Consolidated peptide/protein databases including markers for application II

Deliverable D9.9

21 June 2022

Author(s): Karim Arafah[1], Dalel Askri[1], Sébastien Voisin[1] & Philippe Bulet[2]

[1]Plateforme BioPark d'Archamps, Archamps, France
[2]CR, University Grenoble Alpes, IAB Inserm 1209, CNRS UMR5309, Grenoble, France

PoshBee

**Pan-european assessment, monitoring, and mitigation
of stressors on the health of bees**

_____

Prepared under contract from the European Commission

Grant agreement No. 773921
EU Horizon 2020 Research and Innovation action

Project acronym:	**PoshBee**
Project full title:	**Pan-european assessment, monitoring, and mitigation of stressors on the health of bee**
Start of the project:	June 2018
Duration:	60 months
Project coordinator:	Professor Mark Brown
	Royal Holloway and Bedford New College www.poshbee.eu

Deliverable title:	Consolidated peptide/protein databases including markers for application II
Deliverable n°:	D9.9
Nature of the deliverable:	Other
Dissemination level:	Public

WP responsible:	WP9
Lead beneficiary:	CNRS

Citation:	Arafah, K., Askri, D., Voisin, S. & Bulet, P. (2022). *Consolidated peptide/protein databases including markers for application II* Deliverable D9.9 EU Horizon 2020 PoshBee Project, Grant agreement No. 773921.

Due date of deliverable:	Month n°49
Actual submission date:	Month n°49

Deliverable status:

Version	Status	Date	Author(s)
1.0	Final	21 June 2022	Karim Arafah[1], Dalel Askri[1], Sébastien Voisin[1] & Philippe Bulet[2]
			[1]*Plateforme BioPark d'Archamps, Archamps, France* [2]*CR, University Grenoble Alpes, IAB Inserm U1209, CNRS UMR 5309, Grenoble, France*

The content of this deliverable does not necessarily reflect the official opinions of the European Commission or other institutions of the European Union.

_____

Table of contents

_____

Introduction

This database, referred to as BOMDBase-1.0, lists the *Bombus* proteins identified during the proteomics analysis of the bee haemolymph samples (Table 1) provided by the PoshBee consortium, as part of the experiments of work packages (WPs) 5 "Effects of agrochemical-nutrition interactions on bee health in the laboratory" & 6 "Effects of agrochemical-pathogen interactions on bee health in the laboratory", or collected from other bumblebees, as part of the experiments done within WP9 "OMICS of agrochemical responses in bees".

BOMDBase-1.0 has been made public. BOMDBase-1.0 is intended to be mined by any researcher looking for specific proteins or interested in cross-referencing their findings with ours. BOMDBase-1.0 will be further implemented and updated as more samples are provided by the PoshBee partners and analysed by 10-BIOP.

The list of the identified proteins will be refined. For example, a number of identified proteins have their sequence annotated as hypothetical, low quality, and/or uncharacterized (not annotated) in the reference databases we used during the MS/MS spectra identification step. We intend to complete BOMDBase-1.0 with the description of these proteins that are currently unidentified/putative in the available reference databases (e.g. NCBI, UniProtKB, BeeBase).

Table 1: List of haemolymph pools analysed by 10-BioP and used for BOMDBase-1.0

WP	Pools
5-Agrochemical-Nutrition interactions	38
6-Agrochemical_pathogen interactions	32
9-Bacterial infection	41
9-Thermal stress	16

1. Bottom-up proteomics workflow to fill the BOMDBase-1.0

Haemolymph is the circulating body fluid in invertebrates, equivalent to human blood. As summarized in Figure 1, haemolymph samples collected from *Bombus terrestris* were regrouped into pools of five individual haemolymphs based on the individual Mass Fingerprints generated with MALDI BeeTyping®. The pools were dried by vacuum centrifugation before being analysed by a bottom-up proteomics approach, according to the protocol reported in Masson *et al.,* 2018 and Houdelet *et al.,* 2020. Briefly, dried samples were suspended in 20 µL of Rapigest 0.1% in 50 mM ammonium bicarbonate (ABC) buffer, and the proteins' cysteine residues were reduced (disulfide bonds are opened) and alkylated (blocked) using dithiothreitol and 4-vinyl-pyridine, respectively. The reduced-alkylated proteins were then digested by trypsin.

After an overnight incubation, samples were acidified, centrifuged, and the supernatant transferred into an HPLC autosampler vial. Samples were separated on a reverse-phase C_{18} capillary column installed on a U3000 nano-HPLC connected to a high-resolution mass spectrometer, a Q-Exactive Orbitrap (all instruments Thermo Scientific). A 155-min long chromatographic method using a linear gradient of acidified acetonitrile was used to separate the peptide digests. The separated peptides were analysed online by the electrospray interface connected to the Q-Exactive Orbitrap for detection and acquisition of MS/MS spectra.
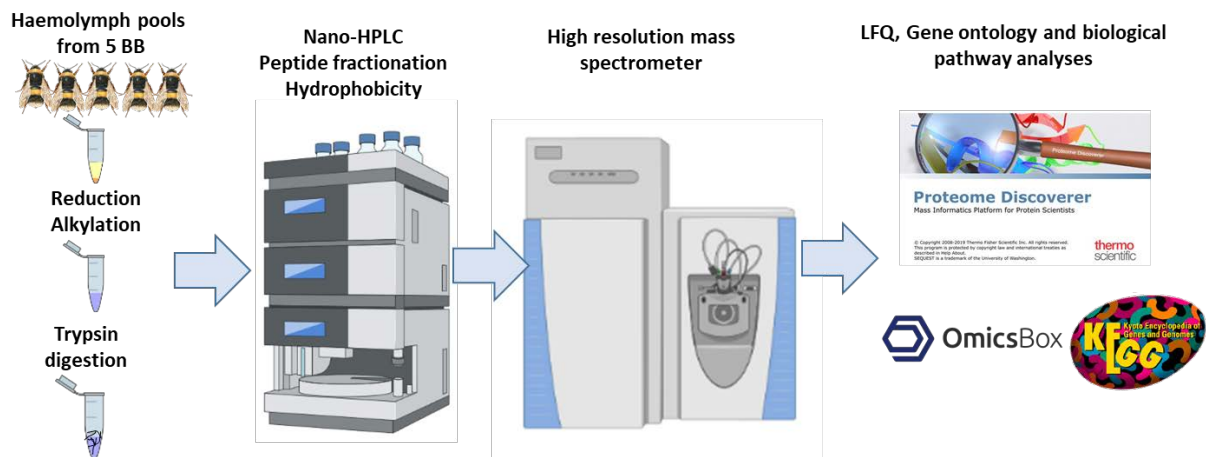
_____

Figure 1: Workflow of bottom-up proteomic analysis of haemolymph samples

2. MS/MS spectra matching against public protein databases

The search algorithm Sequest HT was run by Proteome Discoverer 2.5 (Thermo Fisher Scientific) to match the acquired peptide MS/MS spectra to a protein sequence database made of entries aggregated from NCBI (https://www.ncbi.nlm.nih.gov/protein) and Uniprot (https://www.uniprot.org). See Table 2 below for a detailed list of the entries (April 2022 version). The following parameters were used: trypsin digest with two maximum missed cleavages; a tolerance of 10 ppm/0.02 Da for precursors and fragment ions, respectively; cysteine pyridyl-ethylation was set as a fixed modification (4-VP); C-terminal protein amidation, methionine and tryptophan oxidation were set as variable modifications. The identification confidence was set at a false discovery rate (FDR) of 1%.

Table 2: Organisms added in the protein sequence database used for matching MS analysis

Host	*Pathogen*	UniprotKB
Apis mellifera	*Acarapis woodi*	TrEMBL (20)
Apis mellifera	*Achroia grisella*	TrEMBL (07)
Apis mellifera	*Actinobacteria*	restriced to apis - TrEMBL (12)
Apis mellifera	Acute bee paralysis virus	Swiss-prot (3) TrEMBL (381)
Apis mellifera	*Aethina tumida*	Swiss-prot (1) TrEMBL (132)
Apis mellifera	*Alphaproteobacteria*	TrEMBL (818) – focus on Alphaproteobacteria+Hymenoptera
Apis mellifera	*Apocephalus borealis*	TrEMBL (05)
Apis mellifera	*Ascosphaera apis*	TrEMBL (6492)
Apis mellifera	*Aspergillus flavus*	Swiss-prot (300)
Apis mellifera	*Aspergillus fumigatus*	Swiss-prot (2916)
Apis mellifera	*Aspergillus niger*	Swiss-prot (600) + Q7M3M5 (Q7M3M5_APIME) TrEMBL (01)
Apis mellifera	*Bacillus cereus*	Swiss-prot (7394)
Apis mellifera	*Bacillus megatherium*	TrEMBL (01)
Apis mellifera	*Bacillus oleronius*	TrEMBL (4898) – New name *Heyndrickxia oleronia*
Apis mellifera	*Bacillus pulvifaciens*	TrEMBL (02)
Apis mellifera	*Bacillus thuringiensis*	Swiss-prot (1699)
Apis mellifera	*Bacteriocin*	TrEMBL (17)
Apis mellifera	*Bartonella apis*	TrEMBL (6660)
Apis mellifera	*Betaproteobacteria*	Swiss-prot 24527
Apis mellifera	*Bifidobacterium animalis*	Swiss-prot (153) - TrEMBL (11249)
Apis mellifera	*Bifidobacterium asteroides*	TrEMBL (18241)
Apis mellifera	*Bifidobacterium coryneforme*	TrEMBL (1385)
Apis mellifera	*Bifidobacterium cuniculi*	TrEMBL (2203)
Apis mellifera	*Bifidobacterium globosum*	TrEMBL (18871)

Apis mellifera	Black queen cell virus	TrEMBL (861)
Apis mellifera	*Braula coeca*	TrEMBL (09)
Apis mellifera	*Brevibacillus laterosporus*	Swiss-prot (01) - TrEMBL (33034)
Apis mellifera	Chronic bee paralysis virus	TrEMBL (139)
Apis mellifera	Cloudy wing virus	TrEMBL (01)
Apis mellifera	*Crithidia mellificae*	TrEMBL (63)
Apis mellifera	Deformed wing virus	TrEMBL (3546)
Apis mellifera	*Delftia acidovorans*	Swiss-prot (267) - TrEMBL (44309)
Apis mellifera	*Enterococcus cloacae*	Swiss-prot (01)
Apis mellifera	*Enterococcus faecalis*	Swiss-prot (557) - TrEMBL (112284)
Apis mellifera	*Enterococcus avium*	Swiss-prot (01) - TrEMBL (22554)
Apis mellifera	*Escherichia coli*	*TrEMBL (01) - restricted to Apis*
Apis mellifera	*Filamentous virus*	*TrEMBL (270) - restricted to Apis*
Apis mellifera	*Firmicutes*	*TrEMBL (2860) - restricted to Apis*
Apis mellifera	*Frischella perrara*	*TrEMBL (32) - restricted to Apis*
Apis mellifera	*Galleria mellonella*	*Swiss-prot (66) - TrEMBL (22317)*
Apis mellifera	Gammaproteobacteria (class)	*TrEMBL (6766) - restricted to Apis*
Apis mellifera	*Gilliamella apicola*	*TrEMBL (61479)*
Apis mellifera	*Gluconacetobacter*	*Swiss-prot (239) - TrEMBL (95481)*
Apis mellifera	*Iflavirus*	*Swiss-prot (01) - TrEMBL (4214)*
Apis mellifera	invertebrate iridescent virus type 6	Swiss-prot (86) - TrEMBL (265)
Apis mellifera	Israeli acute paralysis virus	Swiss-prot (01) - TrEMBL (161)
Apis mellifera	Kakugo virus	TrEMBL (61)
Apis mellifera	Kashmir bee virus	TrEMBL (189)
Apis mellifera	*Klebsiella oxytoca*	Swiss-prot (73) - TrEMBL (74923)
Apis mellifera	*Lactobacillus acidophilus*	Swiss-prot (296) - TrEMBL (7812)
Apis mellifera	*Lactobacillus crispatus*	Swiss-prot (01) - TrEMBL (39162)
Apis mellifera	*Lactobacillus Firm 4*	TrEMBL (2384)+lactobacillus apis TrEMBL (2908)

_____

Apis mellifera	*Lactobacilluskunkeei*	TrEMBL (10154)
Apis mellifera	*Lactobacillus plantarum*	Swiss-prot (513) - TrEMBL (61915)
Apis mellifera	*Lactobacillus sp.*	TrEMBL (22638)
Apis mellifera	*Lactobacillus Firm 5*	TrEMBL (15713)
Apis mellifera	Lake Sinai virus (LSV)	TrEMBL (377)
Apis mellifera	*Lotmaria passim*	TrEMBL (103)
Apis mellifera	*Melissococcus plutonius*	Swiss-prot (01) - TrEMBL (3616)
Apis mellifera	Moku virus	TrEMBL (48)
Apis mellifera	*Morganella morganii*	Swiss-prot (16) - TrEMBL (62922)
Apis mellifera	*Neisseria*	Swiss-prot (2837)
Apis mellifera	*Nosema apis*	TrEMBL (3045)
Apis mellifera	*Nosema ceranae*	Swiss-prot (23) - TrEMBL (5806)
Apis mellifera	*Paenibacillus alvei*	Swiss-prot (4) - TrEMBL (23938)
Apis mellifera	*Pandora delphacis*	TrEMBL (3)
Apis mellifera	*Pantoea agglomerans*	Swiss-prot (54) - TrEMBL (37970)
Apis mellifera	*Pantoea ananatis*	Swiss-prot (25) - TrEMBL (13948)
Apis mellifera	*Parasaccharibacter apium*	TrEMBL (3476)
Apis mellifera	*Pediococcus acidilactici*	Swiss-prot (9) - TrEMBL (18043)
Apis mellifera	*Proteus mirabilis*	Swiss-prot (576) - TrEMBL (47433)
Apis mellifera	*Pseudomonas oryzihabitans*	TrEMBL (30443)
Apis mellifera	*Rahnella aquatilis*	Swiss-prot (04) - TrEMBL (10030)
Apis mellifera	*Rahnella terrigena*	no entry – if google search *Roultella terrigena* was found:-)
Apis mellifera	*Raoutella ornithinolytica*	ou *Raoultella - we considered Raoultella. Swiss-prot (03) TrEMBL (25164)*
Apis mellifera	*Sacbrood bee virus (SBV)*	TrEMBL (134)
Apis mellifera	*Saccharibacter sp*	TrEMBL (3960)
Apis mellifera	*Saccharomycotina*	Swiss-prot (14463)
Apis mellifera	*Serratia*	liste limited to ATCC, Swiss-prot (57) - TrEMBL (9175)
Apis mellifera	*Serratia liquefaciens*	Swiss-prot (05) - TrEMBL (13379)

Apis mellifera	*Serratia marcescens*	Swiss-prot (159) - TrEMBL (84600)
Apis mellifera	*Simonsiella*	TrEMBL (2220)
Apis mellifera	Slow paralysis virus	TrEMBL (100)
Apis mellifera	*Snodgrassella alvi*	TrEMBL (24796)
Apis mellifera	*Sphingomonas parapaucimobilis*	TrEMBL (3700)
Apis mellifera	*Spiroplasma apis*	Swiss-prot (01) - TrEMBL (1000)
Apis mellifera	*Spiroplasma melliferum*	Swiss-prot (24) - TrEMBL (2427)
Apis mellifera	*Staphylococcus epidermidis*	Swiss-prot (2481) - TrEMBL (32084)
Apis mellifera	Tobacco ringspot virus	Swiss-prot (11) - TrEMBL (271)
Apis mellifera	*Tropilaelaps clareae*	Swiss-prot (05)
Apis mellifera	*Tropilaelaps mercedesae*	TrEMBL (14507)
Apis mellifera	*Varroa destructor*	Swiss-prot (01) - TrEMBL (20811)
Apis mellifera	*Varroa destructor virus 1*	TrEMBL (520)
Apis mellifera	*Varroa jacobsoni*	TrEMBL (48)
Apis mellifera	*Vespa crabro*	Swiss-prot (06) - TrEMBL (32)
Apis mellifera	*Vespa orientalis*	Swiss-prot (06) - TrEMBL (32)
Apis mellifera	*Vespa velutina*	Swiss-prot (09) - TrEMBL (217)
Apis mellifera	X virus	*Apis* (83 entries)
Apis mellifera	Y virus	*Apis* (415 entries)
Bombus terrestris	*Crithidia bombi*	TrEMBL (32)
Bombus terrestris	Deformed wing virus (DWV)	TrEMBL (3546)
Bombus terrestris	*Lotmaria passim*	TrEMBL (103)
Bombus terrestris	*Actinobacteria*	TrEMBL (65)
Bombus terrestris	*Actinomycetales*	Swiss-prot (23)
Bombus terrestris	Acute bee paralysis virus (ABPV)	Swiss-prot (03) - TrEMBL (381)
Bombus terrestris	Alphaproteobacteria	TrEMBL (818) limited to hymenoptera
Bombus terrestris	*Ampelomyces quisqualis*	Swiss-prot (01) - TrEMBL (11232)
Bombus terrestris	*Apicystis bombi*	TrEMBL (01)

_____

Bombus terrestris	*Ascosphaera sp*	TrEMBL (64)
Bombus terrestris	*Aureobasidium pullulans*	Swiss-prot (05)
Bombus terrestris	*Bacillaceae*	Swiss-prot (18083)
Bombus terrestris	*Bacillus amyloliquefaciens*	Swiss-prot (542)
Bombus terrestris	*Bacillus thuringiensis*	Swiss-prot (1699)
Bombus terrestris	*Bacillus thuringiensis aizawai*	Swiss-prot (09) - TrEMBL (77)
Bombus terrestris	*Bacillus thuringiensis israelensi*	Swiss-prot (12) - TrEMBL (19678)
Bombus terrestris	*Bacillus thuringiensis kurstaki*	Swiss-prot (24) - TrEMBL (6566)
Bombus terrestris	*Bacteroidetes*	Swiss-prot (6775)
Bombus terrestris	*Beauveria bassiana*	Swiss-prot (24) - TrEMBL (43861)
Bombus terrestris	*Betabaculovirus*	Swiss-prot (23) - TrEMBL (5625)
Bombus terrestris	*Betaproteobacteria*	Swiss-prot (24527)
Bombus terrestris	*Bifidobacteriaceae*	Swiss-prot (1054)
Bombus terrestris	Black queen cell virus (BQCV)	TrEMBL (861) - present in *Apis* pathogens DB
Bombus terrestris	*Burkholderiales*	Swiss-prot (17714)
Bombus terrestris	Chronic bee paralysis virus	TrEMBL (139)
Bombus terrestris	*Clonostachys rosea*	TrEMBL (18098)
Bombus terrestris	*Conopidae*	TrEMBL (498)
Bombus terrestris	*Cordyceps fumosorosea*	Swiss-prot (04) - TrEMBL (10160)
Bombus terrestris	*Enterobacteriaceae*	Swiss-prot (47498)
Bombus terrestris	*Enterococcaceae*	Swiss-prot (615)
Bombus terrestris	*Firmicutes_Bombus*	TrEMBL (6234)
Bombus terrestris	Gammaproteobacteria	Gammaproteobacteria+insects Swiss-prot (290) - TrEMBL (4733) and Gammaproteobacteria+hymenoptera TrMBL (05)
Bombus terrestris	*Gliocladium catenulatum*	TrEMBL (06)
Bombus terrestris	*Granulovirus*	Swiss-prot (23) - TrEMBL (5671)
Bombus terrestris	*Heterorhabditis sp*	TrEMBL (140)
Bombus terrestris	Kashmir bee virus	TrEMBL (189) - Already present in *Apis* pathogens DB

Bombus terrestris	*Lactobacillaceae*	Limited lactobacillaceae bombus TrEMBL (09) + 9286 entries TrEMBL *bombilactobacillus*
Bombus terrestris	Lake Sinai virus (LSV)	TrEMBL (377) - Already present in *Apis pathogens DB*
Bombus terrestris	*Metarhizium anisopliae*	Swiss-prot (86) - TrEMBL (44270)
Bombus terrestris	*Metarhizium brunneum*	Swiss-prot (08) - TrEMBL (11565)
Bombus terrestris	*Neisseriaceae*	Swiss-prot (2763)
Bombus terrestris	*Nosema bombi*	Swiss-prot (06)
Bombus terrestris	*Nosema ceranae*	Swiss-prot (23) - TrEMBL (5806)
Bombus terrestris	*Orbaceae*	TrEMBL (1861) with *Orbaceae bombi* - Already present in *Apis* (TrEMBL 4613)
Bombus terrestris	*Pseudomonadaceae*	Swiss-prot (11304)
Bombus terrestris	Sacbrood bee virus (SBV)	TrEMBL (134)
Bombus terrestris	*Steinernema kraussei*	TrEMBL (04)
Bombus terrestris	*Steinernema spp*	TrEMBL (114)
Bombus terrestris	*Streptococcaceae*	Swiss-prot (14143)
Bombus terrestris	*Trichoderma atroviride*	Swiss-prot (15) - TrEMBL (12067)
Bombus terrestris	*Trichoderma harzianum*	Swiss-prot (36) - TrEMBL (51219)
Bombus terrestris	*Verticillium lecanii*	Swiss-prot (01) - TrEMBL (74)
Osmia bicornis	*Agromyces*	Swiss-prot (05)
Osmia bicornis	*Arthrobacter*	Swiss-prot (1263)
Osmia bicornis	*Conexibacter*	Swiss-prot (04) - TrEMBL (20164)
Osmia bicornis	*Crithidia expoeki*	Swiss-prot (08)
Osmia bicornis	*Deformed wing virus*	TrEMBL (3546) - already present in Apis pathogens DB
Osmia bicornis	*Diaphorobacter*	Swiss-prot (249)
Osmia bicornis	*Lamia*	Swiss-prot (29) - TrEMBL (90)
Osmia bicornis	*Massilia*	Swiss-prot (05) - TrEMBL (294877)
Osmia bicornis	*Metschnikowia*	TrEMBL (31054)
Osmia bicornis	*Microlunatus*	TrEMBL (43542)
Osmia bicornis	*Nosema bombi*	Swiss-prot (06) - already present in *Bombus* pathogens DB
Osmia bicornis	*Nosema ceranae*	Swiss-prot (23) - TrEMBL (5806) - already present in *Bombus* pathogens DB

_____

Osmia bicornis	*Nosematidae*	Swiss-prot (31) - TrEMBL (15575)
Osmia bicornis	*Paenibacillus*	Swiss-prot (68)
Osmia bicornis	*Pseudomonas*	Swiss-prot (12728)
Osmia bicornis	*Rhabdovirus*	Swiss-prot (155) - TrEMBL (17162)
Osmia bicornis	*Rubrobacter*	Swiss-prot (244) - TrEMBL (20523)
Osmia bicornis	*Skermanella*	TrEMBL (21932)
Osmia bicornis	Slow bee paralysis virus	TrEMBL (100)
Osmia bicornis	*Solirubrobacter*	TrEMBL (6590)
Osmia bicornis	*Sphingomonas*	Swiss-prot (582)
Osmia bicornis	*Trypanosomatidae*	Swiss-prot (662) - TrEMBL (406728)
Osmia bicornis	*Tubulinosema*	TrEMBL (2996)

3. Structure of BOMDBase-1.0

The protein identification reports generated as described above for each analysed haemolymph pool were aggregated together and the protein entries sorted by accession numbers. Redundant entries with the same accession numbers were removed. Different entries corresponding to different isoforms of the same protein were kept. For this initial version of BOMDBase-1.0, the protein lists were restricted to those belonging to *Bombus*. This merged database contains all identified proteins, with no distinction between the experimental condition or quantification data. BOMDBase-1.0 contains **1353 accession numbers**, corresponding to **1042 proteins** of *Bombus* species (646 from *Bombus terrestris*).

In the next section, the content of each column in the BOMDBase-1.0 is described.

The names in bold are the column headers:

Accession: Reference code of the protein entry into the original protein sequence database. Entries of type P81463 or A0A6P3UBV8 are from UniprotKB, other entry types are from NCBI.

Description: The description of that protein in the UniprotKB or NCBI database.

Species: The organism to which that protein belongs.

NbAAs; MW [kDa]: The number of amino acids (**NbAAs**) and the molecular weight in kilodaltons (**MW [kDa]**) of the full protein sequence. *Caution! The sequence used for these calculations is the full protein sequence deduced from the precursor form in the original UniprotKB/NCBI database entry.* As mentioned for the coverage, to have a corresponding molecular mass, additional calculation needs to be conducted (e.g., deduction of 2 Da per cysteine pairing, and/or elimination of the molecular mass of the signal peptide if predicted by [SignalP-5.0 server](#), and/or the molecular mass of a pro-domain predicted by [ProP1.0 Server](#)).

CalcpI: Calculated isoelectric point of the full protein sequence. *Caution! The sequence used for this calculation is the full protein sequence in the original UniprotKB/NCBI database entry, based on the full genomic sequence.*

Biological Process, Cellular Component, Molecular Function, GO Accessions: Gene Ontology (GO) terms recorded in the protein entry. http://geneontology.org/ .

Pfam IDs: Pfam protein domains recorded in the entry database. http://pfam.xfam.org/ .

Entrez Gene ID; Gene Symbol: Genetic information in the protein entry.

4. References

Masson V, Arafah K, Voisin S, Bulet P. (2018) Comparative Proteomics Studies of Insect Cuticle by Tandem Mass Spectrometry: Application of a Novel Proteomics Approach to the Pea Aphid Cuticular Proteins. Proteomics. 2018 Feb;18(3-4). [doi: 10.1002/pmic.201700368](#). Epub 2018 Feb 2. PMID: 29327416

Houdelet C, Sinpoo C, Chantaphanwattana T, Voisin SN, Bocquet M, Chantawannakul P, Bulet P. (2021) Proteomics of Anatomical Sections of the Gut of *Nosema*-Infected Western Honeybee (*Apis mellifera*) Reveals Different Early Responses to *Nosema* spp. Isolates. J Proteome Res. 2021 Jan 1;20(1):804-817. [doi: 10.1021/acs.jproteome.0c00658](#). Epub 2020 Dec 11. PMID: 33305956