



HAL
open science

Comparative Proteomics Studies of Insect Cuticle by Tandem Mass Spectrometry: Application of a Novel Proteomics Approach to the Pea Aphid Cuticular Proteins

Victor Masson, Karim Arafah, Sébastien Voisin, Philippe Bulet

► To cite this version:

Victor Masson, Karim Arafah, Sébastien Voisin, Philippe Bulet. Comparative Proteomics Studies of Insect Cuticle by Tandem Mass Spectrometry: Application of a Novel Proteomics Approach to the Pea Aphid Cuticular Proteins. *Proteomics*, 2018, 18 (3-4), pp.1700368. <10.1002/pmic.201700368>. <hal-02996346>

HAL Id: hal-02996346

<https://hal.science/hal-02996346v1>

Submitted on 9 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Comparative proteomics studies of insect cuticle by tandem mass spectrometry: Application of a novel proteomics approach to the pea aphid cuticular proteins

Masson Victor¹, Arafah Karim¹, Voisin Sébastien¹, Bulet Philippe^{1,2*}

¹Platform BioPark Archamps, Archamps, France

²Institute for Advanced Biosciences, CR Inserm U1209, CNRSUMR 5309, University of Grenoble-Alpes, Grenoble, France

***CORRESPONDING AUTHOR**

Dr Bulet Philippe

260, Avenue Marie Curie

Archamps Technopole

74166 Saint Julien-en-Genevois

France

E-mail: philippe.bulet@univ-grenoble-alpes.fr

Tel: +33450432521

Abstract

The exoskeleton of arthropods, called cuticle, is a biological composite material consisting principally of N-acetylglucosamine polymer embedded in cuticular proteins (CPs). The CPs have been studied and characterized by mass spectrometry in several cuticular structures and in many arthropods. Such analyses were usually carried out by protein extraction using SDS detergent followed by electrophoresis, allowing detection and identification of a large number of CPs. To build a repertoire of cuticular structures from *Bombyx mori*, *Apis mellifera*

Received: 10 03, 2017; Revised: 12 21, 2017; Accepted: 01 05, 2018

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/pmic.201700368](https://doi.org/10.1002/pmic.201700368).

This article is protected by copyright. All rights reserved.

and *Anopheles gambiae* we avoided the use of SDS and electrophoresis. Based on the combination of hexafluoroisopropanol solvent and of a surfactant solution fully compatible with MS, we successfully identified a high number of CPs in wings, legs and antennae of *An. gambiae*, as well as in the thoracic integument cuticle of *Ap. mellifera* pupae. The complete exoskeleton analysis of *B. mori* larvae allowed us to identify 85 CPs from the cuticle of a single larva. Finally, we tested our novel proteomics approach on cuticles left behind after the molt from the fourth instar of the pea aphid *Acyrtosiphon pisum*. Analysis of these cast cuticles allowed us to identify 100 *Ac. pisum* CPs as authentic constituents of the cuticle. These correspond to 68% of the total putative CPs previously annotated for this pea aphid. While this paper analyzes only the recovered cuticular proteins, peptides from many other proteins were also detected.

Keywords

Chitin, Cuticle, Insects, Mass spectrometry, Proteomics

Significance of the study

Cuticular proteins (CPs) have been identified in many arthropods. This study focused on comparative cuticular proteins extracted from different cuticular structures of insects by LC-MS/MS. To perform such analysis we combined the use of a solvent (HexaFluoroIsoPropanol, HFIP) and of a surfactant solution in order to eliminate SDS and the tedious electrophoresis approach often used to reduce biological matrix complexity. Using this fast and adaptable method for small quantities of biological material, we were able to apply this approach to notably improve our knowledge on the cuticular proteins that are involved in the composition of molts from the pea aphid *Acyrtosiphon pisum*. This novel proteomics approach should accelerate the investigation of the composition in CPs for cuticular structures difficult to collect in large amount.

1. Introduction

The cuticle is the outermost layer of insects and covers almost their entire body, serving as an articulated exoskeleton. It is the first physical barrier, protecting the body against dehydration, predators, chemicals, injuries and entomo-pathogenic microorganisms [1]. This biological composite material is produced by epidermal cells and organized principally as a network of chitin nano-fibers, a polymer of N-acetylglucosamine, embedded with CPs. Also associated are some lipids, and catecholamines [2]. CPs have been identified in many arthropods and their classification resulted in different families according to the presence of conserved consensus or semi-consensus motifs within their amino acid sequences. Thirteen distinct families of CPs have been described to date in arthropods [3-7]. The largest CP family is the CPR family, identified by Rebers & Riddiford in 1988, giving the name of R&R to the consensus sequence. The original R&R Consensus is part of a longer conserved sequence, pfam00379 (<http://pfam.janelia.org/family/>). In fact, the pfam sequence that allows annotators to classify a protein as a cuticular protein in the CPR family is pfam00379. This family has been subdivided into three sub-families (RR1/2/3), based on variations in the extended consensus motif. A hallmark of arthropod CPs is the absence of disulfide bonds [8, 9]. Exceptions are the CPAP1 and CPAP3 (peritrophin A-like proteins), which are characterized by the presence of six distinct separated cysteine residues forming a chitin-binding domain ChBD2. These proteins have one or three ChtBD2 domains [10]. The CPF and CPFL (CPF-like) families are also common in arthropods. While initially, the CPF family was considered to have a 51 amino acid consensus motif [11], the work of Togawa and colleagues [12] showed that the conserved region is composed of only 42-44 amino acids. CPFL share protein motifs with CPFs and could be defined by three conserved C-terminal residues [12]. The TWDL family, initially identified in *Drosophila melanogaster* [13], appears to be widespread among insects [3]. Four families of low complexity CPs, CPLCA (rich in alanine residues), CPLCG (presence of two glycine residues in a shared motif close to the C-terminus of the proteins), CPLCW (invariant tryptophan residue) and CPLCP (presence of

proline residues occurring as PV and PY frequently adjacent to each other), were first described in *An. gambiae* [14]. Three additional CP families have been described the (1) CPGs family (glycine-rich CPs organized into repeats of GGYGG or GGxGG) [15], (2) Apidermins identified in *Ap. mellifera* [16] and (3) CPTC family identified in *An. gambiae* carrying two cysteines [17]. Additional CPs identified in insect species have not been assigned yet to any CP families [14].

CPs, involved in tissue and body shapes, confer specific physicochemical and mechanical properties such as stiffness, flexibility, rigidity or structural integrity [9, 18-20]. Due to a limited capability of extension of the cuticle, the molting process compensates for this limitation and CPs have been shown to play key roles [21-23]. However, the precise roles of individual CPs are still poorly understood. Indeed, some CPs seem to be involved in insecticide resistance, as their expression has been increased in some insect strains resistant to insecticides [24-26]. CPs from the TWDL family are associated with defense reactions of insects against biotic and abiotic stressors by participating in the recognition of bacteria, and indirectly, in the activation of the synthesis of antimicrobial peptides [27]. Moreover, some CPs were also shown to be involved in the transmission of plant viruses by their insect vectors [28, 29]. CPs have been mainly characterized using molecular biology tools and *in silico* genomic studies on several arthropods species. New bioinformatics tools such as CuticleDB [30] and CutProtFam-Pred [5] have been developed to predict whether a protein is cuticular or not and to classify them according to their consensus sequences.

Many things about CPs have been learned using conventional transcriptomic techniques such as *in situ* hybridization or isolation of transcripts from epidermis underlying cuticle [7, 22]. But since the cuticle is a cell-free structure, transcriptomic techniques cannot be directly used on the latter, and the presence of a putative CP transcript in the epidermis is not sufficient to establish that the corresponding protein will be in the cuticle. However, due to its molecular complexity and to its physicochemical properties, a cuticle is a very resilient and challenging material from which to extract proteins. To date, proteomics studies on cuticular

Accepted Article

structures have been essentially based on SDS extraction/precipitation, sometimes including a separation by 1D- or 2D-SDS-PAGE, and subsequent trypsin digestion of the obtained supernatants, gel bands or SDS-insoluble pellets prior to structural elucidation by bottom-up proteomics of the CPs fractions. Such approaches have been used to identify CPs in different species and conditions [11, 17, 31-37]. In recent studies, between 1,100 and 1,500 *An. gambiae* adults were used to study the CPs of the complete body and between 80 and 330 insects to study the CPs composition of their appendages [32, 33]. Micas and colleagues used a combination of 2D electrophoresis and MALDI-MS to investigate the proteins present in 15 integuments of *Apis mellifera* workers [34]. Even if the approach of electrophoreses is widespread in bottom-up proteomics, this approach does have limitations because of the number of steps that must be completed, resulting in loss of information. So it is difficult to identify less abundant proteins and it requires sufficient amount of biological material to identify these.

Here, we developed a new sample preparation requiring limited amounts of cuticular tissues to perform proteomic analyses of insect cuticle from small appendages and structures (legs, antennae, wings, cast cuticle). This novel method avoids the use of detergents, reagents and separation step by gel electrophoresis. We applied our method to study the protein composition of the cuticle by tandem mass spectrometry. In order to compare our approach to already published data, we referred to the three following insect species *B. mori*, *An. gambiae* and *Ap. mellifera*. We compared the list of CPs identified using our analytical procedure to those previously published. Finally, we applied our experimental protocol to generate and present here new datasets on the CPs composition of the exoskeleton cuticle of the pea aphid, *Acyrtosiphon pisum*.

2. Materials and methods

2.1 Chemicals reagents

For sample preparation, *RapiGest*TM SF surfactant was purchased from Waters (Milford, USA), and ammonium bicarbonate (NH_4HCO_3), Hexafluoroisopropanol (HFIP),

iodoacétamide (IAA), dithiothreitol (DTT) were reagent grade from Sigma-Aldrich (St. Louis, MO). For LC-MS/MS, MilliQ water (Merck Millipore, Billerica, MA) was used. Acetonitrile (ACN) and trifluoroacetic acid (TFA) were of HPLC grade or higher and obtained from Carlo-Erba Reagents (Val de Reuil, France). LC-MS-grade formic acid (FA) was obtained from Fluka (Sigma-Aldrich, St. Louis, MO). For protein digestion, sequencing-grade modified trypsin (Promega, Madison, WI) was used. PBS buffer was purchased from Thermo Fisher Scientific (Waltham, MA, USA).

2.2 Tissue collection

Anopheles gambiae specimens and organs were a generous gift from Dr S. Blandin from the University of Strasbourg (France). *An. gambiae* appendages were collected from frozen adults stored at -20°C. Legs and wings were collected from five adults, while antennae (without Johnston's organ) were collected from 10 adults. *B. mori* specimens were a gift from Mr Perret from Soucieu-en-Jarrest (France). Once fifth instar larvae reached day-5, the complete exoskeleton including head and prolegs was dissected by longitudinal incision and stored at -20°C until use. *Apis mellifera* thoracic integuments were collected from pupae of honeybee workers collected in cells from a beehive set up at our laboratory and stored at -20°C until use. Pea Aphid cast cuticles left behind after the 4th instar molted to the adult stage were provided by Dr Uzest of the National Institute of Agronomic Research (Montpellier, France). Proteomic analyses of *B. mori*, *Ap. mellifera*, and *An. gambiae* cuticles were performed in duplicate, while that of *Ac. pisum* was performed in triplicate.

2.3 Sample preparation

All cuticular structures were washed twice with PBS buffer in an ultrasonic bath (Branson, Danbury, CT) for 15 minutes to remove the internal tissues. Cuticles were intact after treatment. Samples were pelleted by a quick centrifugation step, and dried by centrifugation under vacuum (Labconco, Kansas City, MO). Dried samples were incubated in pure HFIP with a sufficient volume to cover the biological matrices for 4h at 4°C under gentle shaking. HFIP was evaporated by centrifugation under vacuum, and samples were incubated

overnight at 4°C in ammonium bicarbonate buffer (50 mM, pH 7.8) supplemented with 0.1% RapiGest™ surfactant. Proteins were reduced with 10 mM DTT in 50 mM NH₄HCO₃ for 1h in the dark, at 50°C, prior to alkylation with 30 mM IAA in 50 mM NH₄HCO₃ for 30 minutes at 50°C. Digestion was carried out overnight at 37°C by addition of 0.5 µg of trypsin prepared in 50 mM NH₄HCO₃. To stop proteolysis and cleave the RapiGest™ surfactant, samples were acidified with TFA to obtain a 0.5% concentration and incubated for 30 minutes at 37°C. Finally, samples were dried by centrifugation under vacuum and dry pellets re-suspended in 2% ACN/0.1% TFA final concentration prior to MS analysis.

2.4 nano-LC-MS/MS analyses

Nano-LC-MS/MS analyses were carried out using an Ultimate 3000 nano-HPLC, coupled with a Q-Exactive Orbitrap high-resolution mass spectrometer (Thermo Fisher Scientific, MA). The digests were loaded onto a C₁₈ PepMap100 precolumn (5 µm, 300 µm x 5 mm) at 10 µL/min, separated onto a Acclaim C₁₈ PepMap100 column (3 µm, 75 µm x 250 mm) at 300 nL/min (Thermo Fisher Scientific) in a linear gradient of water/acetonitrile/0.1% formic acid (v/v) and analyzed by ESI-MS. Peptides were eluted in biphasic linear gradient of 2-32% and of 32-65% acetonitrile (0.1% formic acid) in 100 and 5 min, respectively.

For MS analysis, the Q-Exactive mass spectrometer, equipped with a nanospray ion source, was used in positive mode and data-dependent acquisition. The voltage applied to the nano-tips was adjusted to produce 0.3 µA and the entrance capillary was maintained at 320°C. The Q-Exactive Orbitrap acquired a full-range scan from 380 to 2,000 *m/z* (70,000 resolution, AGC target 3.10⁶, maximum IT 200 ms) and then fragmented the top ten-peptide ions in each cycle (17,500 resolution, AGC target 2.10⁵, maximum IT 100 ms, intensity threshold 4.10⁴, excluding charge-unassigned ions, Normalized Collision Energy selected at 27). Parent ions were then excluded from MS/MS for the next 15s. The software Chromeleon™ Xpress and Xcalibur™ 2.2 were used to control the HPLC and the mass spectrometer, respectively.

2.5 Database searching, CPs identification and presentation

Sequest HT was run by Proteome Discoverer™ 1.4 to match the acquired MS/MS spectra to protein database described below, with the following parameters: trypsin digest with two maximum missed cleavages, six and 144 amino acids as minimum and maximum peptide length, respectively, a tolerance of 10 ppm / 0.02 Da for precursors and fragment ions respectively; carbamidomethylation of cysteine was set as a fixed modification, C-terminal protein amidation, methionine and tryptophan oxidation were set as variable modifications.

To identify CPs from the model organisms we studied, and to compare our results with those previously obtained by other groups, we proceeded as follows: one database was constructed per organism. For *An. gambiae*, the identification was performed against a database “AGamP4.4” (<https://www.vectorbase.org/>), containing 13,021 sequences. CPs were identified using the list of 282 CPs provided by Zhou et al (2016) [33].

For *Ap. mellifera*, the identification of the proteins was carried out against the database “Amel_OGS v3.2” (http://hymenoptera.genome.org/beebase/?q=download_sequences). This base contains the 15,272 predicted proteins of *Ap. mellifera*'s genome. The database used for the identification of *B. mori* proteins was composed of 19,970 sequences, downloaded in April 2017 on NCBI (<https://www.ncbi.nlm.nih.gov/>).

Finally, “ACYPI proteins v2.0”, the database used for the identification of *Ac. pisum* proteins, was downloaded from the AphidBase (bipaa.genouest.org/is/aphidbase/). This database contains 31,908 sequences.

Detection of CPs in the protein databases of *Ap. mellifera* and *Ac. pisum*, were performed using the CutProtFam-Pred tool [5], allowing us to identify structural CPs, corresponding to eight CP families out of 13 and to assign them to specific families. In order to be able to make the direct comparison between the data produced by our methodology, with data published in the literature, the CPs handling and presentation is achieved in different ways, depending on each insect studied. In summary, identifications of the CPs from the cuticles of *Ap. mellifera* were confirmed when at least one unique peptide was characterized in each

replicate. Identifications of CPs from *Ac. pisum* cast cuticle were confirmed by a unique peptide and/or a shared peptide in each replicate and for *B. mori* and *An. gambiae*, the identified CPs were compiled from any replicate and confirmed by the presence of a unique peptide for *B. mori* and by a unique peptide and/or a shared peptide for *An. gambiae*.

3. Results and Discussion

3.1 Proteins identified in different cuticular structures

The improved experimental procedure described in the present work successfully lead to the identification of proteins in the different cuticular structures (Table 1). Among the three biological models considered, *An. gambiae* was the insect in which the highest number of proteins in the appendages was identified (1,151 to 1,623 distinct proteins identified in legs, antennae and wings). The percentage of identified CPs compared to the total number of identified proteins was the lowest in these structures, ranging from 4.93% in wings to 7.52% in antennae. These results are consistent with the general composition of these appendages, which contain not only cuticle but also additional tissues such as muscles and hemolymph rich in proteins other than CPs. The highest percentages of CPs were obtained for *Ac. pisum* molts (25.58%), for the complete exoskeleton of *B. mori* (25.91%) and for the thoracic segment of *Ap. mellifera* pupae (29.17%).

3.2 CPs identified

85 CPs have been identified in *B. mori* exoskeleton (Figure 1), including 56 CPR (29 RR-1, 25 RR-2 and two RR-3), five CPAP3, three CPFL, six CPG, 12 CPH (Cuticular Protein Hypothetical) and three TWDL. All CPs have been identified with at least one unique peptide (Supplementary data 1). 28 CPs have been identified in the thoracic integument of *Ap. mellifera* pupae (Figure 1), among which 22 are defined as CPs by the CutProtFam-Pred tools, 16 CPR (eight RR-1 and eight RR-2), two CPAP3, two CPF, and two TWDL. The six others proteins have already been described as CPs [34], two proteins that are not recognized by the CutProtFam-Pred tool, GB53110 (apidermins-3) and GB53119 (apidermin-2). For the other proteins (GB48841, GB48844, GB45151, GB48842), no

information about their family is available. As with *B. mori*, all CPs have been identified through the characterization of at least one unique peptide (Supplementary data 2).

Regarding the different *An. gambiae*'s appendages, between 67 and 105 CPs were identified (Figure 1 and Table 1). 105 distinct CPs have been identified in antennae, 80 in wings, and 67 in legs. Several CPs were exclusively identified in one appendage, with the largest number in antennae (Supplementary data 3).

3.3 Comparison with published data on model organisms (*B. mori*, *Ap. mellifera*, *An. gambiae*).

Previous works on the exoskeleton of *B. mori*, allowed the authors to identify 103 CPs [31] and 23 CPs [36]. Using our experimental protocol and identification procedure, only 85 CPs could be identified (Supplementary data 1) from a single larval exoskeleton per replicate (see Table 2). However, despite the number of diverging parameters between the published data [31] and those used with our analytical procedure, we were able to identify 61 CPs of the 103 published CPs [31], corresponding to 59%, using a time-saving and straightforward methodology that avoids the use of SDS and of a purification step with affinity beads.

Regarding the work published by Fu *et al.* [36] we identified in our results 12 CPs out of the 23 previously published, corresponding to 52% (supplementary data 1) but 73 additional CPs were identified without the use of SDS and followed or not by a electrophoresis step. Moreover, the fact we have generated proteomic data at the individual scale, gives an interesting asset to this method to decipher CPs with a minimum amount of material and a comfortable analytical resolution.

A proteomic study of the thoracic cuticle of *Ap. mellifera* was previously performed by 2D electrophoresis coupled with MALDI-TOF MS [34] where a differential expression of CPs was reported at different developmental stages. Nine CPs were described in the thoracic cuticle of the honeybee pupae while 14 distinct CPs were reported from all of the investigated development stages. Using the new developed protocol, we could increase the number of CPs detected up to 28 CPs, all being identified with a unique peptide (Table 2).

Among these CPs, seven were previously identified by Micas *et al.* in pupae. We also identified four other proteins that had not been identified at this development stage but at others: one previously described in L5 Larval instar, two in pharate adults and one that had been identified at all developmental stages except in the pupae [34] (See Supplementary data 2). In view of these results we can say that our experimental strategy allowed us to establish a far more exhaustive list of CPs than the strategy used by Micas *et al.* [34], which only analyzed spots from 2D gels. In addition, our CP list was obtained starting with just a single individual whereas Micas *et al.* carried out their analysis from a pool of 15 individuals. The proteomic-based identification of *An. gambiae*'s CPs, particularly within antennae, legs and wings has markedly improved [33] (Table 2). In order to compare these CPs with the ones obtained using HFIP and RapiGest™ surfactant, we applied the same method for database searching, data handling and presentation with minor changes in the database version of AgamP from version 4.2 to 4.4 (Supplementary data 3). We have found 80 CPs in wings (Table 2 and Supplementary data 4), of which 57 were reported previously [33]. Over the previously reported CPs only six were not identified in this work. Interestingly, we were able to identify 23 additional proteins, not reported in this previous work [33]. In antennae, we have identified 105 CPs (Supplementary data 5) *versus* 94 as previously reported [33]. The CPLCG25 identified by Zhou *et al.* (2016) was not taken into account because this protein was not present in the database AgamP 4.4. Of the 105 CPs identified, 74 were previously reported by Zhou *et al.* [33]. Interestingly while 20 CPs were not detected with our method, 31 additional CPs were identified for the first time from *An. gambiae*'s antennae. For legs, 67 CPs (Supplementary data 6) were identified with our method *vs* 61 in the previous works. Zhou's group [33] has identified 12 CPs that we did not identify in our analyses and conversely, we have identified 18 CPs that were not described by Zhou and colleagues. In terms of biochemistry of the CPs, these two approaches did not reveal a group or a family of CPs preferentially identified.

The complication of identifying CPs by MS, because of peptides shared among proteins, results in several CP annotations based on shared tryptic peptides. This has been previously discussed [33]. We therefore investigated the ratio of the identified CPs with at least one unique characterized peptide, over the number of total CPs identified exclusively with shared peptides (Supplementary data 3). Ratios were similar between our methodology and the data previously published [33]. Starting with a quantity of biological material at least ten times lower for each appendage, we demonstrated that the methodology we developed and presented in these proteomic studies allowed to increase our knowledge on the diversity of CPs that are involved in the different cuticular structures of *An. gambiae* (Table 2).

3.4 CPs identification in the pea aphid *Acyrtosiphon pisum*.

After evaluating the effectiveness of our simplified analytical procedure for CPs identification on different insect species, and several types of cuticle structures, we explored the CPs of the pea aphid *Ac. pisum*. In this insect, no extensive proteomic studies on cuticle structures have been reported yet. To evaluate the relevance of this new proteomics approach on totally unexplored cuticle, we applied the protocol on a pool of five pea aphid cast cuticles (whole exoskeleton containing antennae, legs and stylets bundle). The CPs presented in Figure 1, Table 1 and in Supplementary data 7 and 8 have all been identified in triplicate.

The CutProtFam-Pred tool allowed us to identify 150 CPs in the database “ACYPI Proteins V2.0”, corresponding to 147 distinct genes. Among them, 15 are belonging to CPR_RR-1, 108 to CPR_RR-2; 10 to CPAP1, eight to CPAP3, three to TWDL, two to CPF, and one to the CPCFC. In our study, 100 CPs were identified, including 10 RR-1, 72 RR-2, six CPAP1, eight CPAP3, two TWDL and two CPF (Figure 1), thus belonging to almost all CPs families expressed in the pea aphid. The only characterized cuticular protein family for which no member has been identified in this study is the CPCFC family. Our methodology allowed us to identify 60% of CPAP1, 66.66% of CPR_RR-1, CPR_RR-2 and TWDL, and 100% of CPAP3 and CPF, and a total of 68% of the putative CPs.

Among the 100 CPs found in *Ac. pisum* cast cuticles, only 52 were identified with a unique peptide. The percentage of identified proteins with at least one unique peptide differed among the different CP families, reaching 100% for CPAP1, CPAP3 and for TWDL. The two CPFs were identified with the same peptide. Finally, for the CPRs, 60% of the RR-1 and 41.66% of RR-2 sub-families were identified with a unique peptide. As evidenced by Zhou's group [33] for *An. gambiae*, it is clearly more difficult to discriminate the different CPRs and CPFs within their respective families, and to identify proteins unambiguously within these families because of their highly conserved amino acid sequences, especially for the CPR at their chitin-binding domain.

3.5 CPR expression in *Ac. pisum*.

CPRs represent 83.7% of the total CPs identified by the CutProtFam-Pred tool. We identified 66% of them in *Ac. pisum* cast cuticles with unique and shared peptides. Interestingly, through our strategy and for each biological replicate, when the proteins are ranked by their score of identification, the 25 first identified CPs were all members of the CPR family (See supplementary data 8). We can hypothesize either these proteins are more easily detected under our experimental conditions, or they are more abundant in the cast cuticles. Another hypothesis we could consider is that, as the CPR genes have been found in the insect genomes in much higher numbers compared to other CP gene categories, we could posit that the protein population resulting from these genes would show this abundance: while CPR proteins may not necessarily be more abundant individually, together they could represent a major part of the present proteins. This would result in these CPR proteins being statistically more likely to be extracted and identified by the LC-MS/MS analysis.

3.6 CPAP and Chitin-binding domain in *Ac. pisum*

The second most abundant CP families identified in *Ac. pisum*, are the CPAP1 and CPAP3 families with ten and eight proteins characterized in the pea aphid genome, respectively. These families are characterized by their chitin-binding domain (ChBD2), also known as Peritrophin-A chitin-binding domain, containing six cysteines [10]. CPAP1 have a single

ChBD2 domain, while CPAP3 have three. In this study, six CPAP1 and eight CPAP3 were identified in this cuticular structure, corresponding respectively to 60% and 100% of the total proteins within these families.

In 2015, CPAP1 and CPAP3 proteins were reported to have originated from a common ancestral gene that spread to different insect orders through evolution [6]. The same authors showed that proteins possessing ChBD2 domains, such as CPAP1 and CPAP3, peritrophic membrane proteins (PMP) or chitinase, could actually be classified into different groups, depending on the sequence of their ChBD2 domains. Based on this classification, a study of the *Manduca sexta* CPs [38] highlighted by orthology the proteins belonging to these different groups in other insects, including *Ac. pisum*. We identified in *Ac. pisum* cast cuticles 6 CPAP1 proteins and 8 CPAP3 proteins. Most of them, but not all, were already described in the 2015 study by Dittmer et al. [38] as homologues to CPAP of *Manduca sexta*; the two exceptions being the protein ACYPI45536 for the CPAP1, and ACYPI6835 for CPAP3. The proteins we found in *Ac. pisum* belong to the groups CPAP1-A, CPAP1-H, CPAP1-J, CPAP1-K and CPAP1-L, and CPAP3-A1, CPAP3-A2, CPAP3-B, CPAP3-C, CPAP3-D1, CPAP3-D2 and CPAP3-E according to Dittmer et al [38].

Other proteins possessing ChBD2 domain not recognized as members of the CPAP1 or CPAP3 families by the CutProtFam-Pred tool (See supplementary data 10), were still identified in the molt of *Ac. pisum*. According to the classification of proteins established for *Tribolium castaneum* [10], these proteins possessing 1, 3, 4 or 6 ChBD2 domains belong to the PMP family, usually located in the digestive tract. In order to confirm that these proteins possessing one or three ChBD2 domains are not CPs, we examined them against phylogenetic analysis as described [6]. These domains did not fit the phylogenetic model, hence confirming that they may not be part of the CPAP1 or CPAP3 families.

The protein ACYPI003231-PA, which harbored a chitin-binding domain of type 4, was also identified in this study. This domain is not surprising to find in a cast cuticle as the proteins could come from a remnant product of the molting process or molting fluid itself [39].

3.7 Others CP families identified in *Ac. pisum*.

Only five CP families defined by CutProtFam-Pred were found to be present in *Ac. pisum* genome [5]. CPs from four of these families were identified in our study. Two CPs belonging to the TWDL family have been detected in the molt, ACYPI001451-PA and ACYPI006519-PA. Previous investigations have also shown that TWDL proteins were involved in the molting process in *Locusta migratoria* [22]. The identification of Tweedle proteins in the pea aphid cast cuticle is not aberrant, as already mentioned above for ACYPI003231-PA. Finally, two CPFs were identified in *Ac. pisum* molt: ACYPI002788-PA and ACYPI010088-PA. The CPF and CPFL proteins were described to be involved in development and molting process [12]. Therefore, as with the two aforementioned families, it is not surprising to have identified this family in the *Ac. pisum* sample.

4 Concluding remarks

The use of a low polarity highly fluorinated compound such as the fluorinated alcohol hexafluoroisopropanol (HFIP) is offering a new interesting solvent to improve peptide and protein extraction under fully compatible MS analyses. Even if the entire mechanistic of the effect of HFIP on peptides and proteins remains to be fully elucidated [40], we are confident that the strategy to introduce HFIP in proteomics approaches represents a potent alternative to modify the structure of peptides and proteins and to minimize the usage of SDS, the most widely used amphiphilic detergents. We did not see any evidence of limitation to the use of HFIP in the process of proteomics as this chemical is fully compatible with MALDI-MS and LC-MS/MS. The new developed procedure, which is also including the use of *RapiGest*TM SF surfactant in solution rather than in-gel digestion, allows one to limit the number of tedious steps of pre-fractionation (1D- or 2D-SDS-PAGE, affinity chromatography or bead enrichment) and to avoid any selective procedure resulting in maximizing the recovery of peptide and proteins from complex biological matrices. As all our procedure of treatment is performed within a same container, it is also possible to work on very small quantities of tissues and organs. In addition to be time and cost effective, our procedure minimized the

possibility to lose materials during tube transfer and the complex manipulations required for 1D- or 2D-SDS-PAGE of affinity chromatography before bottom-up or top-down proteomics analyses. To conclude, we have developed a new proteomics approach to investigate the protein composition of cuticular structures by high-resolution mass spectrometry. This novel approach, compared to the other methods discussed in this comparative work, significantly reduces the number of sample processing steps. We have shown that 68% of the *Acyrtosiphon pisum*'s annotated putative CPs were identified from cast cuticle collected after one molt. Most importantly, this approach can be applied to lower amounts of cuticular materials, and is therefore promising for all samples for which large quantities are not always possible to collect. The number of CPs identified with our methodology showed that it was possible to identify more CPs in small appendages, as well as in larger cuticular structures such as the complete exoskeleton of *B. mori*, rather than using more conventional sample preparations (SDS extraction and multiple purification steps). CPs identified in the different insects studied in this work have been assigned to a large number of CP families, showing that our analytical methodology was not restricted to peculiar CP families, and could be applied to exhaustively study CPs. Finally, we have generated the first proteomic repertoire within an aphid cuticle.

Acknowledgements: We thank Dr Uzest and Mr Cayrol from INRA (Montpellier, France) for providing *Acyrtosiphon pisum*. We are grateful to Dr. Blandin and Mr. Fraiture from the University of Strasbourg (France) for *Anopheles gambiae*, and to M. Perret for *Bombyx mori* larvae (Soucieu-en-Jarrest, France). We would like to thank Mr. Bocquet (Apimedia, France) for providing us *Apis mellifera* pupae. We are also grateful to Dr Uzest for comments, suggestions and advice and Miss Blangarin for critical reading of the manuscript. Finally, We would like to acknowledge the Silk Museum from St Hippolyte du Fort (France) for providing a *Bombyx mori* picture free of copyright. This work was supported by the Bill & Melinda Gates Foundation, Seattle, WA (contract ID: OPP1130147).

Conflict of Interest: The authors have declared no conflict of interest.

This article is protected by copyright. All rights reserved.

References

- [1] Vincent, J.F., Wegst, U.G., Design and mechanical properties of insect cuticle, *Arthropod Struct. Dev.* 2004, 33, 187-199.
- [2] Moussian, B., Molecular Model of Skeletal Organization and Differentiation, in: E. Cohen, B. Moussian, (Eds.), *Extracellular Composite Matrices in Arthropods*. Springer International Publishing Switzerland 2016, pp 67-87.
- [3] Willis, J.H., Structural CPs from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem. Mol. Biol.* 2010, 40, 189-204.
- [4] Willis, J.H., Papandreou, N.C., Iconomidou, V.A., Hamodrakas, S.J., Cuticular proteins in: Gilbert, L.I. (Ed.), *Insect Mol. Biol. Biochem.* 2012. Academic Press, London, Waltham & San Diego, pp. 134-166.
- [5] Ioannidou, Z.S., Theodoropoulou, M.C., Papandreou, N.C., Willis, J.H., Hamodrakas, S.J., CutProtFam-Pred: detection and classification of putative structural CPs from sequence alone, based on profile hidden Markov models. *Insect Biochem. Mol. Biol.* 2014, 52, 51-59.
- [6] Tetreau, G., Dittmer, N.T., Cao, X., Agrawal, S., Chen, Y.R., Muthukrishnan, S., Haobo, J., Blissard, G.W., Kanost, M.R., Wang, P., Analysis of chitin-binding proteins from *Manduca sexta* provides new insights into evolution of peritrophin A-type chitin-binding domains in insects, *Insect Biochem. Mol. Biol.* 2015, 62, 127-141.
- [7] Vannini, L., Bowen, J.H., Reed, T.W., Willis, J.H., The CPCFC cuticular protein family: anatomical and cuticular locations in *Anopheles gambiae* and distribution throughout *Pancrustacea*. *Insect Biochem. Mol. Biol.* 2015, 65, 57-67.
- [8] Andersen, S.O., Hojrup, P., Roepstorff, P., Insect Cuticular Proteins, *Insect Biochem. Mol. Biol.* 1995, 25, 153-176.
- [9] Willis, J.H., Iconomidou, V.A., Smith, R.F., Hamodrakas, S.J., Cuticular Proteins, in: Gilbert, L.I. Iatrou, K., Gill, S.S., (Eds.), *Comprehensive molecular insect sciences*. Vol. 4. Elsevier B.V., Amsterdam; New York, 2005, pp. 79-110.

- [10] Jasrapuria, S., Arakane, Y., Osman, G., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped into three distinct families based on phylogeny, expression and function. *Insect Biochem. Mol. Biol.* 2010, 40, 214-227.
- [11] Andersen, S.O., Rafin, K.P., Roepstorff, P., Sequence studies of proteins from larval and pupal cuticle of the yellow meal worm, *Tenebrio molitor*, *Insect Biochem. Mol. Biol.* 1997, 27, 121-131.
- [12] Togawa, T., Dunn, W.A., Emmons, A.C., Willis, J.H., CPF and CPFL, two related gene families encoding CPs of *Anopheles gambiae* and other insects. *Insect Biochem. Mol. Biol.* 2007, 37, 675-688.
- [13] Guan, X., Middlebrooks, B.W., Alexander, S., Wassermann, S.A., Mutation of TweedleD, a member of tan unconventional cuticle protein family, alters body shape in *Drosophila*. *Proc. Natl. Acad. Sci. USA.* 2006, 103, 16794-16799.
- [14] Cornman, R.S., Willis, J.H., Annotation and analysis of low-complexity protein families of *Anopheles gambiae* that are associated with cuticle, *Insect Mol. Biol.* 2009, 18, 607-622.
- [15] Futahashi, R., Okamoto, S., Kawasaki, H., Zhong, Y.S., Iwanaga, M., Mita, K., Fujiwara, H., Genome-wide identification of cuticular protein genes in the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* 2008, 38, 1138-1146.
- [16] Kucharski, R., Maleszka, J., Maleszka, R., Novel cuticular proteins revealed by the honey bee genome. *Insect Biochem. Mol. Biol.* 2007, 37, 128-134.
- [17] He, N., Bothelho, J.M.C., McNall, R.J., Belozarov, V., Dunn, W.A., Mize, T., Orlando, R., Willis, J.H., Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass spectrometry, *Insect Biochem. Mol. Biol.* 2007, 37, 135-146.
- [18] Noh, M.Y., Kramer, K.J., Muthukrishnan, S., Kanost, M.R., Two major CPs are required for assembly of horizontal laminae and vertical pore canals in rigid cuticle of *Tribolium castaneum*. *Insect Biochem. Mol. Biol.* 2014, 53, 22-29.

- [19] Noh, M.Y., Muthukrishnan, S., Kramer, K.J., Arakane, Y., *Tribolium castaneum* RR-1 cuticular protein TcCPR4 is required for formation of pore canals in rigid cuticle, *PLoS Genet.* 2015, 11:e1004963.
- [20] Tajiri, R., Ogawa, N., Fujiwara, H., Kojima, T., Mechanical control of whole body shape by a unique cuticular protein obstructor-E in *Drosophila melanogaster*, *PLoS Genet.* 2017, 13(1):e1006548.
- [21] Mun, S., Noh, M.Y., Dittmer, N.T., Muthukrishnan, S., Cuticular protein with a low complexity sequence becomes cross-linked during insect cuticle sclerotization and is required for the adult molt. *Sci. Rep.* 2015, 5:10484.
- [22] Song, T.Q., Yang, M.L., Wang, Y.L., Liu, Q., Cuticular protein LmTwdl1 is involved in molt development of the migratory locust, *Insect Sci.* 2016, 23, 520-530.
- [23] Wu, F., Wang, P., Zhao, Q., Kang, L., Mutation of a Cuticle Protein Gene, BmCPG10, Is Responsible for Silkworm Non-Moulting in the 2nd Instar Mutant, *PLoS One* 2016, 11(4):e0153549.
- [24] Vontas, J., David, J.P., Nikou, D., Hemingway, J., Transcriptional analysis of insecticide resistance in *Anopheles stephensi* using cross-species microarray hybridization. *Insect Mol. Biol.* 2007, 40, 189-204.
- [25] Awolola, T.S., Oduola, O.A., Strode, C., Koekemoer, L.L., Evidence of multiple pyrethroid resistance mechanisms in the malaria vector *Anopheles gambiae sensu stricto* from Nigeria. *Trans. R. Soc. Trop. Med. Hyg.* 2009, 103, 1139-1145.
- [26] Højland, D.H., Kristensen, M., Analysis of Differentially Expressed Genes Related to Resistance in Spinosad- and Neonicotinoid-Resistant *Musca domestica* L. (Diptera: Muscidae) Strains. *PLoS One* 2017, 12(1):e0170935.
- [27] Liang, J., Wang, T., Xiang, Z., He, N., TWDL cuticular protein BmCPT1 is involved in innate immunity by participating in recognition of *Escherichia coli*. *Insect Biochem. Mol. Biol.* 2015, 58, 76-88.

[28] Brault, V., Uzest, M., Monsion, B., Jacquot, E., Blanc, S., Aphids as transport devices for plant viruses. *C. R. Biol.* 2010, 333, 524-538.

[29] Liu, W., Gray, S., Huo, Y., Li, L., Proteomic Analysis of Interaction between a Plant Virus and Its Vector Insect Reveals New Functions of Hemipteran Cuticular Protein, *Mol. Cell Proteomics* 2015, 14, 2229-2242.

[30] Magkrioti, C.K., Spyropoulos, I.C., Iconomidou, V.A., Willis, J.H., Hamodrakas, S.J., CuticleDB: a relational database of Arthropod CPs, *BMC Bioinformatics* 2004, 5.

[31] Dong, Z., Zhang, W., Zhang, Y., Zhang, X., Identification and Characterization of Novel Chitin-Binding Proteins from the Larval Cuticle of Silkworm, *Bombyx mori*, *J. Proteome Res.* 2016, 15, 1435-1445.

[32] Zhou, Y., Badgett, M.J., Billard, L., Bowen, J.H., Properties of the CPs of *Anopheles gambiae* as revealed by serial extraction of adults, *PLoS One* 2017, 12(4):e0175423.

[33] Zhou, Y., Badgett, M.J., Bowen, J.H., Vannini, L., Distribution of CPs in different structures of adult *Anopheles gambiae*, *Insect Biochem. Mol. Biol.* 2016, 75, 45-57.

[34] Micas, A.F., Ferreira, G.A., Laure, H.J., Rosa, J.C., Bitondi, M.M., Proteins of the integumentary system of the honeybee, *Apis mellifera*, *Arch. Insect Biochem. Physiol.* 2016, 93, 3-24.

[35] Tang, L., Liang, J., Zhan, Z., Xiang, Z., He, N., Identification of the chitin-binding proteins from the larval proteins of silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* 2010, 40, 228-34

[36] Fu, Q., Li, P., Xu, Y., Zhang, S., Jia, L., Zha, X., Xiang, Z., He, N., Proteomic analysis of larval integument, trachea and adult scale from the silkworm, *Bombyx mori*. *Proteomics.* 2011, 18, 3761-3767.

[37] Dittmer, N.T., Hiromasa, Y., Tomich, J.M., Lu, N., Beeman, R.W., Kramer, K.J., Kanost, M.R., Proteomic and transcriptomic analyses of rigid and membranous cuticles and epidermis from the elytra and hindwings of the red flour beetle, *Tribolium castaneum*. *J Proteome Res.* 2012, 1, 269-78.

[38] Dittmer, N.T., Tetreau, G., Cao, X., Jiang, H., Wang, P., Kanost, M.R., Annotation and expression analysis of cuticular proteins from the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol.* 2015, 62, 100-113.

[39] Forsberg, Z., Mackenzie, A.K., Sørli, M., Røhr, Å.K., Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases, *Proc. Natl. Acad. Sci. USA* 2014, 111, 8446-8451.

[40] Hong, D.P., Masaru Hoshino, M., Kuboi, R., and Goto, Y., Clustering of Fluorine-Substituted Alcohols as a Factor Responsible for Their Marked Effects on Proteins and Peptides. *J. Am. Chem. Soc.* 1999, 121, 8427-8433

Table 1: Summary of identified CPs as fraction of all proteins recovered.

Organisms	Cuticular structures	# Total proteins	# CPs recovered	% CPs
<i>B. mori</i>	Complete exoskeleton	328	85	25.91
<i>Ap. mellifera</i>	Thoracic integument of pupae	96	28	29.17
<i>An. gambiae</i>	Wings	1,623	80	4.93
	Antennae	1,397	105	7.52
	Legs	1,151	67	5.82
<i>Ac. pisum</i>	Cast cuticle	391	100	25.58

Table 2: Summary of CPs identified in published data and methodology used to perform these analyses.

Species	Cuticular structures	# of specimens	Sample preparation			MS analysis	# identified CPs	Ref.
			Extraction	Protein fractionation	Reduction-alkylation & Trypsin digestion			
<i>B. mori</i>	Complete exoskeleton	Undefined	SDS 2%	With Affinity Beads	In ultrafiltration tube digestion	LC-ESI-MS/MS	103	[31]
				Without Affinity Beads				
			SDS 2%	With SDS-PAGE	In-gel digestion	LC-ESI-MS/MS	23	[36]
				Without SDS-PAGE	Direct digestion of the final pellets			
<i>Ap. mellifera</i>	Thoracic integument of pupae	15	Urea 7.7 M Thiourea 2.2 M / CHAPS 4.4%	2D SDS-PAGE	In-gel digestion	MALDI-TOF MS/MS	9	[34]
<i>An. gambiae</i>	Legs	80	SDS 1% (2016)	With SDS-PAGE	In-gel digestion	LC-ESI-MS/MS	61	[33]
	Antennae	220		Without SDS-PAGE	Direct digestion of the final pellets		95 (with CPLCG25)	
	Wings	100					63	
<i>B. mori</i>	Complete exoskeleton	1	<i>Rapigest</i> TM 0.1% Pure HFIP	None	In-solution digestion	LC-ESI-MS/MS	85	Present work
<i>Ap. mellifera</i>	Thoracic integument of pupae	1					28	

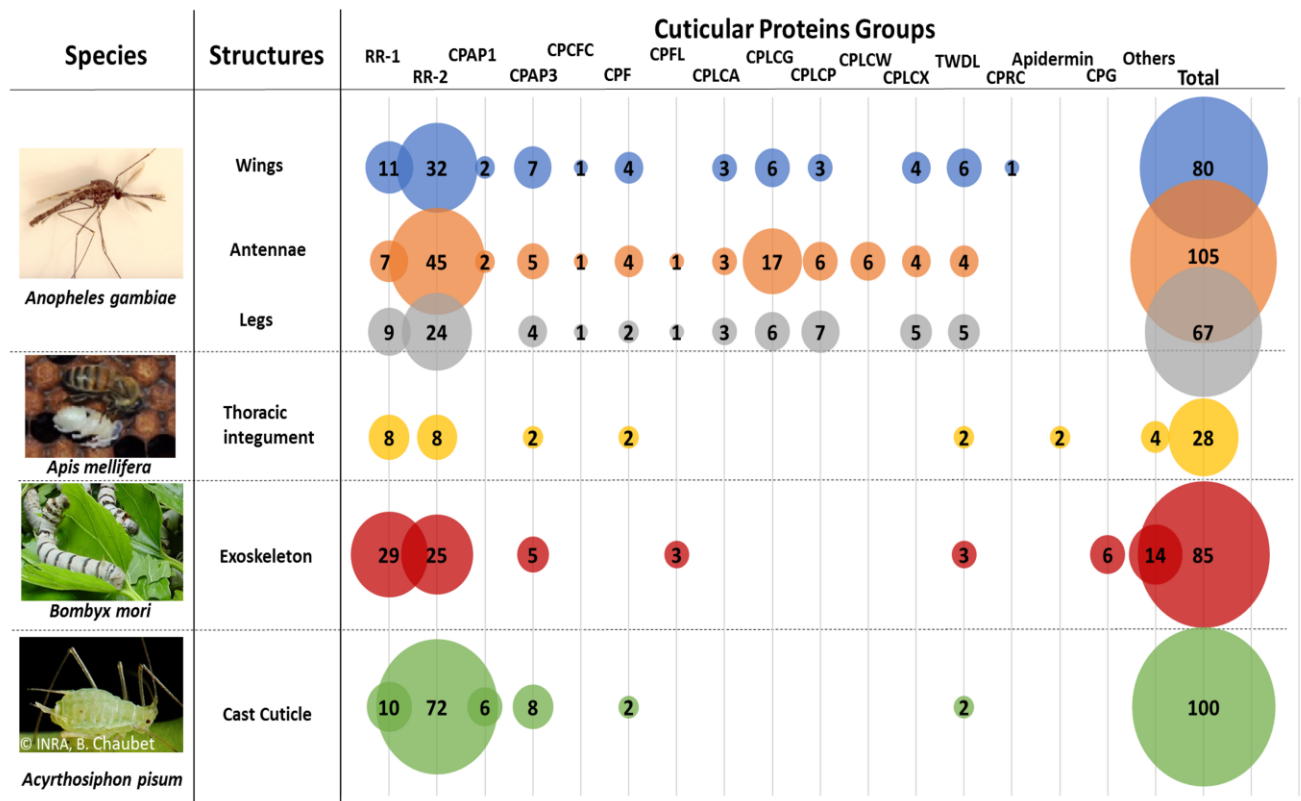
Received: 10 03, 2017; Revised: 12 21, 2017; Accepted: 01 05, 2018

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/pmic.201700368](https://doi.org/10.1002/pmic.201700368).

This article is protected by copyright. All rights reserved.

An. gambiae	Legs	5					67	
	Antennae	10					105	
	Wings	10					80	
Ac. pisum	Cast cuticle	5					100	

Figure 1: Inventory count of CPs families in cuticular structures characterized in *An. Gambiae*, *Ap. mellifera*, *B. mori* and *Ac. Pisum*. CPs families for each CPs identified were assigned with the CutProtFam-Pred tool



Received: 10 03, 2017; Revised: 12 21, 2017; Accepted: 01 05, 2018

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/pmic.201700368](https://doi.org/10.1002/pmic.201700368).

This article is protected by copyright. All rights reserved.