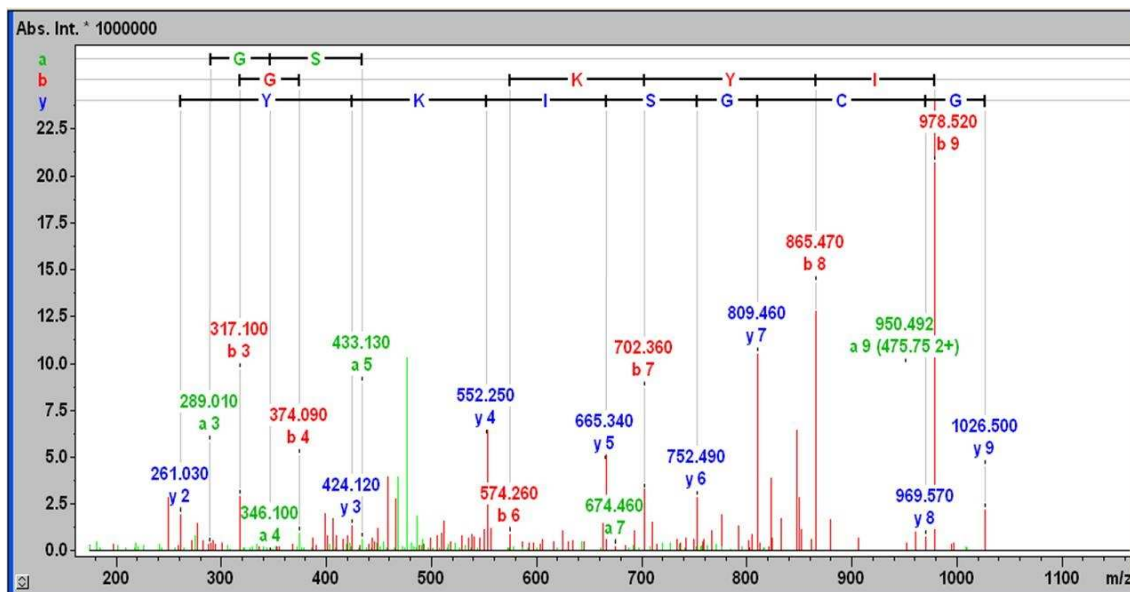
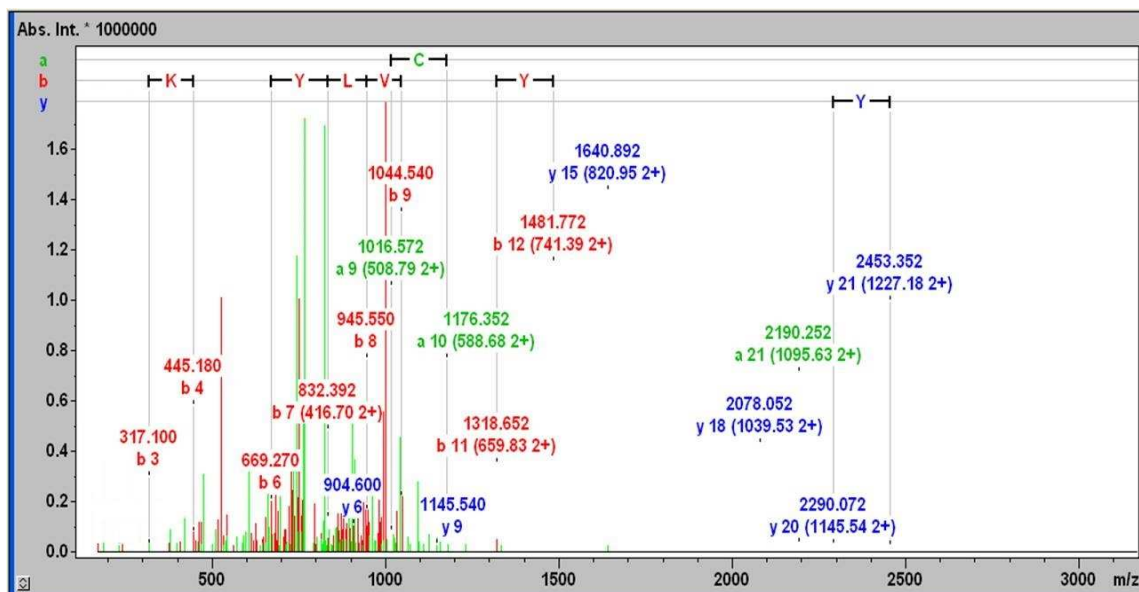


**Figure S40.** Representative CID MS/MS spectrum of proteolytic peptide from Glu-C/V8 protease E.VGCGSIKYIE.K (171-180), selecting the  $m/z$  563.260  $[M + 2H]^{2+}$  as precursor ion.



**S41.** Representative CID MS/MS spectrum of proteolytic peptide from Glu-C/V8 protease E.KGMKSHYLVLCNYGPAGNVLGAQIYEIK (181-207), selecting the  $m/z$  781.143  $[M + 4H]^{4+}$  as precursor ion.