Matthias Van Vaerenbergh

Honeybee (*Apis mellifera*) and Bumblebee (*Bombus terrestris*) Venom: Analysis and Immunological Importance of the Proteome

Thesis submitted to obtain the academic degree of
Doctor in Science: Biochemistry and Biotechnology

Ghent University

Department of Physiology (WE15)

Laboratory of Zoophysiology

Honeybee (*Apis mellifera*) and bumblebee (*Bombus terrestris*) venom: analysis and immunological importance of the proteome

Het gif van de honingbij (*Apis mellifera*) en de aardhommel (*Bombus terrestris*): analyse en immunologisch belang van het proteoom

Matthias Van Vaerenbergh

Ghent University, 2013

Supervisors:

Promotor: **Prof. Dr. Dirk C. de Graaf**
Laboratory of Zoophysiology
Department of Physiology
Faculty of Sciences
Ghent University

Co-promotor: **Prof. Dr. Bart Devreese**
Laboratory for Protein Biochemistry and Biomolecular Engineering
Department of Biochemistry and Microbiology
Faculty of Sciences
Ghent University

Reading Committee:

Prof. Dr. Geert Baggerman (University of Antwerp)
Dr. Simon Blank (University of Hamburg)
Prof. Dr. Bart Braeckman (Ghent University)
Prof. Dr. Didier Ebo (University of Antwerp)

Examination Committee:

Prof. Dr. Johan Grooten (Ghent University, chairman)
Prof. Dr. Dirk C. de Graaf (Ghent University, promotor)
Prof. Dr. Bart Devreese (Ghent University, co-promotor)
Prof. Dr. Geert Baggerman (University of Antwerp)
Dr. Simon Blank (University of Hamburg)
Prof. Dr. Bart Braeckman (Ghent University)
Prof. Dr. Didier Ebo (University of Antwerp)
Dr. Maarten Aerts (Ghent University)
Prof. Dr. Guy Smagghe (Ghent University)

Dean: Prof. Dr. Herwig Dejonghe

Rector: Prof. Dr. Anne De Paepe

Please refer to this work as:

Van Vaerenbergh, M. (2013) Honeybee (*Apis mellifera*) and bumblebee (*Bombus terrestris*) venom: analysis and immunological importance of the proteome. PhD thesis, Ghent University.

Author's email address:

matthiasvanvaerenbergh@hotmail.com

Front cover

Photographs

© http://www.abeezhoney.net/wp-content/uploads/2011/12/HoneyStock1.jpg

© http://commons.wikimedia.org/wiki/File:Bumblebee_2007-04-19.jpg

Design: Quinten De Corte

Curriculum vitae

PERSONAL DETAILS

Name	Matthias Van Vaerenbergh
Date of birth	19/04/1986
Nationality	Belgian
Telephone	0498/679827
E-mail	matthiasvanvaerenbergh@hotmail.com

EDUCATION

2009-2013	Ph.D. candidate, Biochemistry and Biotechnology, Ghent University (Grant: IWT-Flanders)
	PhD thesis: Honeybee (*Apis mellifera*) and bumblebee (*Bombus terrestris*) venom: analysis and immunological importance of the proteome.
2007-2009	Master in Science (MSc), Biology, Ghent University (magna cum laude)
	Master thesis: Venom of honeybees (*Apis mellifera*): immunological importance of new venom components.
2004-2007	Bachelor in Science (BSc), Biology, Ghent University (cum laude)
	Bachelor thesis: Study of zebrafish (*Danio rerio*) mutants with an aberrant tooth phenotype.
1998-2004	Secondary school: Science-mathematics, KTA "De Rijdtmeersen", Brakel

SCHOLARSHIPS AND AWARDS

Ph.D. scholarship of the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT-Flanders), Belgium, 01/01/2010 - 31/12/2013

INTERNSHIPS

28/01/2012 - 10/03/2012: Institute of Biochemistry and Molecular Biology, Hamburg University, Germany

SCIENTIFIC OUTPUT

A1 peer reviewed publications

1. Cardoen D., Ernst U. R., <u>Van Vaerenbergh M.</u>, Boerjan B., de Graaf D. C., Wenseleers T., Schoofs, L. and Verleyen, P. (2011) Differential proteomics in dequeened honeybee colonies reveals lower viral load in hemolymph of fertile worker bees. PLoS ONE 6(6): e20043.
2. <u>Van Vaerenbergh M.,</u> 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. (2013) Extending the honey bee venome with the antimicrobial peptide apidaecin and a protein resembling wasp antigen 5. Insect Molecular Biology 22(2): 199–210.
3. <u>Van Vaerenbergh M.,</u> Debyser G., Devreese B., de Graaf D. C. Exploring the hidden honeybee (*Apis mellifera*) venom proteome by integrating a combinatorial peptide ligand library approach with FTMS. Journal of Proteomics: in press.
4. <u>Van Vaerenbergh M.,</u> Debyser G., Smagghe G., Devreese B., de Graaf D. C. Unraveling the venom proteome of the bumblebee (*Bombus terrestris*) by integrating a combinatorial peptide ligand library approach with FT-ICR MS. Toxicon: in press.

Participation at international conferences

1. BEEDOC Workshop: Diagnostics in honeybees: from sampling to data analyses, Ghent, Belgium, August 2010.
   Van Vaerenbergh M., Jacobs F. J., Devreese B., de Graaf D. C. Molecular cloning reveals the existence of multiple alternative splice variants of hyaluronidase and icarapin in honeybee venom glands. p. 22. (poster)
2. 4th International Symposium on Molecular Allergology, Munich, Germany, October 2010, p. 26.
   Van Vaerenbergh M., Devreese B., de Graaf D. C. Exploring the honeybee (*Apis mellifera*) venom transcriptome using a PCR-based cloning approach. (poster)
3. 17th Benelux Congress of Zoology: Classic Biology in Modern Times, Ghent, Belgium, October 2010, p. 156.
   Van Vaerenbergh M., Devreese B., de Graaf D. C. Identification of novel isoforms of PVF-1 and icarapin in honeybee venom glands using an RT-PCR-based approach. (poster)
4. Bee–together: conference on pollinators with emphasis to stimulate interactions within the field, Ghent, Belgium, December 2010, p. 17.
   Van Vaerenbergh M., Devreese B., de Graaf D. C. Mass spectrometry as a powerful tool for Hymenoptera venom analysis. (oral presentation)
5. 17th Congress of the European Section of the International Society on Toxinology. Valencia, Spain, September 2011, p. 235.
   Van Vaerenbergh M. An in-depth analysis of the honeybee (*Apis mellifera*) and bumblebee (*Bombus terrestris*) venom proteome. (poster)
6. EAACI-WAO World Allergy & Asthma Congress. Milan, Italy, 22-26 June 2013, p. 44.
   Van Vaerenbergh M., De Smet L., Blank S., Spillner E., Ebo D., Jakob T., Devreese B., de Graaf D.C. IgE recognition of multiple novel Api m 10 isoforms evaluated by protein array technology. (oral presentation)

GUIDANCE OF STUDENTS

Educational support (practicals, exercises, seminars)

First Bachelor of Medicine and Dentistry: Aspects of peripheral blood research: 2009-2010, 2010-2011

Second Bachelor of Biochemistry and Biotechnology: Practical exercises of biomedical physiology: 2009-2010, 2010-2011, 2012-2013

Thesis support

Master thesis: The venom of honeybees (*Apis mellifera*): immunological importance of the venom compounds icarapin and PVF1. (Kim De Crem, 2009-2010)

Master project: Recombinant expression of honeybee venom compounds in an insect cell expression system. (Nele Van Der Steen, 2010-2011)

Bachelor thesis: Determination of the serum IgG4-titers against honeybee venom and Api m 10 variants. (Sarah Delbaere, Stef Deyaert, Lieven Lancksweerdt, 2012-2013)

Dankwoord

Ongelooflijk, de tijd vliegt! Meer dan vier jaar geleden stond ik vol zenuwen dit doctoraatsproject te verdedigen voor de IWT-jury. Met succes, zo bleek achteraf. Nu, met een afgewerkt proefschrift voor mij, kan ik met grote voldoening terugkijken op de fantastische ervaring die dit doctoraatsproject is geweest. Natuurlijk zou dit niet gerealiseerd zijn zonder de hulp van anderen. Bij deze wil ik dus beginnen om iedereen te bedanken die een bijdrage heeft geleverd aan dit proefschrift.

Eerst en vooral bedank ik mijn beide promotoren, Prof. Dirk de Graaf en Prof. Bart Devreese, voor hun excellente begeleiding, hun vertrouwen in mij, hun niet-aflatende steun en het ter beschikking stellen van hun labo.

Mijn dank gaat ook uit naar de overige leden van mijn examencommissie en begeleidingscommissie: Prof. Grooten, Prof. Baggerman, Prof. Braeckman, Prof. Ebo, Prof. Smagghe, Dr. Aerts en Dr. Blank. Dankzij jullie terechte opmerkingen is de kwaliteit van dit werk aanzienlijk verbeterd.

Hiernaast dank ik IWT-Vlaanderen voor de financiële steun die dit project mogelijk maakte.

Ook dank ik alle co-auteurs van de verschillende hoofdstukken van dit proefschrift nog eens voor hun bijdrage aan dit onderzoek. In het bijzonder wil ik Griet Debyser bedanken om mij wegwijs te maken in de wereld van massaspectrometrie en Dr. Lina De Smet voor de praktische begeleiding doorheen het project.

Of course I also want to thank Dr. Simon Blank and Prof. Edzard Spillner to allow me to do a 6-week internship in the Institute of Biochemistry and Molecular Biology at Hamburg University. Special thanks to Simon, Frankie and Yvonne who guided me through the lab work. Thanks Laura, Julia, Thorsten, Yvonne, Andrea, Helene, Simon, Edzard, Frankie, Melanie, Sara, Tim, Dirk and Thomas for the pleasant stay in the lab, for teaching me very interesting German words and for organizing social activities such as walking on the frozen

Alster lake, movie night, for having a spicy lunch in the Mensa and going out for dinner. It was a great experience for me to live and work in Hamburg!

Verder bedank ik de collega's van het Laboratorium voor Zoöfysiologie voor alle hulp tijdens het doctoraat, maar ook voor het aangenaam gezelschap, de goede werksfeer en toffe babbels. Marleen, bedankt voor de administratieve ondersteuning, Lina voor de taxi-service, en de doctoraatsstudenten Ellen D, Ellen F, Jorgen en Tine voor het uitstaan van mijn nerveus gedoe in ons bureau, inclusief drummen en kapot nijpen van stressballen. En laat ons zeker ook Dieter en zijn fantastische droge humor nooit vergeten.

Ook wil ik 'de mensen van de Sterre' bedanken voor hulp allerhande. Dries, Jeroen, Wilfried, Mikalai en Patrick, heel erg bedankt voor de koninginnenkweek en om de bijenstand te onderhouden zodat de bijtjes altijd ter beschikking waren voor het onderzoek.

Ik mocht ook heel wat praktisch werk uitvoeren in het massaspectrometrie en proteomics labo van L-PROBE. Bedankt Griet, Laurence, Isabel, Silke, Sara, Pablo, Gonzalez, Simon, Agata en Isaak voor de hulp en de interessante discussies over proteomics.

Ook bedank ik graag Quinten voor het maken van de cover.

Dankzij mijn doctoraat kreeg ik ook de kans om via verschillende binnenlandse en buitenlandse congressen mijn werk voor te stellen aan andere onderzoekers en ideeën met hen uit te wisselen. Zo heb ik heel wat toffe mensen leren kennen die ik hopelijk ooit nog eens terugzie: Daniel, Bettie, Carol, Raquel, Gema, Davinia en Carlinhos (Valencia), Heidi (München en Milaan) en Maaike (Milaan).

Dit resultaat zou er natuurlijk nooit geweest zijn zonder de steun van familie en vrienden. Ik wil dan ook graag in de eerste plaats mijn ouders, zussen en toekomstige schoonbroers, grootouders en schoonouders bedanken. Ook bedank ik de vrienden en familieleden die steeds met interesse naar mijn doctoraat in de bloemetjes en de bijtjes hebben gevraagd.

Last but not least zou ik uiteraard mijn vriendin Fien willen bedanken die steeds begrip toonde voor mijn late werkuren en mij bleef steunen tijdens mijn doctoraat, zelfs toen dit werd gecombineerd met het zoveelste verbouwproject. Binnenkort kunnen we eindelijk samen ontspannen en genieten in ons 'hoevetje' in het hartje van de Vlaamse Ardennen.

Matthias

List of abbreviations

List of abbreviations

1D	One dimensional
2D	Two dimensional
AA	Amino Acids
ACN	Acetonitrile
Ag5	Antigen 5
AHB	Africanized honeybee
BAT	Basophil activation test
BiP	Binding immunoglobulin protein
Blast	Basic Local Alignment Search Tool
bp	basepair
CCD	Cross-reactive Carbohydrate Determinant
CPLL	Combinatorial Peptide Ligand Library
CRD	Component-resolved diagnosis
CUB	Complement C1r/C1s, Uegf, Bmp1
Da	Dalton
DPP	Dipeptidyl peptidase
DTT	Dithiothreitol
EL	Elution
ELISA	enzyme-linked immunosorbent assay
ESI	Electrospray Ionization
EST	Expressed Sequence Tags
FDR	False Discovery Rate
FT	Flow-through
FT-ICR	Fourier Transform-Ion Cyclotron Resonance
FTMS	Fourier Transform Mass Spectrometry
GlcNAc	N-Acetylglucosamine
GO	Gene Ontology
HEX	Hexamerin
HPLC	High Performance Liquid Chromatography
HSP	Heat Shock Protein
ICA	icarapin
ICR	Ion Cyclotron Resonance
IL	Interleukin
LC	Liquid Chromatography
LD50	median Lethal Dose for 50% of subjects
LTQ	Linear Trap Quadrupole
MALDI-TOF	Matrix Assisted Laser Desorption Ionization – Time Of Flight
MCDP	Mast cell degranulating peptide
MRJP	Major Royal Jelly Protein

mRNA	messenger RNA
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MW	Molecular Weight
nt	Nutcleotide
OGSv	Official Gene Set version
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PLA$_2$	Phospholipase A$_2$
PVF1	Platelet-derived Growth Factor 1
RT	Room Temperature
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SDS-PAGE	Sodium Dodecylsulphate Polyacrylamide Gel Electorphoresis
Sf9	*Spodoptera frugiperda* cell line 9
Th cell	T helper cell
TMB	3,3',5,5'-tetramethylbenzidine
TOF	Time Of Flight
UFP	Unknown Function Protein
VEGF	Vascular Endothelial Growth Factor
VIT	Venom Immunotherapy
WHO	World Health Organization

General introduction

1. ANALYSIS OF THE HONEYBEE AND BUMBLEBEE VENOM PROTEOME

1.1 The venom apparatus: an evolutionary perspective

Venoms are toxic substances containing a broad range of compounds including organic molecules, amines and alkaloids, salts and minerals, amino acids, peptides and proteins [1]. A specialized apparatus, such as a stinger, fang, hollow spine or other mechanical delivery system delivers these compounds into the victim [1;2]. Venomous animals include sea anemones, jellyfish, gastropods, cephalopods, centipedes, insects, echinoderms, amphibians, reptiles, fish and five mammalian species. The venom composition, its delivery system and physiological target can differ considerably, as these venoms serve different functions and evolved independently between the phylogenetically divergent lineages [1].

When people think of insects, images of stings and bites often flash into mind [2]. Venoms are widely distributed throughout the class of insects, but the ants, wasps and bees belonging to the order of Hymenoptera are the characteristic groups of venomous insects [2]. Remarkably, hymenopterans share a common parasitic ancestral origin [3]. Many parasitoid wasps of the Terebrantia (paraphyletic suborder [4]) still use their stinging organ (terebra) to deposit their eggs inside (endoparasitoids) or outside (ectoparasitoids) the body cavity of invertebrate hosts. In addition, venom is released which can target the host's immunity, physiology, mobility, reproductive capacity and even their behavior in order to guarantee the offspring's development [3;5]. The venom composition in parasitoids varies between species, even within the same genus, reflecting their important functional diversification during evolution [3]. In contrast, in the Aculeata, a monophyletic suborder derived within the Terebrantia [4], the ancestral ovipositor has evolved to a highly specialized device (aculeus) for injection of venom and no longer serves for the egg-laying function, except in more basal lineages (e.g. Dryinidae and Chrysididae) [3]. Only species belonging to three families sting humans with a high frequency: Vespidae (wasps), Apidae (bees) and Formicidae (ants) [6;7]. Vespids include the genera *Vespula* (yellow jackets),

1

Dolichovespula (aerial yellow jackets), *Vespa* (hornets) and *Polistes* (paper wasps), while apids include the genera *Apis* (honeybees) and *Bombus* (bumblebees). Also few ant species belonging to the genera *Solenopsis* (fire ants) and *Pogonomyrmex* (harvester ants) retained their stinger and are capable of delivering painful stings. Some well-known stinging hymenopteran species are depicted in Figure 1. While some of the aculeate species use venom only for prey capture purposes, others are also effective against predators, including humans [2]. In contrast, bees no longer have a predatory lifestyle and shifted to a diet of pollen and nectar. Therefore, the major function of their venom evolved to personal defense or defense of the colony by inflicting pain. As the possession of a venomous sting is unique in its power against very large enemies, it is believed to have facilitated the evolution towards eusociality, the condition of living in colonies with cooperative brood care, reproductive castes and overlapping of adult generation, which has evolved multiple times within the Aculeata [8]. Remarkably, some eusocial living bees have lost a functional stinger through

Figure 1: Some stinging hymenopteran species: *Apis mellifera* (a), *Bombus terrestris* (b), *Vespula vulgaris* (c), *Vespa crabro* (d), *Polistes dominulus* (e), *Solenopsis invicta* (f). Figure adapted from [7]. Pictures remain copyright of Informatiecentrum voor Bijenteelt, Ghent University (a), Moniotte Philippe (c), ww.zwolle-insecten.nl (d), Devalez Jelle (e), alexanderwild.com (f)

evolution. The stingless Meliponini have acquired defensive behaviour which differs from that of the *Apis* species. While both queens and workers of primitive stingless bee species retained the venom sac homologues of stinging bees, only the queens of evolved meliponine species have preserved this vestigial character [9].

1.2 Stingers and venoms of bees

This PhD project focused on the venom composition of two eusocial aculeate species, the European honeybee (*Apis mellifera*) and European buff-tailed or large earth bumblebee (*Bombus terrestris*). Both belong to the Apidae, the largest family of bees with over 5600 described species [10]. Honeybees an bumblebees have diverged about 77-95 million year ago [11]. The basic sting apparatus structure is anatomically similar, consisting of a long, thin, distally bifurcated venom gland which produces the venom compounds (Figure 2) [12]. Venom glands are epidermal glands that have evolved from female accessory reproductive glands [12]. Venom compounds are stored in the venom sac/reservoir and are released upon stinging. However, the life cycle of both species differs. As this can have significant implications for the venom composition and sting apparatus, relevant differences are described.

Honeybees live in large hives, which can reach up to 80,000 individuals during the summer months. Only one queen is present in the hive which lays the eggs, while workers are sterile and engage in various tasks depending on their age. While drones lack a stinger, both the queen and workers possess a morphologically distinct stinger, which is used for a different purpose. Workers use their venom for defending the hive, which is a rewarding target for many predators searching for honey, pollen, immature brood and adults. The stinger of workers is barbed and possesses an associated set of muscles and nerves, allowing to penetrate the target deeper independent of the rest of the body. In case of stinging mammals with a thick skin, the stinger is torn from the body and the muscles surrounding the venom sac will continue to pump venom into the attacker (Figure 3). This process of autotomy maximizes venom delivery [2]. The venom sac of European honeybees contains about 150 μg of dry weight venom protein, which is almost completely delivered from the sting within one minute [13]. However, tearing the stinger loose damages the honeybee's

Figure 2: Morphology of the honeybee worker sting apparatus. Venom is produced by the venom glands, stored in the venom sac and released upon stinging via the stinger and its associated set of muscles and nerves.

abdomen which leads to its death. In contrast, queens use their stinger only in deadly fights with rival queens [14]. Therefore, their stinger is curved and smooth which allows them to sting multiple times (Figure 3). Both the venom glands and reservoir are much larger than those of workers, enabling the storage of up to 5 times more venom [15]. In contrast to the worker venom composition which has been the subject of many studies, less is known about the queen venom. Queen venom has been reported to be only half as lethal to mice as worker venom [16], which points to a different venom composition. Moreover, the start of venom gland activity differs between both castes which has evolved due to the fact that queens and workers need a functional venom at a different time in their life [17]. Queens use the venom upon emergence to fight with other queens. They can live up to 5 years, but by the time they reach the age of 1 to 2 years, their venom has become inactive [16]. In contrast, workers use the venom when performing tasks outside the hive, which starts around the 20th day of adult life. The venom glands of both castes have only one secretory cycle, which starts at the end of pupation in queens and just after emergence in workers. In workers, the highest secretory activity of the venom glands is reached around the 16th day. Therefore, the venom composition is also age-dependent. In addition to these quantitative and qualitative caste- and age-related alterations, seasonal variation in venom composition has been demonstrated [18]. Also variability due to geographical and colony-dependent factors may exist in honeybee venom, but this has not been studied in-depth.
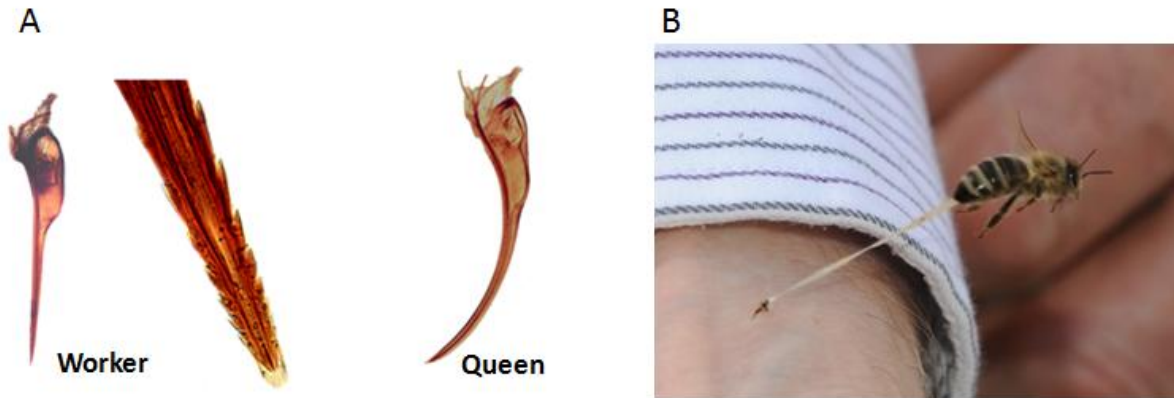
Figure 3: (A) Morphology of the stinger of the honeybee worker and queen. The stinger of the workers is straight and barbed, while the queen stinger is curved and smooth. (B) Upon stinging, the barbs present on the stinger of workers cause the release of the sting apparatus, which maximizes venom delivery. Picture B remains copyright of Kathy Keatley Garvey.

Bumblebee nests are generally much less extensive than those of honeybees. As the queen is the only individual surviving through the winter, these are the first ones you see during spring. The queen constructs a nest (usually underground) and lays eggs developing in workers which forage and live for several weeks. By the end of the summer, new queens and drones develop which mate. Only the young queens will survive the winter through diapause. Although the amount of venom released during a sting varies from species to species and within a species, generally, the amount of venom protein released by a honeybee sting is approximately five times greater than that released by a bumblebee sting [19]. However, in contrast to honeybees, both the queen and workers possess a smooth and curved stinger which is strongly attached and enables them to sting more than once [20]. In contrast to honeybee venom which has received much interest, bumblebee venom has been investigated only very scarcely.

1.3 Toxicity of bee venom

A single honeybee sting induces death in other insects and pain and inflammation in higher organisms. However, when (accidentally) disturbing a honeybee colony, it responds aggressively which can result in mass stinging events. Between different colonies, striking variation in the intensity of the aggressive response is noticeable. In docile colonies only a few bees may respond, whereas in more aggressive colonies, the response may involve hundreds or even thousands of stinging individuals [21]. The most severe aggressive

behavior is observed in the Africanized honeybee (AHB), originated in Brazil by introduction and cross-breeding of African honeybees (*A. mellifera scutellata*) with European honeybees (predominantly *A. mellifera ligustica*) [21]. AHBs reached the wild in Brazil in 1957 and spread north and south (Figure 4). They appear to produce lower amounts of venom and the general biochemical composition and lethality of their venom does not differ from their European counterparts [22]. However, AHBs have been found to sting human airways more often and are able to release higher amounts of venom upon stinging, suggesting a reason for fatalities from relatively small numbers of stings [23;24]. Since bee stings have become a public health concern in the American continent, where massive attacks by AHBs in humans and animals have been documented, interest in the physiopathological effects of honeybee venom increased [22]. Mass envenomations lead to systemic toxicity with a potential fatal outcome. Receiving more than 500 stings is usually fatal, although many victims have survived more than 1000 honeybee stings by receiving medical treatment [23;25]. A series of acute toxic effects were described in mice by inducing a severe, sublethal systemic envenomation in response to honeybee venom subcutaneous injections. A variety of increased biochemical markers revealed liver, skeletal muscle and kidney damage. Also disturbances in the coagulation system and a hemoconcentration (hematocrit and
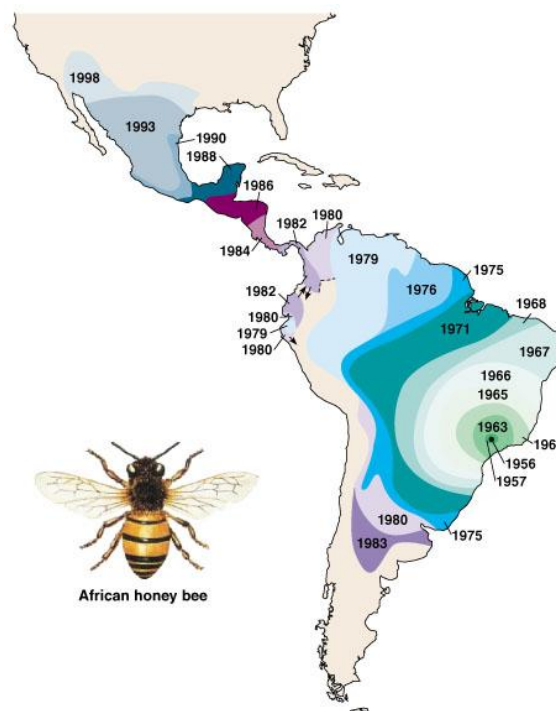
Figure 4: Spreading of the Africanized honeybee on the American continents starting from its release in the wild in 1957 until 1998. Picture remains copyright of Pearson Education, Inc.

hemoglobin increase) effect was observed, while circulating platelet and leukocyte numbers remained unaltered. In addition, an inflammatory response including edema, lipid peroxidation, nitric oxide production and systemic release of cytokines (IL-1β, IL-6, TNF-α) was demonstrated. These findings seem to be in concordance with the reported clinical effects in humans [22]. Immediate effects include localized pain, swelling and erythema at individual sites. Next, large envenomations cause early systemic symptoms which include fatigue, dizziness, nausea, vomiting and diarrhea. Within 24 hours, hemolysis, hemoglobinuria, rhabdomyolysis and hepatic transaminase enzyme elevations develop. Also subendocardial damage and cardiac enzyme elevation may occur, while renal insufficiency and electrolyte abnormalities develop secondary to rhabdomyolysis, hemolysis and acute tubular necrosis [26].

The search for an effective treatment for bee envenomation has been the subject of many studies. Already in 1999, Jones *et al.* [27] produced the first antivenom against AHB venom by immunizing sheep. However, the use of heterologous sera can cause anaphylaxis and serum sickness. Recently, a new antivenom was developed by the use of phage display technology, which enabled the production of human antibody fragments binding to two major honeybee venom compounds [28]. As currently no commercially available antivenom exists, massive envenomations have to be treated with supportive care, involving antihistamines, steroids, epinephrine and airway assistance. Aggressive hydration reduces the likelihood of rhabdomyolysis-induced renal insufficiency [26].

Massive bee envenomations by other honeybee subspecies are less common but have been described in regions such as England, India, and Hawaii, where Africanized bees are non-endemic [26]. Venoms of several species of the *Apis* genus (*A. mellifera*, *A. dorsata*, *A. cerana*, *A. florea*) exhibit an almost identical median lethal dose (LD50) for mice [16], which indicates that their venom composition is very similar. Subcutaneously injected honeybee venom has a LD50 of 41.6 mg/kg, which is nearly one order of magnitude weaker in comparison to LD50 values obtained by intravenous or intraperitoneal routes. This suggests that a substantial fraction of the lethal components of bee venom are prevented to enter the systemic circulation when administered subcutaneously, which may be caused by venom inactivation by host factors and/or binding to locally available tissue target sites [22].

Compared to honeybees, bumblebees rarely sting humans. Intraperitoneal injection of venom of the bumblebee *B. impatiens* (LD50= 7.2 mg/kg) in mice showed that it is about half as lethal to vertebrates as honeybee venom (3.5 mg/kg) [29]. However, the venom composition of this bumblebee species has not yet been investigated. The LD50 of the venom of *B. terrestris* has not yet been determined.

1.4 Biological function of bee venom compounds

The European honeybee, *A. mellifera*, is the most important managed pollinator and is domesticated by humans for production of honey. Therefore, a honeybee sting has been an unpleasant experience for many people, and many researchers became intrigued by the honeybee venom composition and toxic effects of individual venom compounds. More than 1700 scientific publications on the composition and effects of bee venom in animals and humans have been published [30]. Earliest efforts to unravel its venom composition even date back to the work of Langer in 1897, who found that it consists of active and hemolytic basic components [31]. Today, up to 28 venom proteins and peptides (Table S1) and 4 biogenic amines are described.

The two highest abundant venom compounds, melittin and phospholipase A_2 (PLA_2) constitute ~50% and ~10-12% of the venom dry weight respectively [32]. Both have direct cell lytic activities. PLA_2 activity leads to cell lysis by cleaving plasma membrane phospholipids. It hydrolyzes the 2-acyl bonds of phosphatidylcholines, phosphatidylethanolamines, phosphatidylinositols and phosphatidylserines, releasing lysophospholipids and fatty acids which themselves may further damage the membrane [33]. Melittin is a helical amphipatic peptide with hemolyzing activity due to its ability to interact with and disrupt cell membranes [33]. It also causes pain by activating specific receptors in primary nociceptive neurons [34]. In addition, melittin appears to be the main trigger of inflammasome activation [35], which induces a caspase-1-dependent inflammatory response. This is characterized by recruitment of neutrophils to the site of envenomation, which can protect against the damaging effects of envenomation. In addition, melittin is well-known for its antimicrobial properties. Its activity may contribute to the recently suggested function of honeybee venom in social immunity of the hive [36;37]. Indeed, venom peptides are smeared on the body surface of females and on wax combs, which may protect the bees against pathogens attracted to the constant and relatively high temperature and humidity

levels maintained in the hive [36]. Additionally, mast cell degranulating peptide, a neurotoxin which mediates mast cell degranulation at low concentration [38], was found to possess antimicrobial activity [39]. In contrast, the neurotoxic peptide apamin was also suggested to be an antimicrobial agent [36], but recently it was shown to lack antiseptic activity [39]. Many solitary hymenopteran species have evolved a venom with antimicrobial properties, which additionally suggests an important function at the organismic level [3]. Indeed, only few pathogens seem to be able to use venom-producing organs of Hymenoptera as a natural route of infection, which shows that microbial diffusions into these organs and from here to the rest of the body are possibly limited by the biochemical venom properties, although also physical barriers may contribute to this observation [3]. In addition, it is remarkable that no significant infectious disease of medical importance has been described to be vectored by hymenopteran stings, while many insects are vectors for microbial pathogens and parasites [3]. In contrast to insect bites, stings are quite rare, which may explain why transmission mechanisms of infectious diseases through venoms have not evolved.

Besides melittin and PLA$_2$, honeybee venom contains multiple compounds of lower abundance. Only few of them have been functionally characterized. Hyaluronidase cleaves hyaluronic acid, which is a large and highly abundant glycosaminoglycan of the vertebrate extracellular matrix. This cleavage facilitates the penetration of venom constituents into the body [40]. Additionally, for some well-known but functionally non-characterized honeybee venom compounds we propose a function based on similar venom compounds from other species. First, in snake venom, acid phosphatase has been suggested to play a role in liberating purines (mainly adenosine). It acts as a multitoxin and potentiates venom-induced hypotension and paralysis [41]. Second, the platelet-derived growth factor (PVF1) may act similar to snake venom vascular endothelial growth factor (VEGF)-like molecules, which are the most potent vascular permeability factors known and which can facilitate venom spreading [42]. Two other enzymes, carboxylesterase and serine carboxypeptidase may play a role in degradation of insect neurotransmitters [43] and a wide range of proteins [44], respectively. The CUB serine protease is similar to the *B. terrestris* venom homologue and snake venom proteases, which act as a fibrin(ogen)olytic enzyme, decreasing the concentration of blood fibrinogen and facilitating the spread of bee venom components throughout the bloodstream in mammals [45]. In addition, it may modulate the innate

immunity by acting as a prophenoloxidase-activating factor, which triggers the phenoloxidase cascade and induces a lethal melanization response in target insects [46;47]. Dipeptidyl peptidase IV (DPP IV) was suggested to play a role in the conversion of venom components into their active forms in the venom gland, while it may also enhance or decrease the chemotactic activity of immune cells after the sting [48].

For others, the exact function remains elusive. Based on its 3-dimensional structure, Api m 6 is highly likely to be a serine protease inhibitor. However, its natural function and putative binding partners remain unknown [49]. Major royal jelly protein (MRJP) 8 and 9 were found to be the most ancient members of the MRJP protein family, which lack the later evolved repetitive regions suggested to have a nutritious function. Therefore, both MRJPs may possess the original but yet unknown pre-royal jelly function [33]. Vitellogenins are transported into oocytes to serve as food for the developing embryo. However, honeybee vitellogenin is thought to act as a multifunctional compound involved in many processes, such as hormone signaling, food-related behavior, immunity, stress resistance and longevity. Therefore, assumptions for its role as a venom protein remain speculative [50]. Also a role for the hexamerin 70a (HEX 70a) is not clear. Hexamerins are larval amino acid storage proteins important during metamorphosis. In addition, HEX 70a was found to be expressed in the fat body of adult workers and gonads from workers, queens and drones, suggesting other undefined tissue-specific functions [51;52]. Others, such as icarapin and secapin, possess no functional domains elucidating their exact function. We hypothesize that they have a toxic function as they were shown to be moderately/highly abundant [53]. Secapin is also a major component of queen bee venom [54]. In addition, for multiple compounds described to be present in the venom by studies of the 1970s and 80s, no sequence data have been generated (Table S1). Therefore, determining their function remains impossible.

In addition, the venom may contain lowly and very lowly abundant compounds without toxic activity. These so called venom trace molecules only have a local function in the venom duct or reservoir (maturation or stabilization of the secretes, protection and recovery of the gland tissue) or are released by leakage of the gland tissue [7]. Both DPP IV and the C1q-like protein were suggested to be members of the venom trace molecules [55], although for DPP IV also a toxic function has been proposed (see higher) [48].
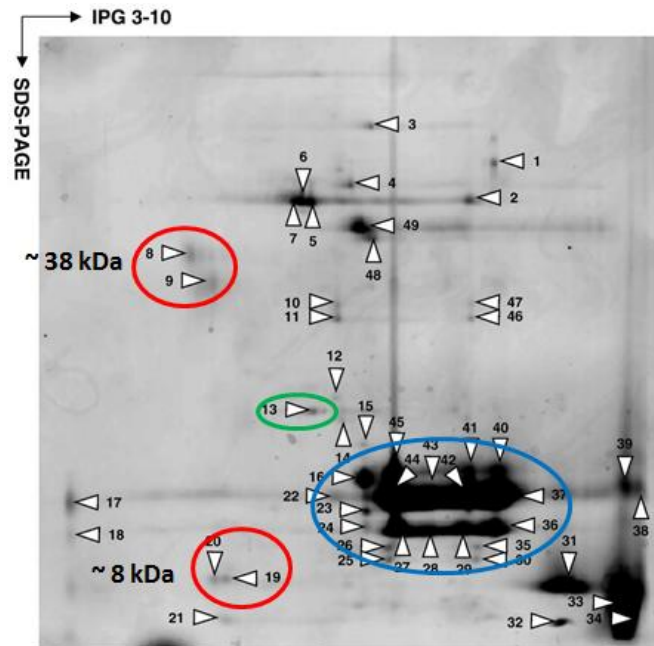
Figure 5: In 2005, Peiren and coworkers analyzed the honeybee worker venom by 2D-PAGE separation of the venom proteins and mass spectrometry analysis of excised spots (see numbers on gel) [53]. Icarapin was identified in four spots (indicated in red). Some spots have a variation in molecular weight of about 30 kDa. A large part of the gel is obscured by phospholipase A_2 (indicated in blue), comprising about 10-12% of the venom proteins. In addition, this study newly identified PVF1, a platelet-derived growth factor (indicated in green).

Besides the identification of 28 honeybee venom proteins and peptides, another level of complexity of the venom composition has been revealed. Protein heterogeneity or post-translational modifications generate different isoforms of several compounds. Protein heterogeneity can be caused by allelic variation at a single gene, the occurrence of multiple genes encoding highly homologous proteins or by alternative splicing of a single transcript [56]. In 2001, Kettner and coworkers [57] identified four isoforms of Api m 6, only differing in the amino- and carboxy-terminal ends. Using genome-level information, this heterogeneity was later found to be caused by allelic variation [56]. Also, upon 2D-gel separation of pure venom, multiple icarapin protein isoforms were found to exist (Figure 5) [53]. In addition, two highly similar icarapin transcripts were found [58], which differ by only 0.3 kDa in their theoretical molecular weight (MW) (Figure 6). As some 2D-gel spots differ by about 30 kDa (Figure 5), we suggest that additional transcripts are generated by the honeybee venom glands. Finally, also post-translational modifications, such as glycosylations and other enzymatic processes, contribute to the venom complexity. SDS-PAGE separation of purified PLA_2 reveals three protein bands. While the lowest MW band represents non-glycosylated

PLA$_2$, the two higher MW bands correspond to different PLA$_2$ glycoforms [59;60]. In addition, melittin's inactive precursor promelittin and the complete series of ten conversion intermediates to the mature peptide were detected in crude honeybee venom [61]. This maturation process is executed by DPP IV activity in the venom.

```
variant1      MKTLGVLFIAAWFIACTHSFPGAHDEDSKEERKNVDTVLVLPSIERDQMMAATFDFPSLS 60
variant2      MKTLGVLFIAAWFIACTHSFPGAHDEDSKEERKNVDTVLVLPSIERDQMMAATFDFPSLS 60
              ************************************************************

variant1      FEDSDEGSNWNWNTLLRPNFLDGWYQTLQSAISAHMKKVREQMAGILSRIPEQGVVNWNK 120
variant2      FEDSDEGSNWNWNTLLRPNFLDGWYQTLQT----HMKKVREQMAGILSRIPEQGVVNWNK 116
              *****************************:    *************************

variant1      IPEGANTTSTTKIIDGHVVTINETTYTDGSDDYSTLIRVRVIDVRPQNETILTTVSSEAD 180
variant2      IPEGANTTSTTKIIDGHVVTINETTYTDGSDDYSTLIRVRVIDVRPQNETILTTVSSEAD 176
              ************************************************************

variant1      SDVTTLPTLIGKNETSTQSSRSVESVEDFDNEIPKNQGDVLTA 223
variant2      SDVTTLPTLIGKNETSTQSSRSVESVEDFDNEIPKNQGDVLTA 219
              *******************************************
```

Figure 6: Sequence alignment of the two identified alternative splice variants of Api m 10. The molecular weight of these variants differs by only 0.3 kDa.

In addition to proteins and peptides, honeybee venom presents biologically active amines. Serotonin (5'hydroxytryptamine) injection may cause an increase of pain in large animals, lethal vasoconstriction in smaller predators and neurotoxicity in insects [62]. Histamine [63] produces dilatation and increased permeability of the blood vessel capillaries. Also dopamine and noradrenaline [64] increase the venom distribution by elevating the rate of the heart beat [32]. Finally, honeybee venom contains carbohydrates and lipids with an unspecified function [65].

In contrast to *A. mellifera*, the venoms of other species belonging the Apidae family have received only little interest. The few studies focusing on the venom composition of other honeybee species (*A. cerana*, *A. dorsata*, *A. florea*) only provided evidence for highly similar *A. mellifera* homologues (Table S2). Also the venom composition of bumblebees has only been poorly resolved (Table S3). So far, sequences are available of only five *B. terrestris* venom compounds. The PLA$_2$ [66] and serine protease [45] proteins appear to be homologous to their honeybee venom counterparts, while the two described bombolitin peptides ([67]; GenBank: ADY75782.1) show structural and biological properties similar to honeybee venom melittin. The Kunitz-type serine protease inhibitor [68] affects the victim's

hemostatic system via its antifibrinolytic activity. In addition, *B. terrestris* venom was shown to exert hyaluronidase, acid phosphatase and casein hydrolyzing protease enzymatic activity [66]. Venom proteomes of few other bumblebee species (*B. ignitus*, *B. pennsylvanicus*, *B. lapidarius, B. hypocrita sapporoensis* and *B. ardens ardens*) have been investigated (Table S3), identifying PLA$_2$s [66;69], serine proteases [46;66;70], bombolitins [71-73], a mast cell degranulating peptide [74] and a Kunitz-type serine protease inhibitor [46;75].

1.5 Unraveling the venom proteome

For venom research, venom can be collected by manual milking or electrostimulation (Figure 7). Manual milking is preferred as samples collected by electrostimulation may contain contaminants derived from saliva or digestive tract fluids [7]. Moreover, the protein composition of a venom sample collected by manual milking may closely resemble that of venom injected during a natural honeybee sting: in addition to the release of venom proteins which are produced by the venom glands, also proteins originating from other tissues such as the sting apparatus cell lining, stinger lancets and/or stinger lubricant, which has been hypothesized to be generated by the Dufour gland [76], may be released.

The proteomic technologies used to explore Hymenoptera venom compositions have gone through a remarkable evolution. Most research conducted in the 1970s and 80's relied on single compound-oriented, time-consuming and lowly sensitive techniques, identifying only a few, primarily highly abundant compounds. Venom constituents were often isolated by chromatographic means, followed by bio-assays to determine their function (Table S1, S2 and S3). However, as co-purified highly abundant venom compounds may have influenced these bio-assays, their determined function may be unreliable. Moreover, amino acid sequences are often lacking, e.g. for honeybee venom minimine [77], cardiopep [78] and adolapin [79] (Table S1), and bumblebee venom hyaluronidase and acid phosphatase (Table S3) [66]. Other studies applied chemical sequencing via Edman degradation to obtain sequence information of individual compounds (Table S1, S2 and S3). One of the first Hymenoptera venom profiling studies involving more sophisticated techniques was conducted in the early 80's. The work used 2D-PAGE to characterize the venom proteins of the honeybee and several wasp species [80]. Only since 2000, mass spectrometry was used for profiling of Hymenoptera venoms [80]. The shotgun proteomics strategy, based on digesting proteins into peptides and sequencing them using tandem mass spectrometry
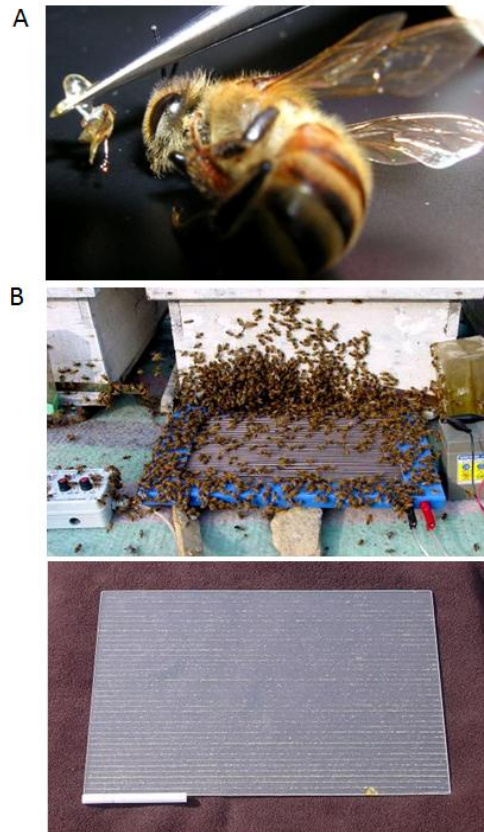
Figure 7: Honeybee venom collected by manual milking (A) and electrical milking (B). (A) The stinger is pulled out and gently compressing the venom sac releases the venom via the stinger. Venom appearing at the tip of the stinger is collected in phosphate buffer. (B) The collector is placed in front of the hive and generates low voltage electric pulses, which stimulates worker bees to sting through a latex film onto a glass collector plate. Afterwards, the collector plate is removed and the venom can be collected in the form of a white powder.

(MS/MS), has been widely adopted. In 2002, Stöcklin and Favreau [61] published the first mass spectrometry analysis of honeybee venom, identifying several major compounds (Table S1). The use of multiple proteomic methods and the availability of the *Apis mellifera* genome since 2006 [81] has boosted the detection of new bee venom compounds significantly [36;53;61;82;83].

Two studies, published by a cooperation of the Laboratory of Zoophysiology and L-PROBE, successfully identified new compounds of the honeybee venom proteome by the combination of 2D-PAGE with MALDI-TOF/TOF MS [18;20]. A third gel-based proteome study analyzed the honeybee venom gland tissue and identified several compounds putatively involved in protecting the venom gland secretory cells from the toxins they produce [33]. However, the gel-based approach lacks dynamic range and sensitivity to allow the detection

of lowly abundant compounds. Therefore, previous studies suggested the existence of yet unknown venom compounds in the honeybee venom proteome. For example, although overloading the 2D-gel enabled to improve the spot intensity and resolution of some minor proteins, several spots remained unidentified [55] (Figure 8). In addition, a large part of a typical honeybee venom 2D-gel is obscured due to highly abundant compounds, such as PLA$_2$, that mask the detection of lowly abundant compounds with similar molecular weight and pI (Figure 5 and 8). Besides, fractions with an extreme pI or molecular weight remain largely unexplored in 2D-PAGE separation.
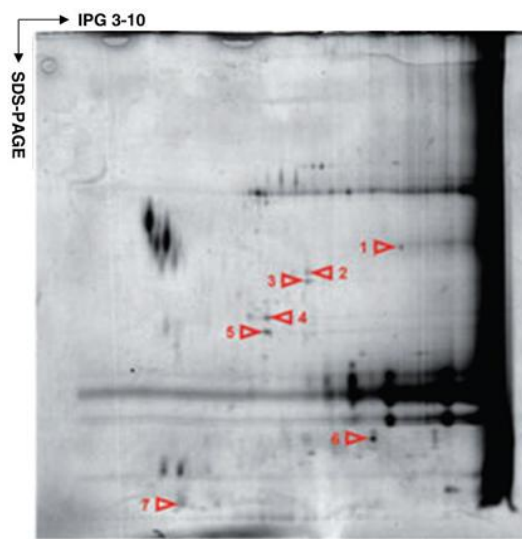
Figure 8: Overloading a 2D-PAGE gel with honeybee venom reveals additional spots of lowly abundant compounds (indicated in red). Spot 6 could be identified as a C1q-like venom protein by mass spectrometry [55]. The other spots remained unidentified. Due to protein overloading, the high isoelectric point (pI) region of the gel is covered by phospholipase A$_2$.

Mass spectrometry studies of in-liquid digested liquid chromatography (LC) venom fractions may overcome these issues related to gel-based proteomics and gain deeper insights in venom proteomes/peptidomes. In addition, several recent studies have shown that the use of a combinatorial peptide ligand library (CPLL) can significantly improve the coverage of proteomic analyzes as this allows access to many lowly abundant compounds in complex proteomes [84]. The solid-phase CPLL consists of a bead-bound set of possibly 64 million different hexapeptides, which are bound by constituents of the protein mixture upon incubation. Highly abundant proteins saturate their high-affinity binding sites and excess protein is washed away, while lowly abundant proteins are enriched by concentration on

their specific affinity ligands (Figure 9). This technology is commercially available since 2007 under the trade name ProteoMiner and has been used in studies of the 'deep' venom proteome of two snake species, namely the Western diamondback rattlesnake (*Crotalus atrox*) [84] and the African puff adder (*Bitis arietans*) [85;86], which led to the discovery of a large number of proteins previously undetected in these proteomes. CPLL has so far not been used in Hymenoptera venomics and may allow to identify many more unidentified lowly abundant compounds.
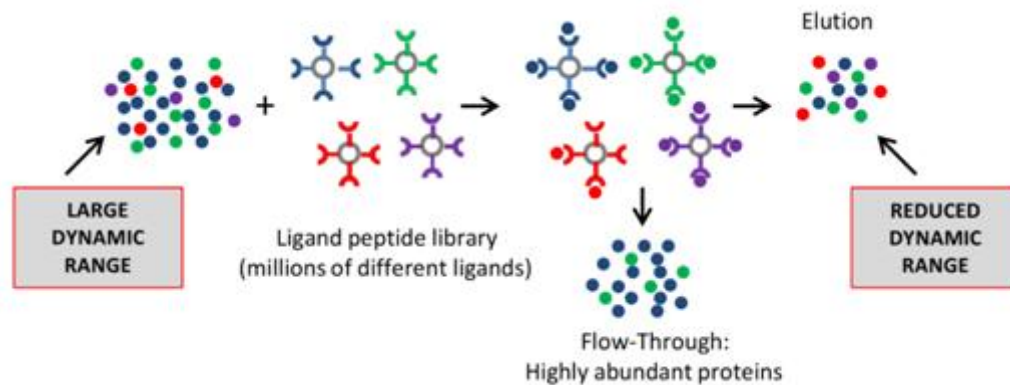
Figure 9: Principle of the combinatorial peptide ligand library (CPLL). A mixture of proteins is presented by colored dots. The initial sample has a large dynamic range of protein concentrations. It contains high amounts of green and blue protein, but only few red and purple protein. This sample is incubated with the CPLL, which is a bead-bound hexapeptide library. Highly abundant proteins saturate their high-affinity binding sites and excess, unbound protein is washed away (flow-through), while lowly abundant proteins are enriched by concentration on their specific affinity ligands. The eluted sample has a reduced dynamic range of protein concentrations compared to the initial protein sample.

The modular arrangement of MALDI and ESI ionization with different types of mass analyzers has resulted in a wide variety of mass spectrometric instrumentation [87]. Many of them have been used in honeybee venom research, except Fourier transform-based mass spectrometers (FTMS), although this equipment provides the highest performance in mass resolution and mass accuracy [88]. In the context of venom research, only few research groups used this technology (Orbitrap or FT-ICR) for protein identifications within the entire venom [5;89-92]. One of these studies was conducted by a cooperation of the Laboratory of Zoophysiology and L-PROBE, which used 2D-LC-ESI-FT-ICR-MS/MS to investigate the venom composition of the ectoparasitoid wasp, *Nasonia vitripennis* [5]. Using a shotgun proteomic

strategy, sixty venom proteins were identified starting from the content of 10 venom reservoirs of this minuscule hymenopteran insect only. The discovery of a high number of new venom constituents in these studies points to a highly effective technology for identification purposes in these complex protein mixtures. Therefore, the combination of the CPLL venom sample pre-treatment with FTMS is a promising approach to identify new honeybee venom compounds.

Obtaining peptide sequences from acquired MS/MS spectra is most often performed using the database search approach. However, the database sequence content is important for successful application of this method [93]. Venom proteins can be identified through cross-species protein identifications, but the success of this approach depends on the level of protein homology [7]. As homology decreases, MS/MS data need de novo sequencing techniques combined with database blasting [7]. Alternatively, mass spectra can be searched against venom gland transcriptome sequence datasets [94]. However, these have not been generated for honeybee and bumblebee species. Since 2006, the honeybee genome became available [81], which enabled the production of protein prediction datasets, providing significant benefits for protein identifications. However, the first generated genome sequence was noted to have a bimodal GC content that affected the quality of the assembly in some regions and the annotation had fewer genes in the initial gene prediction set (OGSv1.0) than would have been expected based on other insect genomes sequenced since then [95]. Therefore, while the previous genome sequence was obtained by Sanger sequencing and a whole-genome-shotgun model, the honeybee genome was recently re-sequenced using next-generation sequencing which allows a much deeper sequence coverage. This resulted in an improved genome assembly (Amel_4.5), which is more contiguous and complete, and a new gene annotation set (OGSv3.2), which includes ~5000 more protein-coding genes, increasing the gene set by about 50% [95]. Therefore, searching generated venom mass spectra against this improved dataset may identify a new set of venom proteins.

In contrast to honeybee venom, the venom of *B. terrestris* has never been investigated using mass spectrometry. So far*, B. lapidarius* is the only bumblebee species from which MS data on its venom proteins are available (Table S3). ESI-MS resulted in the detection of 24

compounds and the three major compounds were identified as three bombolitins [73]. However, this study was hampered by the lack of a well annotated genome. Recently, also the genomes of two bumblebee species, *B. terrestris* and *B. impatiens* (frequent in eastern North-America), were sequenced using a next-generation sequencing approach [96]. This now allows an in-depth proteomic analysis of the venom composition of these species. However, like honeybee venom, *B. terrestris* venom contains some highly abundant compounds [7;45] (Figure 10). Therefore, the same issues of gel-based proteomics as those described for honeybee venom apply for bumblebee venom. Consequently, protein enrichment of lowly abundant compounds and the application of a highly sensitive proteomic technology is also required to obtain in-depth insights in the bumblebee venom proteome.
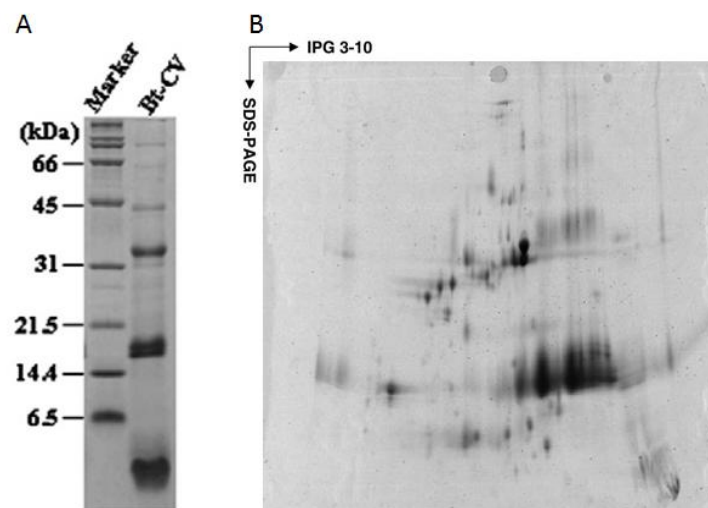
Figure 10: 1D-SDS-PAGE (A; [45]) and 2D-SDS-PAGE (B; [7]) separation of the venom proteins from *Bombus terrestris* reveals the presence of several highly abundant compounds.

2. HYMENOPTERA VENOM ALLERGY

2.1 Allergy mechanism, symptoms and prevalence

While the toxic activity of Hymenoptera venoms is only of medical importance in case of massive sting events, their allergenic properties are of more important concern for human health. In man, early exposure to bee venom evokes IgG1, IgG2 and to a lesser extent IgG4 antibody responses, whereas long-term exposure often found in beekeepers drives the immunity to an IgG4 type of humoral response [97;98]. However, some people develop a

venom allergy, which is an IgE-mediated type 1 hypersensitivity of non-atopic origin [7]. In this case, allergens are taken up and are processed by dendritic cells, which stimulate allergen-specific (CD4+) T helper 2 (Th2) cells, causing the production of Th2 cytokines, such as interleukin 4 (IL-4) and IL-13 (Figure 11). These are responsible for class switching to the ε immunoglobulin heavy chain, allowing IgE production by B cells. IgE binds the high-affinity receptor for IgE, FcεRI, which is expressed at the surface of mast cells and basophils. Upon cross-linking of the IgE–FcεRI complexes by allergen, mast cells and basophils degranulate, releasing vasoactive amines (mainly histamine) and lipid mediators (prostaglandins and cysteinyl leukotrienes), which characterize the immediate phase of the allergic reaction. IgE also binds FcεRI at the surface of dendritic cells and monocytes, as well as the low-affinity receptor for IgE, FcεRII, at the surface of B cells. This process increases the uptake of allergen by these antigen presenting cells and the subsequent presentation of allergen-derived peptides to specific CD4+ T cells, which drive the late phase of the allergic reaction. In IgE-mediated venom allergy, the immediate allergic reaction starts within minutes to one hour after the sting. Late-phase reactions, starting between three to six hours after the sting are exceedingly rare in Hymenoptera venom allergy.

Upon a Hymenoptera sting, patients may suffer from large local or systemic reactions. A large local reaction is defined as a swelling around the site of the sting exceeding a diameter of 10 cm and which lasts longer than 24 hours (Figure 12) [99]. The underlying mechanism of large local reactions is unknown [99]. The prevalence of large local reactions upon Hymenoptera stings varies from 2.4% up to 26.4%. Systemic allergic reactions have been reported to occur in 0.8 to 5% of the general population [7], but people with specific outdoor professions such as beekeepers, gardeners and farmers are at much higher risk [100]. Symptoms include pruritus, urticaria, angioedema, nausea, vomiting, diarrhea, rhinoconjunctivitis, bronchiospasm, hypotension, cardiovascular collapse and loss of consciousness (Figure 12) [101]. Systemic reactions to insect stings can be measured using the Müller grading system, which classifies reactions according to the degree of the severeity of the reaction (Table 1) [99;102;103]. Severe anaphylactic reactions may leave patients with a permanent disability such as hypoxic brain damage with permanent neurologic deficits and myocardial infarction. Even fatal reactions after insect stings may occur [99], although this is rare. Only 0.03–0.48 fatalities per 1,000,000 inhabitants occur each year due to insect
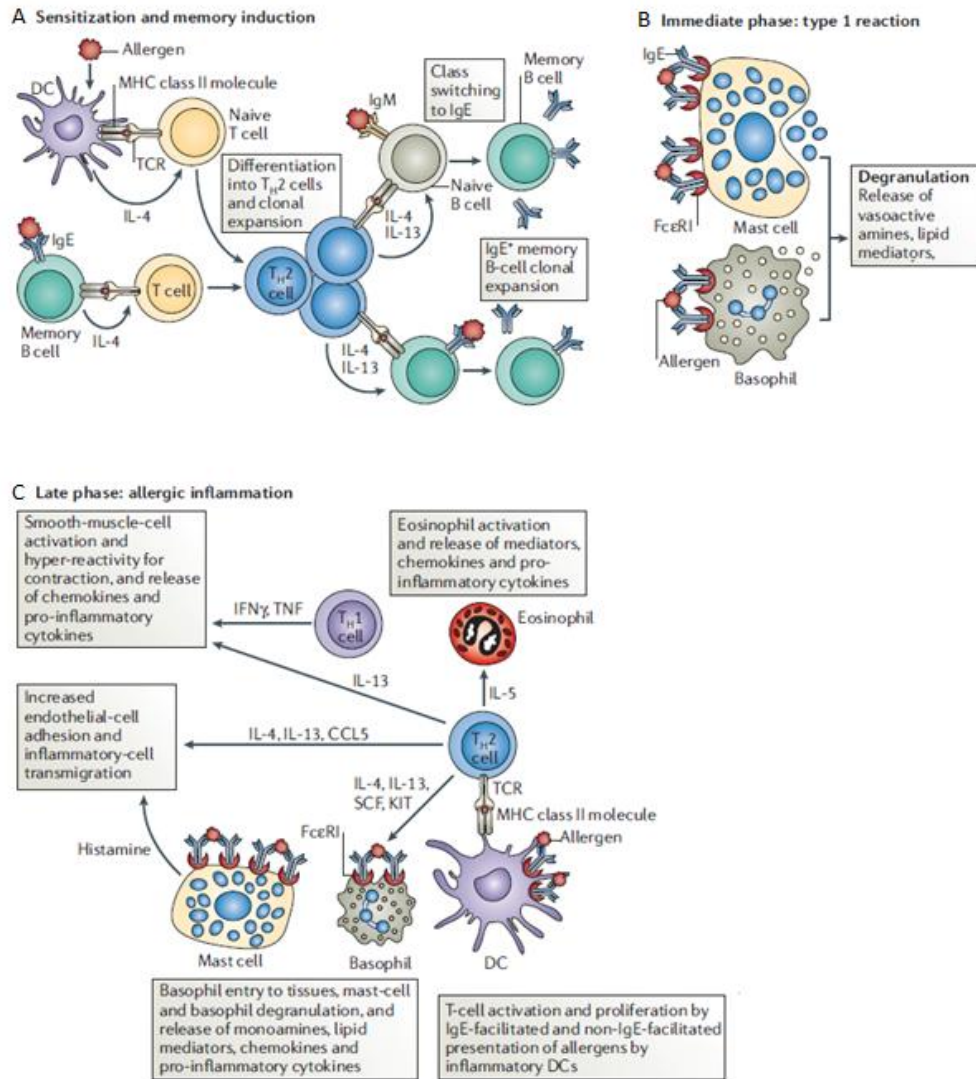
Figure 11: The mechanisms of allergic reactions. (A) Sensitization and memory induction. Allergens are taken up and are processed by dendritic cells, which causes differentiation and clonal expansion of T helper 2 cells (Th2). These produce IL-4 and IL-13 cytokines which induce class switching to IgE and clonal expansion of naive and IgE$^+$ memory B-cells. In addition, IgE at the surface of allergen-specific IgE$^+$ B cells and other IgE-sensitized antigen-presenting cells facilitates antigen presentation to T cells. T-cell activation in the presence of IL-4 increases the differentiation into Th2 cells. (B) Immediate phase of the allergic reaction. IgE binds the high-affinity receptor for IgE (FcεRI), which is expressed at the surface of mast cells and basophils. Upon cross-linking of the IgE–FcεRI complexes by allergen, mast cells and basophils degranulate, releasing vasoactive amines (mainly histamine) and lipid mediators (prostaglandins and cysteinyl leukotrienes), which contribute to the immediate symptoms of allergic reactions. (C) The late phase allergic reaction involves the recruitment, activation and persistence of eosinophils and T-cells at the sites of allergen exposure. Local IgE-facilitated antigen presentation by dendritic cells (DCs) increases T-cell activation. Both eosinophils and activated mast cells and basophils, release allergic mediators. Figure adapted from [97].

stings [3]. In the United States for instance, the probability of dying following contact with hornets, wasps or bees would be in the same order of magnitude (odds of 1 in 71,623) as being struck dead by lightning (1 in 84,079) or legally executed (1 in 96,691). However, the true prevalence of mortality induced by stings may be underestimated, as sting fatalities go unrecognized and misinterpreted [3;104]. Also, for most patients as well as for their families, an anaphylactic reaction after a Hymenoptera sting is a very traumatic event. It has been demonstrated that patients with anaphylactic responses following yellow jacket stings experienced impairment in their quality of life especially because of the emotional distress associated with having to be constantly on the alert while leading their everyday lives [99]. To date, no parameter has been identified that can predict who will have a future reaction and whether it will be a large local reaction or systemic reaction. Several concomitant factors may account for the occurrence of a systemic reaction in individual patients. These include environmental (the frequency of stings and the type of insect), genetical (the persistence of sIgE antibodies and probably other factors) and individual (age, asthma, mastocytosis or ischaemic heart disease, concurrent medication and others) factors [104].
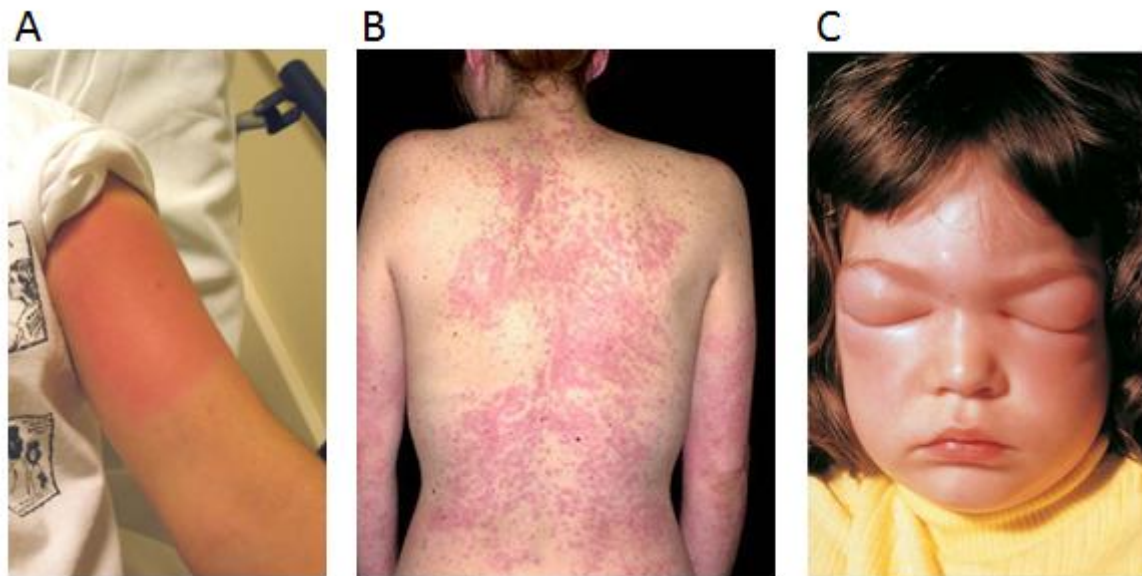
Figure 12: Symptoms of allergic reactions upon stings: (A) large local reaction, (B) urticaria, (C) angioedema. Pictures remain copyright from respectively Dr. P. Marazzi/Photo Researchers, Inc. (A), Dr Adrian Morris (B) and [105] (C).

The prevalence of stings from all Hymenoptera ranges from 56.6% to 85.5% in an adult life [106]. Epidemiologic data specifically addressing honeybee sting are rare. Several studies reported that approximately one-third of the Hymenoptera stings are due to honeybee. No data exist about the prevalence of large local reactions and fatalities attributed solely to honeybees, but one study showed that they are responsible for about half of the systemic reactions related to Hymenoptera stings. Beekeepers are a unique population of those affected by honeybee venom allergy. They have higher local reaction rates of 12% to 76% and higher systemic reaction rates of 4.4% to 43% [99;106]. Interestingly, protection correlates with receiving more than 200 stings per year, although 50 stings per year may also provide benefit. Especially beekeepers (and their family members) receiving fewer than 25 sting per year have a high systemic reaction rate of 45% [106].

As the risk of being stung by a bumblebee is very small, allergic reactions to bumblebee stings are rare. However, allergic reactions have been reported in occupational settings [107]. As bumblebees are increasingly used as pollinators of greenhouse plants, the prevalence of bumblebee venom allergy expanded, especially in greenhouse workers [66;107;108].

Table 1: Classification of allergic reactions modified according to Müller [102;103].

Large local reaction	swelling at site of sting with diameter >10 cm, lasting >24 h
Systemic reaction grade I	generalized urticaria, itching, malaise and anxiety
Systemic reaction grade II	any of the above, plus 2 or more of the following: angioedema, constriction in the chest, nausea, vomiting, diarrhea, abdominal pain and dizziness
Systemic reaction grade III	any of the above, plus 2 or more of the following: dyspnoea, wheezing, stridor, dysphagia, dysarthria, hoarseness, weakness, confusion and fear of death
Systemic reaction grade IV	any of the above plus 2 or more of the following: a drop in blood pressure, collapse, loss of consciousness, incontinence (urine or stools), cyanosis

2.2 Treatment of Hymenoptera venom allergy

The treatment of allergic symptoms depends on the severity of the allergic reaction. Large local reactions are treated with topical/systemic corticosteroids and antihistamines, and by

cooling the swollen area. In case of systemic reactions, also auto-injectable adrenaline should be used as emergency medication [100]. To provide protection from future stings, venom immunotherapy (VIT) is the treatment of choice for patients with grades III and IV. In patients with grades I and II, additional factors, such as high exposure to venoms or impaired health-related quality of life due to venom allergy, are taken into consideration before making a decision of VIT treatment [101;109]. The VIT procedure consists of subcutaneous injections of venom extract in two phases: the incremental and the maintenance dose phase. Different protocols are used for the incremental dose phase, which allow for achieving the maintenance dose phase from within 12 weeks (conventional) to a few days (rush) or a few hours (ultra-rush). The maintenance dose of 100 µg of venom extract is given every 4-6 weeks usually for 3-5 years [101;109].

The exact mechanisms responsible for the beneficial aspects of VIT are not yet fully understood (Figure 13). It appears that regulatory T cells play a significant role for a balanced Th1/Th2 profile by production of IL-10. VIT induces a shift from Th2-type (IL-4) towards a Th1-type (IFN-γ) cytokine response. Increases in the levels of IL-10, IFN-γ and TGF-β lead to decreased mast cell and eosinophil activation, and class switching results in down-regulation of IgE production and increased IgG4 production [97]. In addition, during build-up VIT a transiently reduced number of circulating basophils has been described without significant effect on individual basophil histamine content or release. In contrast, maintenance VIT lowers the content and release of histamine by basophils upon stimulation with allergen [110].

As previously mentioned, receiving a high number of stings correlates with protection of beekeepers, which seems to be mediated through the induction of bee-venom-specific IgG. The natural exposure to large doses of venom proteins resembles VIT as both lead to modulation of peripheral T-cell responses through the generation of allergen-specific IL-10-secreting T-cells and the increased synthesis of IL-10 by monocytes and B cells [97].

VIT is proven effective in the majority of Hymenoptera venom allergic patients. However, as systemic allergic side effects to immunotherapy injections have been reported, as well as patients which were not protected after immunotherapy treatment, there is considerable interest in improving safety and efficacy of VIT [111;112]. Remarkably, honeybee VIT
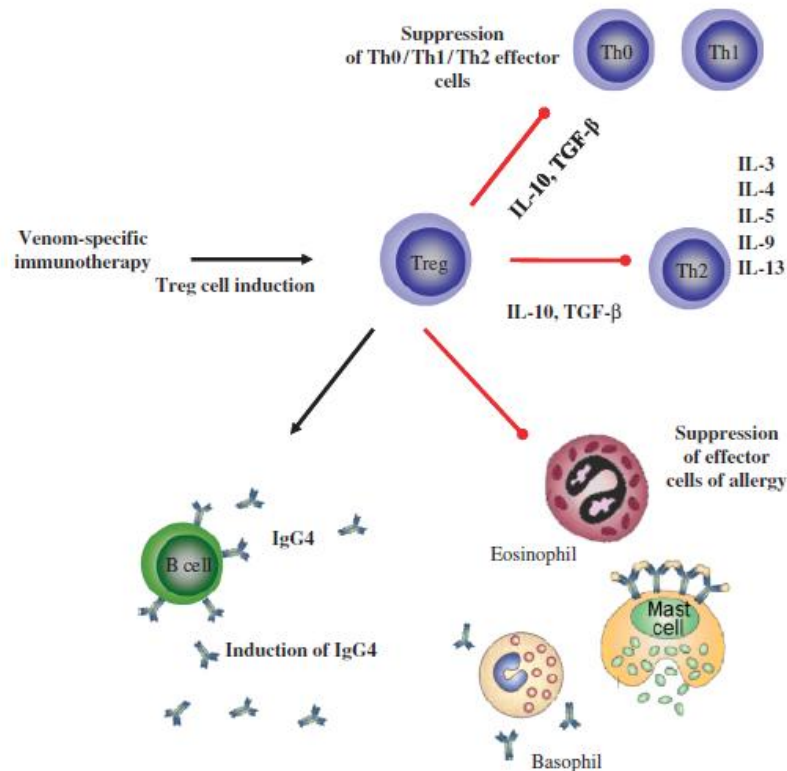
Figure 13: Venom immunotherapy induces the generation of regulatory T (Treg) cells, which play suppressive roles in proliferative and cytokine responses against the venom allergens. Treg cells are characterized by IL-10 and TGF-β secretion capacities that directly or indirectly influence effector cells of allergic inflammation, such as mast cells, basophils and eosinophils. Treg cells have an influence on B cells, suppress IgE production and induce the production of blocking type IgG4 antibodies against venom antigens [113].

appears to be less effective at providing future protection in comparison with other Hymenoptera. According to field stings or sting challenges after VIT, honeybee VIT provides approximately 75% to 85% protection from future stings, which is lower than the approximately 85% to 93% protection for yellow jacket VIT. Also, honeybee VIT is less safe, as the risk for systemic reactions during the full course of treatment ranges from 24% to 41% for honeybee, while this is only 5% to 25% for other Hymenoptera [106]. For many years honeybee venom was also used to treat bumblebee venom allergy, which was not always successful. Since the commercial availability of bumblebee venom extracts and the finding of bumblebee venom-specific IgE lacking cross-reactivity to honeybee venom, bumblebee venom is now the preferred choice for treatment of bumblebee venom allergy [108;114].

Nowadays, allergen extracts are used for immunotherapy. However, due to great variability in the amounts of individual allergens, these extracts are difficult to standardize

[115]. Lowly abundant but major allergens may be even missing due to downstream processing of extracts [116]. In addition, extracts can be contaminated with allergens from other sources or initiate new IgE specificities [115]. Therefore, the use of extracts for immunotherapy may cause the currently insufficient safety and efficacy of VIT. Recombinant allergens aim to overcome these issues as they can be produced in unlimited amounts, at highly standardized quality and with exact physiochemical and immunological properties. Allergens can be modified to have more favourable characteristics, including reduced IgE reactivity or enhanced immunogenicity [117]. Based on the knowledge of the allergen structures, several approaches have been developed, such as recombinant wild-type allergens, hypoallergens, T-cell epitope-based vaccines, carrier-bound peptides, genetic vaccination and gene therapy [115]. For treating honeybee venom allergy, a phase I clinical trial using T-cell epitopes of the Api m 1 allergen has been conducted [118]. Intact T-cell epitopes are required to enable the induction of specific T-cell tolerance. In contrast, IgE-binding B-cell epitopes are prerequisites for sensitisation against the allergen and therefore, their binding efficiency must be reduced. Bee venom allergic patients were treated with three long synthetic peptides encompassing the entire Api m 1 sequence in a rush desensitisation protocol to a maintenance dose of 100 μg. This treatment was safe and induced increases in T-cell proliferation, IFN-γ and IL-10 levels, but no Th2 cytokines. Also allergen-specific IgG4 levels increased, but not IgE levels. No severe adverse reactions were reported [117;118]. Also a prototype of a multi-allergen vaccine including assembled T-cell epitopes of three honeybee venom allergens (Api m 1, Api m 2 and Api m 3) showed a reduction of specific IgE development towards the native allergen in mice [119].

2.3 Hymenoptera venom allergy diagnosis

2.3.1 Conventional diagnosis

2.3.1.1 Principle and methods

A correct allergy diagnosis is required for the initiation of an appropriate immunotherapy. Currently, diagnosis of Hymenoptera venom allergy begins with assessing the clinical history (information on the severity of the reaction, number of stings, sting site, entomological identification,…) [99]. Subsequently, in conventional diagnosis, clinical suspicion is confirmed by several *in vitro* and *in vivo* techniques using venom extracts. In venom skin tests (skin prick or intradermal testing) a small amount of venom is introduced to the patient's skin and

the wheal and flare allergic reactions are measured [99]. In addition, the venom-specific IgE titers in the serum are measured. At present, two FDA-approved diagnostic tests are available [120]. Both ImmunoCAP FEIA (Phadia/Thermo Fisher Scientific, Uppsala, Sweden) and Immulite (Siemens Healthcare Diagnostics, Los Angeles, CA, USA) are enzyme-linked solid-phase immunoassays which give quantitative results of serum IgE levels in kU/L. For the ImmunoCAP FEIA system, the venom extracts of honeybee (*A. mellifera*), bumblebee (*B. terrestris*), common wasp (*Vespula spp.*), European paper wasp (*Polistes dominulus*), paper wasp (*Polistes spp.*), European hornet (*Vespa crabro*), white-faced hornet (*Dolichovespula maculata*) and yellow hornet (*Dolichovespula arenaria*) are available. Also the Immulite system provides several venom extracts (honeybee, wasp, paper wasp, white-faced hornet and yellow hornet). In cases where quantification of sIgE and venom skin tests remain negative or yield contradictory or equivocal results, the European Academy of Allergology and Clinical Immunology Interest Group on Insect Venom Hypersensitivity advises to use cellular tests, such as the basophil activation test (BAT), to demonstrate immunological sensitization [99]. Upon encounter of specific allergen that crosslinks FcεRI-bound IgE, basophils not only synthesize and secrete bioactive mediators, but also up-regulate the expression of certain activation markers that can be quantified flow-cytometrically in the BAT [121;122]. However, entrance of this technique in mainstream use is hampered as it is not always readily accessible and demands particular expertise [123].

2.3.1.2 Difficulties of conventional venom allergy diagnosis

A correct diagnosis is not always straightforward. For example, many patients fail to identify or name the hymenopteran species that stung. In addition, it has been demonstrated that quantification of venom-specific sIgE and venom skin tests generated entirely false-negative results in patients with a history of a severe venom allergy [123]. Besides, although the majority of patients is allergic to a single venom, many patients show double positive sIgE results to multiple Hymenoptera venoms, which relates to immunochemical cross-reactivity [123]. This cross-reactivity can occur on peptide basis due to the presence of similar protein allergens in both venoms, or from cross-reactive carbohydrate determinants (CCDs) which are IgE recognized carbohydrate moieties ubiquitous on many Hymenoptera venom glycoproteins [123].

Many hymenopteran venom proteins contain N- and O-linked glycosylation sites. In contrast to venom O-glycans which have so far not been investigated, the N-glycans of both honeybee venom allergens Api m 1 (PLA$_2$; [124;125]) and Api m 2 (hyaluronidase; [126]), and the wasp venom allergen Ves v 2 (hyaluronidase; [127]) have been well characterized. In total, fourteen N-glycans from honeybee venom Api m 1 were identified [124;125] (Figure 14) and those of Api m 2 and Ves v 2 were found to be very similar [126;127]. They are paucimannosidic and contain fucose α-1,3 and/or α-1,6 linked to the innermost *N*-acetylglucosamine (Figure 14). The α-1,3-core fucoses are currently the only known CCDs from Hymenoptera venoms. They have been found in the venom of *Apis mellifera* and *Vespula vulgaris*, while the venoms of all analyzed American (*P. annularis*, *P. fuscatus*, *P. metricus*, *P. apachus* and *P. exclamans)* and European (*P. dominulus*) *Polistes* species are CCD-free [128]. For bumblebee venom, no data about the presence of venom CCDs were found in literature.

Besides α-1,3-core fucoses, also β-1,2-core xylose and α-1,3-galactose residues linked to N-glycans are known to be CCDs. However, both are not found in insects. Glycosylations of plants and some pathogenic helminths contain both α-1,3-core fucoses and β-1,2-core xylose and therefore cross-reactions between Hymenoptera venoms and pollen, natural rubber latex, vegetables and fruits have been observed in serum investigations [129]. The diagnostic relevance of these structures has been described several times, but their clinical relevance is still discussed [130]. In contrast, a clear clinical relevance of the α-1,3-galactose epitope was confirmed [131]. This CCD is found on glycolipids and glycoproteins of non-primate mammals, prosimians and New World monkeys, but not in apes, Old World monkeys and humans. This CCD is involved in allergy to red meat [132].

In addition, Hymenoptera venoms have several allergens in common which contribute to cross-reactivity (Table 2). *A. mellifera* venom provides the best immunologically characterized model: 12 allergens have been reported, which are named Api m 1 until Api m 12 following the nomenclature guidelines of the International Union of Immunological Societies (IUIS; http://www.allergen.org/Allergen.aspx). In some areas across Asia, *A. mellifera*, *A. cerana* and *A. dorsata* coexist [133]. PLA$_2$ has been found to be an important allergen in these species and sequence identity is higher than 90%. The venom composition of *Apis* species other than *A. mellifera* is only poorly characterized and no other allergens
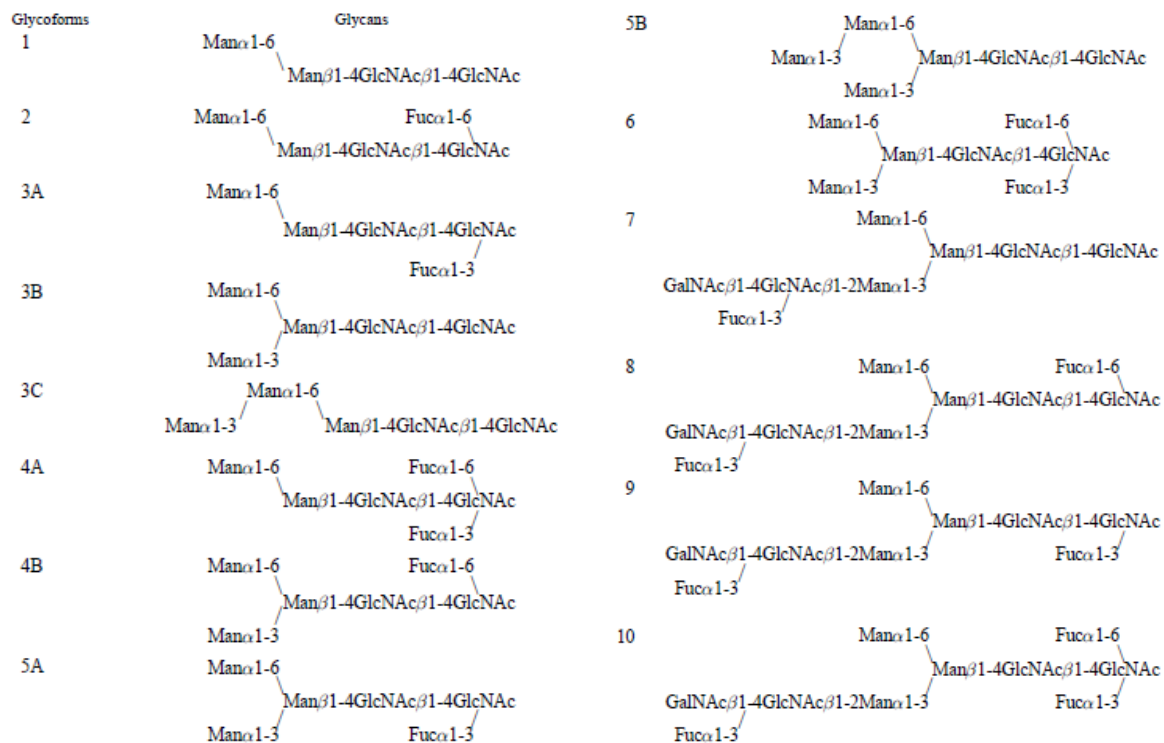
Figure 14: Structures of N-linked glycosylations from phospholipase A$_2$ (Api m 1) of honeybee venom. Figure adapted from [125].

have been identified, but probably honeybee species have a highly similar venom composition containing allergens with high sequence identity. No data are available about cross-reactivity between different honeybee venoms, but this is expected to be high.

Only few bumblebee venom allergens have been characterized (Table 2). Differences in IgE binding between venoms of the European *B. terrestris* and North-American *B. pennsylvanicus* have been reported [134;135]. Although sequence identity between their PLA$_2$ allergens is high (83.8%), they contain partially different IgE epitopes [66]. Also the casein hydrolyzing protease is a major allergen which shows only partial cross-reactivity between both species. Besides, several studies reported a high degree of cross-reactivity between honeybee and bumblebee venom [114]. It has been suggested that there exist two types of patients sensitised to bumblebee venom. The first type of patients has primary earlier exposure and sensitization to honeybee venom and contains IgE highly cross-reactive with honeybee venom. The second type of patients are specifically sensitised to bumblebee venom due to occupational exposure and exhibit IgE with low or absent cross-reactivity to honeybee venom [107]. Species-specific IgE epitopes exist, which is supported by the moderate sequence identity (52.9%) between the major honeybee and *B. terrestris* PLA$_2$

allergens. As treatment by VIT using honeybee venom failed in some bumblebee allergic patients, the use of bumblebee venom for immunotherapy is recommended [114]. This makes a decisive diagnosis necessary.

Vespula, *Dolichovespula* and *Vespa* are three genera belonging to the Vespinae subfamily. In Europe, *Vespula* species are the most important stinging wasps, also called yellow jackets in America. As they are attracted to protein and sugar foods and drinks, these scavenger species sting humans more often than honeybees. Different *Vespula* venoms strongly cross-react, while also substantial cross-reactivity between *Vespula, Vespa* and *Dolichovespula* venoms has been reported [99]. *Polistes* (paper wasps) and *Polybia* species belong to the Polistinae subfamily. *Polistes* species are especially found in the Mediterranean areas, while *Polybia* is a genus from South-America. Cross-reactivity of Vespinae with Polistinae is generally lower than cross-reactivity within the Vespinae. Also cross-reactivity between European species of *Polistes* (*P. dominulus*, *P. gallicus*) is very strong compared to cross-reactivity between European and American *Polistes* species [99]. PLA_1 and antigen 5 proteins are the major venom allergens of many wasp species (Table 2).

Also a small group of ant species is capable of stinging humans and these stings can cause allergic reactions [136]. Although these species are not found in Europe, they are sometimes spread by international transport of cargo. The most medically important aggressive ants are the fire ants of the genus *Solenopsis*. In contrast to the proteinaceous venoms of other Hymenoptera, ant venoms are mostly composed of alkaloids. Nevertheless, four protein allergens have been characterized, including a member of the antigen 5 family (Sol i 3). Sera from patients sensitized to Sol i 3 do not cross-react with wasp antigen 5. In contrast, the PLA_1 allergen, Sol i 1, exhibits cross-reactivity with wasp venom phospholipases. Sol i 2 and Sol i 4 have not been found in other venoms. Other venomous ants belong to the *Pachycondyla*, *Myrmecia* and *Pogonomyrmex* genera [136]. Recently, the genomes of three stinging ant species (*Solenopsis invicta*, *Harpegnathos saltator*, *Pogonomyrmex barbatus*) were sequenced [137-139], which should stimulate more studies on ant venom compositions and immunological characterization of the venom compounds.

Table 2: All Hymenoptera venom allergens from the official list of allergens of the IUIS (http://www.allergen.org/Allergen.aspx) are shown. Homologous allergens within each genus are presented per protein name.

Family	Genus	Protein name	Allergen name						
Apidae	*Apis*		*A. mellifera*	*A.cerana*	*A. dorsata*				
		Phospholipase A$_2$	Api m 1	Api c 1	Api d 1				
		Hyaluronidase	Api m 2						
		Acid phosphatase	Api m 3						
		Melittin	Api m 4						
		Dipeptidyl peptidase IV	Api m 5						
		Protease inhibitor	Api m 6						
		CUB serine protease	Api m 7						
		Carboxylesterase	Api m 8						
		Serine carboxypeptidase	Api m 9						
		Icarapin	Api m 10						
		MRJP8 and MRJP9	Api m 11						
		Vitellogenin	Api m 12						
	Bombus		*B. terrestris*	*B. pennsylvanicus*					
		Phospholipase A$_2$	Bom t 1	Bom p 1					
		Protease	Bom t 4	Bom p 4					
Vespidae	*Vespula*		*V. vulgaris*	*V. maculifrons*	*V.squamosa*	*V. flavopilosa*	*V. germanica*	*V. pensylvanica*	*V. vidua*
		Phospholipase A$_1$	Ves v 1	Ves m 1	Ves s 1				
		Hyaluronidase	Ves v 2	Ves m 2					
		Dipeptidyl peptidase IV	Ves v 3						
		Antigen 5	Ves v 5	Ves m 5	Ves s 5	Ves f 5	Ves g 5	Ves p 5	Ves vi 5
		Vitellogenin	Ves v 6						

Dolichovespula		D. maculata	D. arenaria
	Phospholipase A$_1$	Dol m 1	
	Hyaluroniase	Dol m 2	
	Antigen 5	Dol m 5	Dol a 5

Vespa		V. crabro	V. mandarinia	V. magnifica
	Phospholipase A$_1$	Vesp c 1	Vesp m 1	
	Hyaluronidase			Vesp ma 2
	Antigen 5	Vesp c 5	Vesp m 5	Vesp ma 5

Polistes		P. dominulus	P. exclamans	P. annularis	P. gallicus	P. fuscatus	P. metricus
	Phospholipase A$_1$	Pol d 1	Pol e 1	Pol a 1	Pol g 1		
	Hyaluronidase			Pol a 2			
	Serine protease	Pol d 4	Pol e 4				
	Antigen 5	Pol d 5	Pol e 5	Pol a 5	Pol g 5	Pol f 5	Pol m 5

Polybia		P. paulista	P. scutellaris
	Phospholipase A$_1$	Poly p 1	
	Antigen 5		Poly s 5

Formicidae	Myrmecia		M. pilosula
		Pilosulin-1	Myr p 1
		Pilosulin-3	Myr p 2
		Pilosulin-4	Myr p 3

Solenopsis		S. invicta	S. geminata	S. richteri	S. saevissima
	Phospholipase A$_1$	Sol i 1			
		Sol i 2	Sol g 2	Sol r 2	Sol s 2
	Cystein-rich venom protein	Sol i 3	Sol g 3	Sol r 3	Sol s 3
		Sol i 4	Sol g 4		

2.3.2. Component-resolved diagnosis

2.3.2.1 Principle and methods

As mentioned, conventional tests not always allow to establish correct diagnosis. During the last decade, component-resolved diagnosis (CRD) has entered the field of allergy diagnosis. In contrast to conventional sIgE assays, CRD relies on quantification of sIgE antibodies to single components, purified from natural sources or obtained by recombinant techniques. The use of species-specific unique marker components and cross-reactive determinants can help to distinguish between a true double sensitisation (patient needs immunotherapy with both allergens) and cross-sensitization to several unrelated allergen sources (immunotherapy restricted to sensitizing allergen) [140]. Analyzing sIgE to a well-chosen panel of allergens increases sensitivity and leads to a better discrimination between different allergies than diagnostic tests using extracts [141;142]. As obtaining high amounts of highly standardized allergens is crucial for diagnostic purposes, recombinantly produced allergens are preferred over purified allergens [142]. In addition, careful selection of the expression system allows to obtain allergens without confounding CCDs.

In addition to venom extracts, the ImmunoCAP FEIA and Immulite immunoassays also provide several recombinant CCD-free venom allergens which allow to perform CRD. rVes v 1, rVes v 5, rApi m 1 and rPol d 1 (expression system is not clarified by the manufacturer) can be used for the ImmunoCAP FEIA system, while the Immulite system provides rApi m 1, rApi m 2 and rVes v 5 (produced in Sf9 insect cell line). Each assay requires 40 µl (ImmunoCAP FEIA) or 5µl (Immulite) of serum to test a single allergen. In contrast, sIgE can now be simultaneously determined towards more than 100 allergen compounds of many sources by use of microarray technology, which needs only a very small quantity (20 µl) of serum (ImmunoCAP ISAC, Phadia/Thermo Fisher Scientific; request for FDA approval under way) (Figure 15). This enzyme-linked immunoassay generates semi-quantitative results (ISAC standardised units) [120]. While venom extracts are lacking, rApi m 1, nApi m 4, rVes v 5 and rPol d 5 are included hymenopteran venom allergens. However, this test is almost never applied in venom allergy diagnosis as the spotted selection of components often does not allow a decisive diagnosis.

Figure 15: The ImmunoCAP ISAC (Phadia) test is a microarray-based diagnostic test. Over 100 purified and recombinant allergens are immobilized on the array. IgE antibodies from the patient serum bind to specific allergens. Next, allergen-bound IgEs are recognized by a fluorescently-labeled antibody. The generated fluorescent signal is detected by a microarray scanner. Figure from Phadia.

As CRD can help to distinguish between different Hymenoptera venom allergies in many patients [143], well studied venom proteomes and allergen repertoires of a broad range of stinging species are required for the development of such diagnostic tools. Although for some species several major allergens have been characterized, an in-depth knowledge of their venom composition is often lacking. Besides *A. mellifera*, the venom proteomes of several hymenopteran species known to sting humans have been studied by mass spectrometry (Table S4). However, most studies identified only a limited number of venom proteins. In-depth proteomic insights have only been obtained for the venoms of the neotropical social wasps *Agelaia pallipes pallipes* and *Polybia paulista*, and the red imported fire ant *Solenopsis invicta*. However, the allergenic potential of only few of the identified venom compounds has been determined (Table 2).

2.3.2.2 Distinguishing between honeybee and wasp venom allergy using CRD

In Europe, approximately two thirds of patients with allergy to Hymenoptera venom react to wasp stings and one third to bee stings. Therefore, obtaining a correct diagnosis between both culprit species is important. However, of patients with systemic allergic reactions to Hymenoptera stings, up to 59% have serum-specific IgE antibodies to venoms of both honeybee and wasp [144]. This double positivity can be partially explained by IgE reactivity to CCDs present on both honeybee and wasp venom proteins. Indeed, in our region

sensitization to CCD is found in about 20% of the patients with hymenoptera venom allergy, particularly honeybee venom allergy [145]. In addition, both venoms have several allergens in common. The hyaluronidase enzyme has long been recognised as the most relevant cross-reactive allergen. However, cross-reactivity between Ves v 2 and Api m 2 is mainly induced by CCDs and less often because of shared peptide epitopes [146]. Moreover, Ves v 2 was found to be only a minor wasp venom allergen as it is IgE recognized by only 10-15% of the wasp allergic patient. Recently, additional honeybee and *Vespula vulgaris* venom homologues have been discovered, which may be responsible for cross-reactivity between both venoms. The dipeptidyl peptidase IV allergens of *A. mellifera* (Api m 5) and *V. vulgaris* (Ves v 3) share 53.8% sequence identity. A recombinant anti-Api m 5 human monoclonal IgE antibody reacts to a similar extent with both CCD-free Api m 5 and Ves v 3. This suggests the presence of a conserved protein epitope, which may also be recognized by IgE of venom allergic patients [48]. Also, Blank and co-workers [50] very recently provided indications for cross-reactivity between the venom vitellogenins (Api m 12 and Ves v 6).

Conventional tests are used as the first line of laboratory investigation, but not always allow to distinguish between honeybee and wasp venom allergy. For patients in which the culprit insect is uncertain, and/or double-positive results are obtained with conventional venom extracts, the second-line analysis of IgE to available CCD-free, species-specific recombinant allergens has been found to be helpful in the identification of the relevant sensitization [123;143;147-149]. In patients with allergy to wasp venom, the diagnostic sensitivity of a combination of the currently available wasp venom allergens rVes v 5 and rVes v 1 has been reported to be as high as 92% to 96% [123;141;149]. In contrast, the diagnosis of honeybee venom allergy using solely Api m 1 lacks sensitivity as, depending on the patient population, between 20% and 42% of the patients lacks IgE reactivity to rApi m 1 [141;144;148-150]. From ImmunoCAP data, it can be suggested that sensitisation to multiple honeybee venom allergens is common [144]. For example, in a large patient population (82 honeybee venom allergic patients), it was shown that 75.6% of the patients was sensitized to rApi m 1, 46.3% to rApi m 2, and 26.8% to nApi m 4. Moreover, these data demonstrated that the combination of ImmunoCAPs with Api m 1, Api m 2 and Api m 4 increased the sensitivity of CRD to 89%. It is clear that additional honeybee venom allergens are necessary to further increase sensitivity and allow a better discrimination of bee and wasp venom allergy [144].

As the major honeybee venom allergen rApi m 10 was not recognized by serum IgE of wasp venom allergic patients [116], this is an interesting candidate. The immunological characterization of additional honeybee venom compounds may identify novel species-specific major allergens that allow to develop diagnostic tests with improved sensitivity and specificity. A multiplex diagnostic test, which screens for IgE recognition of multiple honeybee- and wasp-specific venom allergens, should be developed to allow to distinguish between honeybee and wasp venom allergy using a limited amount of serum.

In conventional diagnosis, quantification of venom-specific sIgE sometimes generates false-negative results. Recently, it has been demonstrated that spiking these venom extracts with recombinant venom allergens can increase the sensitivity of these diagnostic tests. Indeed, supplementing wasp venom extract with the major allergen rVes v 5 improved ImmunoCAP sensitivity and allowed a correct diagnosis of wasp venom allergy in patients sensitized to Ves v 5 but demonstrating a negative sIgE to wasp venom [123;151].

2.4 Identification and characterization of honeybee and bumblebee venom allergens

The characterization of novel allergens is an important step towards better diagnostics and immunotherapy. Compounds are incorporated into the IUIS official list of allergens in case IgE binding is demonstrated by 5 sera of patients allergic to the respective allergen source or in case of IgE binding by at least 5% of all tested sera of patients allergic to the respective allergen source. The identification of novel allergens is often performed by 2D-PAGE separation of the source extract, followed by immunoblotting using sera of allergic patients and protein identification of IgE recognized spots by mass spectrometry. Alternatively, individual proteins are purified or produced as recombinants and their IgE recognition is analyzed by ELISA, spot blot or Western blot. Next, in vitro cellular tests such as the basophil activation test (BAT) allow to analyze if IgE recognized compounds can stimulate basophils to release allergenic mediators. Allergens require at least two epitopes to cross-link the high affinity IgE receptor (FcεRI), which causes basophil activation. Once an allergen has been identified, IgE mapping studies using microarrays can reveal relevant information about the antigen structure and epitopes, and the patient's immune response [152].

Several venom allergens of bees, wasps, paper wasps, hornets, and ants have been identified (Table 2). The immunological characterization of novel venom proteins requires large quantities of pure venom proteins, which are often difficult to obtain using purification strategies [153]. As recombinant production solves these issues, recombinantly produced compounds are the favoured choice for immunological characterization. In addition, this technology allows to select an expression host which produces recombinants containing the preferred post-translational modifications. As the clinical relevance of CCDs is still controversial, Hymenoptera venom proteins are preferably obtained without CCDs, which allows to evaluate the allergenicity at the mere protein level. Several cell lines are available which allow the recombinant production of CCD-lacking proteins. However, prokaryiotically produced, non-glycosylated proteins may suffer limitations regarding folding, solubility, activity and IgE epitope conservation, while yeasts and mammalian cell lines produce aberrant glycosylations [154]. Insect cell lines provide expression hosts that are phylogenetically as close as possible to the parental organism, making them indispensable for recombinant insect venom protein production. In addition, expression in insect cells mostly results in secretion of biologically active and soluble proteins, usually in glycosylated form. This added glycosylation is more authentic than that from other hosts [155], although variation between different insect cell types has been noticed. Seismann and co-workers [156] showed that HighFive (*Trichoplusia ni*) and Sf9 (*Spodoptera frugiperda*) cells both produce N-linked glycosylations, but HighFive glycosylations include α-1,3-core fucosylations which are lacking in Sf9 cells. These fucose residues are the only known CCDs of Hymenoptera venoms. Consequently, the evaluation of the allergenicity of novel hymenopteran venom proteins is preferably executed using the baculovirus-mediated infection of Sf9 insect cells, as this cell line produces the natural insect-specific post-translational modifications, but without CCDs which interfere with the identification of proteinous epitopes [156].

The list of honeybee venom allergens (Table 2) includes both major and minor allergens. When the majority (>50%) of the tested population reacts to an allergen, it is described as a major allergen, whereas minor allergens are recognized by a limited number of patients. Although melittin is the highest abundant honeybee venom compound, it is only a minor allergen (Api m 4), active in less than one third of bee allergic patients. As melittin is a highly

abundant, non-glycosylated peptide, it was only immunologically characterized in its natural form. The allergenicity of most other allergens has been determined using the recombinant form of these proteins. However, IgE recognition of individual allergens differs markedly between different studies, which may be explained by variable inclusion criteria of patients, geographical differences, differences in the immunoassay parameters and biochemical properties of the recombinants. For example, for Api m 1 (PLA$_2$) which is the most important allergen in honeybee venom, a large difference in the frequency of Api m 1 sensitization (56.7-97%) is seen between different studies. Previously, it was suggested that sensitivity might depend on the inclusion criteria of patients. For example, patients without detectable sIgE to bee venom and negative skin test results are more likely to be negative for Api m 1. However, this variation has recently been attributed to geographical factors, as a north-south difference in Api m 1 sensitization was demonstrated, with highest levels in Northern Europe (Figure 16) [157]. This observation may be the effect of a variable venom composition. Several other allergens have been recombinantly expressed in *E. coli* and/or Sf9 insect cells, which enabled to evaluate IgE recognition beyond CCD-reactivity. Api m 2 [156], Api m 5 [48], Api m 10 [116], Api m 11 [158] and Api m 12 [50] are typical glycosylated allergens with allergenic relevance beyond their carbohydrate epitopes. As Api m 6 is a non-glycosylated protein, it was immunologically characterized as a bacterial recombinant, which revealed that is a minor allergen [49]. In contrast, IgE recognition of Api m 3 has only been examined as a purified venom protein [153] and as a recombinant produced in the HighFive cell line which adds CCDs [159]. Therefore, its allergenicity should be confirmed by production of a CCD-lacking recombinant. Api m 7, Api m 8 and Api m 9 still need to be immunologically characterized, although they have been added to the official list of honeybee allergens.

The protein structure of several honeybee venom allergens has been determined by X-ray crystallography (Api m 1, Api m 2, Api m 4) [6], while for others (Api m 3, Api m 6 and Api m 7) structures have been determined indirectly by homology modelling based on the known crystallographic structures of related proteins [47;49;160]. Available protein structures are valuable information to predict the location of linear and conformational IgE epitopes using bioinformatics. For IgE binding, surface lysyl residues have been observed to be essential, while linear antigenic determinants should be accessible to the solvent, contain both
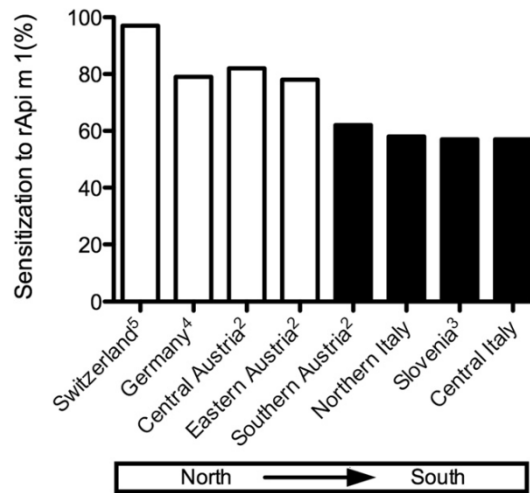
Figure 16: North-south difference in Api m 1 sensitisation in Europe [157].

hydrophobic and hydrophilic residues and preferably be present in loops, avoiding helical regions [160]. The IgE-binding epitopes of Api m 3, Api m 6 and Api m 7 were predicted based on their determined structure. Also peptide arrays using overlapping peptides spanning the complete protein can be used to identify the linear epitopes but these have not been executed for honeybee venom allergens.

Until now, all immunologically characterized honeybee venom compounds were shown to be recognized by IgE antibodies, except the lowly abundant C1q-like venom protein [55]. A preliminary test using recombinant C1q failed to demonstrate IgE recognition by serum from patients with a documented severe honeybee/wasp venom allergy. However, C1q was produced as an insoluble, non-glycosylated recombinant protein in a prokaryotic expression system. As the lack of post-translational modifications may cause an incorrect folding of the bacterial recombinant, its observed lack of IgE recognition may not correspond to that of the natural counterpart. In addition, IgE recognition of C1q was analyzed with sera of only a very limited collection of honeybee and wasp venom allergic patients [55]. Therefore, further research should determine the allergenic nature of this compound.

Additionally, several proteins known to be present in honeybee venom for several years have never been immunologically characterized. Two compounds, PVF1 and hexamerin 70A, were found in venom proteomic studies of 2005 [53; also Figure 5] and 2006 [82] (Table S1) respectively and are interesting candidates for immunological

characterization. Moreover, novel proteomic approaches may reveal additional honeybee venom compounds with allergenic potency.

Allergen protein heterogeneity may make this picture even more complex. Indeed, protein heterogeneity has been reported to be immunologically relevant. For example, the birch genome contains at least 7 pollen-expressed genes that encode distinct Bet v 1 isoforms with varying IgE reactivity [161;162]. Also in honeybee venom, different isoforms of allergens have been described. The four described Api m 6 isoforms differ in their primary structure at the amino and carboxy terminus by a maximum of six amino acids. Immunoblot analyses revealed no isoform-specific IgE [57]. Also two highly similar Api m 10 (icarapin) alternative splice variants were identified (Figure 6) [58], which were both found to be IgE recognized [58;116]. However, as mentioned, 2D-PAGE separation of honeybee venom revealed additional icarapin protein spots (Figure 5) [53]. This strongly suggests that additional Api m 10 isoforms exist. As isoforms differ by their protein sequence and/or conformation, IgE reactivity between different isoforms may vary and result in a variable allergenicity.

So far, for the bumblebee *B. terrestris*, only two venom allergens have been added to the IUIS list (Table 2). Both the PLA$_2$ (Bom t 1) and casein hydrolyzing protease (Bom t 4) were purified from the venom and were shown to be IgE recognized by six sera of occupationally sensitized patients [66]. Additional bumblebee allergens are expected to exist. However, as bumblebee venom allergic patients are still quite rare, further research will be confined by the limited availability of blood samples required for immunological characterization of newly identified venom compounds.

3. ADDENDUM

Supplementary tables can be found on the included CD-ROM or can be requested by e-mail from matthiasvanvaerenbergh@hotmail.com and Dirk.deGraaf@UGent.be.

Table S1 shows all honeybee (*Apis mellifera*) venom compounds described in literature. Allergen names and GenBank accession numbers are presented. In addition, all identification methods used to identify these compounds are included. Sampling method: the methods

used to collect the venom sample (MM= manual milking; EM: electrical milking; VG= venom gland tissue; ND: not defined). Separation methods: methods applied to separate the venom proteins. Identification methods: technology used to identify the venom proteins. NA= no data available in literature.

Table S2 presents all identified venom compounds of *Apis* species other than *Apis mellifera*. Allergen names and GenBank accession numbers are shown. Venom protein evidence has been obtained for only few compounds. For others, venom gland transcript data have been sequenced. NA= no data available in literature.

Table S3 shows all identified venom compounds of bumblebee species. Allergen names and GenBank accession numbers are shown. Separation methods: methods applied to separate the venom proteins. Identification methods: technology used to identify the venom proteins or transcripts. NA= no data available in literature.

Table S4 presents all proteomics studies of venoms of hymenopteran species known to sting humans (honeybee and bumblebees not included). Species names, protein identification methods and the results are shown. Table adapted from [163].

4. REFERENCES

[1]   Wong ESW, Belov K. Venom evolution through gene duplications. Gene 2012 Mar 15;496(1):1-7.
[2]   Schmidt JO. Venoms and Toxins in Insects. In: Capinera JL, editor. Encyclopedia of Entomology . 2008. p. 4076-89.
[3]   Moreau SJM. "It stings a bit but it cleans well": Venoms of Hymenoptera and their antimicrobial potential. Journal of Insect Physiology 2013 Feb;59(2):186-204.
[4]   Sharkey MJ. Phylogeny and classification of Hymenoptera. Zootaxa 2007 Dec 21;(1668):521-48.
[5]   de Graaf DC, Aerts M, Brunain M, Desjardins CA, Jacobs FJ, Werren JH, et al. Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies. Insect Mol Biol 2010 Feb;19 Suppl 1:11-26.
[6]   King TP, Spangfort MD. Structure and biology of stinging insect venom allergens. International Archives of Allergy and Immunology 2000 Oct;123(2):99-106.
[7]   de Graaf DC, 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 Mar 6;72(2):145-54.
[8]   Starr CK. Enabling Mechanisms in the Origin of Sociality in the Hymenoptera-the Stings the Thing. Annals of the Entomological Society of America 1985;78(6):836-40.
[9]   Kerr WE, De Lello E. Sting glands in stingless bees- A vestigial character (Hymenoptera: Apidae). Journal of the New York Entomological Society 1962;70(4):190-214.

[10] Cardinal S, Straka J, Danforth BN. Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. Proceedings of the National Academy of Sciences of the United States of America 2010 Sep 14;107(37):16207-11.

[11] Stolle E, Wilfert L, Schmid-Hempel R, Schmid-Hempel P, Kube M, Reinhardt R, et al. A second generation genetic map of the bumblebee *Bombus terrestris* (Linnaeus, 1758) reveals slow genome and chromosome evolution in the Apidae. Bmc Genomics 2011 Jan 19;12.

[12] Bridges AR, Owen MD. The Morphology of the Honey Bee (Apis-Mellifera L) Venom Gland and Reservoir. Journal of Morphology 1984;181(1):69-86.

[13] Schumacher MJ, Tveten MS, Egen NB. Rate and Quantity of Delivery of Venom from Honeybee Stings. Journal of Allergy and Clinical Immunology 1994 May;93(5):831-5.

[14] Harano K, Obara Y. The role of chemical and acoustical stimuli in selective queen cell destruction by virgin queens of the honeybee *Apis mellifera* (Hymenoptera: Apidae). Applied Entomology and Zoology 2004 Nov;39(4):611-6.

[15] Roat TC, Nocelli RCF, Landim CD. The venom gland of queens of *Apis mellifera* (Hymenoptera, Apidae): morphology and secretory cycle. Micron 2006;37(8):717-23.

[16] Schmidt JO. Toxinology of Venoms from the Honeybee Genus Apis. Toxicon 1995 Jul;33(7):917-27.

[17] Nocelli RCF, Roat TC, Cruz-Landim C. Alterations induced by the juvenile hormone in glandular cells of the *Apis mellifera* venom gland: Applications on newly emerged workers (Hymenoptera, Apidae). Micron 2007;38(1):74-80.

[18] Ferreira RS, Sciani JM, Marques-Porto R, Lourenco A, Orsi RD, Barraviera B, et al. Africanized honey bee (*Apis mellifera*) venom profiling: Seasonal variation of melittin and phospholipase A(2) levels. Toxicon 2010 Sep 1;56(3):355-62.

[19] Hoffman DR, Jacobson RS. Allergens in Hymenoptera Venom.12. How Much Protein Is in A Sting. Annals of Allergy 1984;52(4):276-8.

[20] Steen CJ, Janniger CK, Schutzer SE, Schwartz RA. Insect sting reactions to bees, wasps, and ants. International Journal of Dermatology 2005 Feb;44(2):91-4.

[21] Alaux C, Sinha S, Hasadsri L, Hunt GJ, Guzman-Novoa E, DeGrandi-Hoffman G, et al. Honey bee aggression supports a link between gene regulation and behavioral evolution. Proceedings of the National Academy of Sciences of the United States of America 2009 Sep 8;106(36):15400-5.

[22] Prado M, Solano-Trejos G, Lomonte B. Acute physiopathological effects of honeybee (*Apis mellifera*) envenoming by subcutaneous route in a mouse model. Toxicon 2010 Nov;56(6):1007-17.

[23] Schumacher MJ, Schmidt JO, Egen NB, Lowry JE. Quantity, Analysis, and Lethality of European and Africanized Honey-Bee Venoms. American Journal of Tropical Medicine and Hygiene 1990 Jul;43(1):79-86.

[24] Funari SRC, Zeidler PR, Rocha HC, Sforcin JM. Venom production by africanized honeybees (*Apis mellifera*) and africanized-european hybrids. Journal of Venomous Animals and Toxins 2001;7(2):190-8.

[25] Vetter RS, Visscher PK, Camazine S. Mass envenomations by honey bees and wasps. Western Journal of Medicine 1999 Apr;170(4):223-7.

[26] Betten DP, Richardson WH, Tong TC, Clark RF. Massive honey bee envenomation-induced rhabdomyolysis in an adolescent. Pediatrics 2006 Jan;117(1):231-5.

[27] Jones RGA, Corteling RL, To HP, Bhogal G, Landon J. A novel Fab-based antivenom for the treatment of mass bee attacks. American Journal of Tropical Medicine and Hygiene 1999 Sep;61(3):361-6.

[28] Funayama JC, Pucca MB, Roncolato EC, Bertolini TB, Campos LB, Barbosa JE. Production of Human Antibody Fragments Binding to Melittin and Phospholipase A2 in Africanised Bee Venom: Minimising Venom Toxicity. Basic & Clinical Pharmacology & Toxicology 2012 Mar;110(3):290-7.

[29] Schmidt JO, Blum MS, Overal WL. Comparative Lethality of Venoms from Stinging Hymenoptera. Toxicon 1980;18(4):469-74.

[30] Ali MAASM. Studies on bee venom and its medical uses. International Journal of Advancements in Research & Technology 2012;1(2):1-14.

[31] Zalat S, Nabil Z, Hussein A, Rahka M. Biochemical and haematological studies of some solitary and social bee venoms. Egyptian Journal of Biology 1999;1:57-71.

[32] Hider RC. Honeybee Venom - A Rich Source of Pharmacologically Active Peptides. Endeavour 1988;12(2):60-5.

[33] Peiren N, de Graaf DC, Vanrobaeys F, Danneels EL, Devreese B, Van BJ, et al. Proteomic analysis of the honey bee worker venom gland focusing on the mechanisms of protection against tissue damage. Toxicon 2008 Jul;52(1):72-83.

[34] Du YR, Xiao Y, Lu ZM, Ding J, Xie F, Fu H, et al. Melittin activates TRPV1 receptors in primary nociceptive sensory neurons via the phospholipase A2 cascade pathways. Biochemical and Biophysical Research Communications 2011 Apr 29;408(1):32-7.

[35] Palm NW, Medzhitov R. Role of the inflammasome in defense against venoms. Proceedings of the National Academy of Sciences of the United States of America 2013 Jan 29;110(5):1809-14.

[36] Baracchi D, Turillazzi S. Differences in venom and cuticular peptides in individuals of *Apis mellifera* (Hymenoptera: Apidae) determined by MALDI-TOF MS. J Insect Physiol 2010 Apr;56(4):366-75.

[37] Baracchi D, Francese S, Turillazzi S. Beyond the antipredatory defence: honey bee venom function as a component of social immunity. Toxicon 2011 Nov;58(6-7):550-7.

[38] Ziai MR, Russek S, Wang HC, Beer B, Blume AJ. Mast cell degranulating peptide: a multi-functional neurotoxin. J Pharm Pharmacol 1990 Jul;42(7):457-61.

[39] Baracchi D, Mazza G, Michelucci E, Pieraccini G, Turillazzi S, Moneti G. Top-down sequencing of *Apis dorsata* apamin by MALDI-TOF MS and evidence of its inactivity against microorganisms. Toxicon 2013.

[40] Kreil G. Hyaluronidases--a group of neglected enzymes. Protein Sci 1995 Sep;4(9):1666-9.

[41] Dhananjaya BL, D'Souza CJM. The Pharmacological Role of Phosphatases (Acid and Alkaline Phosphomonoesterases) in Snake Venoms related to Release of Purines - a Multitoxin. Basic & Clinical Pharmacology & Toxicology 2011 Feb;108(2):79-83.

[42] Francischetti IM, My-Pham V, Harrison J, Garfield MK, Ribeiro JM. *Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. Gene 2004 Aug 4;337:55-69.

[43] Yan S, Cui F, Qiao C. Structure, function and applications of carboxylesterases from insects for insecticide resistance. Protein Pept Lett 2009;16(10):1181-8.

[44] Mittapalli O, Wise IL, Shukle RH. Characterization of a serine carboxypeptidase in the salivary glands and fat body of the orange wheat blossom midge, *Sitodiplosis mosellana* (Diptera: Cecidomyiidae). Insect Biochem Mol Biol 2006 Feb;36(2):154-60.

[45] Qiu Y, Choo YM, Yoon HJ, Jia JM, Cui Z, Wang D, et al. Fibrin(ogen)olytic activity of bumblebee venom serine protease. Toxicology and Applied Pharmacology 2011 Sep 1;255(2):207-13.

[46] Choo YM, Lee KS, Yoon HJ, Kim BY, Sohn MR, Roh JY, et al. Dual function of a bee venom serine protease: prophenoloxidase-activating factor in arthropods and fibrin(ogen)olytic enzyme in mammals. PLoS ONE 2010;5(5):e10393.

[47] Georgieva D, Greunke K, Betzel C. Three-dimensional model of the honeybee venom allergen Api m 7: structural and functional insights. Molecular Biosystems 2010;6(6):1056-60.

[48] Blank S, Seismann H, Bockisch B, Braren I, Cifuentes L, McIntyre M, et al. Identification, recombinant expression, and characterization of the 100 kDa high molecular weight Hymenoptera venom allergens Api m 5 and Ves v 3. J Immunol 2010 May 1;184(9):5403-13.

[49] Michel Y, McIntyre M, Ginglinger H, Ollert M, Cifuentes L, Blank S, et al. The Putative Serine Protease Inhibitor Api m 6 From *Apis mellifera* Venom: Recombinant and Structural Evaluation. Journal of Investigational Allergology and Clinical Immunology 2012;22(7):476-84.

[50] Blank S, Seismann H, McIntyre M, Ollert M, Wolf S, Bantleon FI, et al. Vitellogenins are new high molecular weight components and allergens (Api m 12 and Ves v 6) of *Apis mellifera* and *Vespula vulgaris* venom. PLoS ONE 2013;8(4):e62009.

[51] Martins JR, Nunes FMF, Cristino AS, Simoes ZP, Bitondi MMG. The four hexamerin genes in the honey bee: structure, molecular evolution and function deduced from expression patterns in queens, workers and drones. Bmc Molecular Biology 2010 Mar 26;11.

[52] Martins JR, Anhezini L, Dallacqua RP, Simoes ZLP, Bitondi MMG. A Honey Bee Hexamerin, HEX 70a, Is Likely to Play an Intranuclear Role in Developing and Mature Ovarioles and Testioles. PLoS ONE 2011 Dec 9;6(12):e29006.

[53] Peiren N, Vanrobaeys F, de Graaf DC, Devreese B, Van BJ, Jacobs FJ. The protein composition of honeybee venom reconsidered by a proteomic approach. Biochim Biophys Acta 2005 Aug 31;1752(1):1-5.

[54] Vlasak R, Kreil G. Nucleotide-Sequence of Cloned Cdnas Coding for Preprosecapin, A Major Product of Queen-Bee Venom Glands. European Journal of Biochemistry 1984;145(2):279-82.

[55] de Graaf DC, Brunain M, Scharlaken B, Peiren N, Devreese B, Ebo DG, et al. Two novel proteins expressed by the venom glands of *Apis mellifera* and *Nasonia vitripennis* share an ancient C1q-like domain. Insect Mol Biol 2010 Feb;19 Suppl 1:1-10.

[56] Peiren N, de Graaf DC, Evans JD, Jacobs FJ. Genomic and transcriptional analysis of protein heterogeneity of the honeybee venom allergen Api m 6. Insect Molecular Biology 2006 Oct;15(5):577-81.

[57] Kettner A, Hughes GJ, Frutiger S, Astori M, Roggero M, Spertini F, et al. Api m 6: A new bee venom allergen. Journal of Allergy and Clinical Immunology 2001 May;107(5):914-20.

[58] Peiren N, de Graaf DC, Brunain M, Bridts CH, Ebo DG, Stevens WJ, et al. Molecular cloning and expression of icarapin, a novel IgE-binding bee venom protein. FEBS Lett 2006 Sep 4;580(20):4895-9.

[59] Altmann F, Kubelka V, Staudacher E, Uhl K, Marz L. Characterization of the Isoforms of Phospholipase-A(2) from Honeybee Venom. Insect Biochemistry 1991;21(5):467-72.

[60] Blank S, Michel Y, Seismann H, Plum M, Greunke K, Grunwald T, et al. Evaluation of Different Glycoforms of Honeybee Venom Major Allergen Phospholipase A2 (Api m 1) Produced in Insect Cells. Protein and Peptide Letters 2011 Apr;18(4):415-22.

[61] Stöcklin R, Favreau P. Proteomics of venom peptides. In: Ménez A, editor. Perspectives in molecular toxinology.Chichester: John Wiley & Sons Ltd; 2002. p. 107-23.

[62] Owen MD, Sloley BD. 5-Hydroxytryptamine in the venom of the honey bee (*Apis mellifera* L.): variation with season and with insect age. Toxicon 1988;26(6):577-81.

[63] Owen MD, BRAIDWOO.JL. Quantitative and Temporal Study of Histamine and Histidine in Honey Bee (*Apis-mellifera*-L) Venom. Canadian Journal of Zoology-Revue Canadienne de Zoologie 1974;52(3):387-92.

[64] Owen MD, Bridges AR. Catecholamines in Honey Bee (*Apis-mellifera* L) and Various Vespid (Hymenoptera) Venoms. Toxicon 1982;20(6):1075-&.

[65] Chen J, Lariviere WR. The nociceptive and anti-nociceptive effects of bee venom injection and therapy: A double-edged sword. Progress in Neurobiology 2010 Oct;92(2):151-83.

[66] Hoffman DR, El-Choufani SE, Smith MM, de GH. Occupational allergy to bumblebees: allergens of *Bombus terrestris*. J Allergy Clin Immunol 2001 Nov;108(5):855-60.

[67] Qiu Y, Choo YM, Yoon HJ, Jin BR. Molecular cloning and antibacterial activity of bombolitin isolated from the venom of a bumblebee, *Bombus terrestris*. Journal of Asia-Pacific Entomology 2012 Mar;15(1):21-5.

[68] Qiu Y, Lee KS, Choo YM, Kong DX, Yoon HJ, Jin BR. Molecular cloning and antifibrinolytic activity of a serine protease inhibitor from bumblebee (*Bombus terrestris*) venom. Toxicon 2013 Mar 1;63:1-6.

[69] Xin Y, Choo YM, Hu Z, Lee KS, Yoon HJ, Cui Z, et al. Molecular cloning and characterization of a venom phospholipase A(2) from the bumblebee Bombus ignitus. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 2009 Oct;154(2):195-202.

[70] Qiu Y, Choo YM, Yoon HJ, Jin BR. Molecular cloning and fibrin(ogen)olytic activity of a bumblebee (*Bombus hypocrita sapporoensis*) venom serine protease. Journal of Asia-Pacific Entomology 2012 Mar;15(1):79-82.

[71] Argiolas A, Pisano JJ. Bombolitins, A New Class of Mast-Cell Degranulating Peptides from the Venom of the Bumblebee *Megabombus-Pennsylvanicus*. Journal of Biological Chemistry 1985;260(3):1437-44.

[72] Choo YM, Lee KS, Yoon HJ, Je YH, Lee SW, Sohn HD, et al. Molecular cloning and antimicrobial activity of bombolitin, a component of bumblebee Bombus ignitus venom. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 2010 Jul;156(3):168-73.

[73] Favreau P, Menin L, Michalet S, Perret F, Cheyneval O, Stocklin M, et al. Mass spectrometry strategies for venom mapping and peptide sequencing from crude venoms: Case applications with single arthropod specimen. Toxicon 2006 May;47(6):676-87.

[74] Argiolas A, Herring P, Pisano JJ. Amino-Acid-Sequence of Bumblebee Mcd Peptide - A New Mast-Cell Degranulating Peptide from the Venom of the Bumblebee *Megabombus-pennsylvanicus*. Peptides 1985;6:431-6.

[75] Choo YM, Yoon HJ, Jin BR. Effects of the bumblebee (*Bombus ignitus*) venom serine protease inhibitor on serine protease and phospholipase A(2) of *B-ignitus* venom. Journal of Asia-Pacific Entomology 2012 Dec;15(4):543-5.

[76] Martin SJ, Dils V, Billen J. Morphology of the Dufour gland within the honey bee sting gland complex. Apidologie 2005 Oct;36(4):543-6.

[77] Lowy PH, Sarmiento L, Mitchell HK. Polypeptides minimine and melittin from bee venom: effects on *Drosophila*. Arch Biochem Biophys 1971 Jul;145(1):338-43.

[78] Vick JA, Shipman WH, Brooks R, Jr. Beta adrenergic and anti-arrhythmic effects of cardiopep, a newly isolated substance from whole bee venom. Toxicon 1974 Mar;12(2):139-44.

[79] Shkenderov S, Koburova K. Adolapin--a newly isolated analgetic and anti-inflammatory polypeptide from bee venom. Toxicon 1982;20(1):317-21.

[80] Santos LD, Pieroni M, Menegasso ARS, Pinto JRAS, Palma MS. A new scenario of bioprospecting of Hymenoptera venoms through proteomic approach. Journal of Venomous Animals and Toxins Including Tropical Diseases 2011;17(4):364-77.

[81] Weinstock GM, Robinson GE, Gibbs RA, Worley KC, Evans JD, Maleszka R, et al. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 2006 Oct 26;443(7114):931-49.

[82] Hoffman DR. Hymenoptera venom allergens. Clin Rev Allergy Immunol 2006 Apr;30(2):109-28.

[83] Matysiak J, Schmelzer CE, Neubert RH, Kokot ZJ. Characterization of honeybee venom by MALDI-TOF and nanoESI-QqTOF mass spectrometry. J Pharm Biomed Anal 2011 Jan 25;54(2):273-8.

[84] Calvete JJ, Fasoli E, Sanz L, Boschetti E, Righetti PG. Exploring the Venom Proteome of the Western Diamondback Rattlesnake, *Crotalus atrox*, via Snake Venomics and Combinatorial Peptide Ligand Library Approaches. Journal of Proteome Research 2009 Jun;8(6):3055-67.

[85] Fasoli E, Sanz L, Wagstaff S, Harrison RA, Righetti PG, Calvete JJ. Exploring the venom proteome of the African puff adder, *Bitis arietans*, using a combinatorial peptide ligand library approach at different pHs. J Proteomics 2010 Mar 10;73(5):932-42.

[86] Fasoli E, Farinazzo A, Sun CJ, Kravchuk AV, Guerrier L, Fortis F, et al. Interaction among proteins and peptide libraries in proteome analysis: pH involvement for a larger capture of species. Journal of Proteomics 2010 Feb 10;73(4):733-42.

[87] Schuchardt S, Sickmann A. Protein identification using mass spectrometry: a method overview. EXS 2007;97:141-70.

[88] Baykut G, Jertz R, Witt M. Matrix-assisted laser desorption/ionization fourier transform ion cyclotron resonance mass spectrometry with pulsed in-source collision gas and in-source ion accumulation. Rapid Commun Mass Spectrom 2000;14(14):1238-47.

[89] Quinton L, Le Caer JP, Phan G, Ligny-Lemaire C, Bourdais-Jomaron J, Ducancel F, et al. Characterization of toxins within crude venoms by combined use of Fourier transform mass spectrometry and cloning. Anal Chem 2005 Oct 15;77(20):6630-9.

[90] Quinton L, Le Caer JP, Vinh J, Gilles N, Chamot-Rooke J. Fourier transform mass spectrometry: a powerful tool for toxin analysis. Toxicon 2006 May;47(6):715-26.

[91] Fox JW, Ma L, Nelson K, Sherman NE, Serrano SM. Comparison of indirect and direct approaches using ion-trap and Fourier transform ion cyclotron resonance mass spectrometry for exploring viperid venom proteomes. Toxicon 2006 May;47(6):700-14.

[92] Trachsel C, Siegemund D, Kampfer U, Kopp LS, Buhr C, Grossmann J, et al. Multicomponent venom of the spider *Cupiennius salei*: a bioanalytical investigation applying different strategies. FEBS Journal 2012 Aug;279(15):2683-94.

[93] Nesvizhskii AI. Protein Identification by Tandem Mass Spectrometry and Sequence Database Searching. Mass Spectrometry Data Analysis in Proteomics Methods in Molecular Biology 2007;367:87-119.

[94] Fox JW, Serrano SMT. Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures. Proteomics 2008 Feb;8(4):909-20.

[95] Honey bee Genome Sequencing and Analysis Consortium. Finding the missing honey bee genes: lessons learned from a genome upgrade. Unpublished.

[96] International Bumblebee Genomics Consortium. Two bumblebee genomes show the route to advanced social living. Unpublished.

[97] Larche M, Akdis CA, Valenta R. Immunological mechanisms of allergen-specific immunotherapy. Nature Reviews Immunology 2006 Oct;6(10):761-71.

[98] Muller UR. Bee venom allergy in beekeepers and their family members. Current Opinion in Allergy and Clinical Immunology 2005 Aug;5(4):343-7.

[99] Bilo BM, Rueff F, Mosbech H, Bonifazi F, Oude-Elberink JNG. Diagnosis of Hymenoptera venom allergy. Allergy 2005 Nov;60(11):1339-49.

[100] Ruëff F, Chatelain R, Przybilla B. Management of occupational Hymenoptera allergy. Current Opinion in Allergy and Clinical Immunology 2011 Apr;11(2):69-74.

[101] Ebo DG, Hagendorens MM, Stevens WJ. Hymenoptera venom allergy. Expert Rev Clin Immunol 2005 May;1(1):169-75.

[102] von Moos S, Graf N, Johansen P, Mullner G, Kundig TM, Senti G. Risk Assessment of Hymenoptera Re-Sting Frequency: Implications for Decision-Making in Venom Immunotherapy. International Archives of Allergy and Immunology 2013;160(1):86-92.

[103] Müller HL. Diagnosis and treatment of insect sensitivity. The journal of asthma research 1966;3(4):331-3.

[104] Bilo MB, Bonifazi F. The natural history and epidemiology of insect venom allergy: clinical implications. Clinical and Experimental Allergy 2009 Oct;39(10):1467-76.

[105] Fireman P, Slavin RG. Atlas of Allergies. Gower; 1990.

[106] Brown TC, Tankersley MS. The sting of the honeybee: an allergic perspective. Annals of Allergy Asthma & Immunology 2011 Dec;107(6):463-71.

[107] de Groot H. Allergy to bumblebees. Current Opinion in Allergy and Clinical Immunology 2006 Aug;6(4):294-7.

[108] Bucher C, Korner P, Wuthrich B. Allergy to bumblebee venom. Curr Opin Allergy Clin Immunol 2001 Aug;1(4):361-5.

[109] Cichocka-Jarosz E, Brzyski P, Swiebocka E, Lange J, Tobiasz-Adamczyk B, Lis G, et al. Health-related quality of life in Polish adolescents with Hymenoptera venom allergy treated with venom immunotherapy. Archives of Medical Science 2012 Dec;8(6):1076-82.

[110] Nullens S, Sabato V, Faber M, Leysen J, Bridts CH, De Clerck LS, et al. Basophilic histamine content and release during venom immunotherapy: Insights by flow cytometry. Cytometry Part B-Clinical Cytometry 2013 May;84B(3):173-8.

[111] Muller U, Helbling A, Berchtold E. Immunotherapy with Honeybee Venom and Yellow Jacket Venom Is Different Regarding Efficacy and Safety. Journal of Allergy and Clinical Immunology 1992 Feb;89(2):529-35.

[112] Rueff F, Przybilla B, Muller U, Mosbech H. The sting challenge test in Hymenoptera venom allergy. Allergy 1996 Apr;51(4):216-25.

[113] Ozdemir C, Kucuksezer UC, Akdis M, Akdis CA. Mechanisms of immunotherapy to wasp and bee venom. Clinical and Experimental Allergy 2011 Sep;41(9):1226-34.

[114] Cruz S, Vega A, Fernandez S, Marques L, Baltasar M, Alonso A, et al. Report From the Hymenoptera Committee of the Spanish Society of Allergology and Clinical Immunology: Immunotherapy With Bumblebee Venom. Journal of Investigational Allergology and Clinical Immunology 2012;22(5):377-8.

[115] Valenta R, Ferreira F, Focke-Tejkl M, Linhart B, Niederberger V, Swoboda I, et al. From Allergen Genes to Allergy Vaccines. Annual Review of Immunology, Vol 28 2010;28:211-41.

[116] Blank S, Seismann H, Michel Y, McIntyre M, Cifuentes L, Braren I, et al. Api m 10, a genuine A. mellifera venom allergen, is clinically relevant but underrepresented in therapeutic extracts. Allergy 2011 Oct;66(10):1322-9.

[117] Makatsori M, Pfaar O, Lleonart R, Calderon MA. Recombinant Allergen Immunotherapy: Clinical Evidence of Efficacy-A Review. Curr Allergy Asthma Rep 2013 Jun 7.

[118] Fellrath JM, Kettner A, Dufour N, Frigerio C, Schneeberger D, Leimgruber A, et al. Allergen-specific T-cell tolerance induction with allergen-derived long synthetic peptides: Results of a phase I trial. Journal of Allergy and Clinical Immunology 2003 Apr;111(4):854-61.

[119] Karamloo F, Schmid-Grendelmeier P, Kussebi F, Akdis M, Salagianni M, von Beust BR, et al. Prevention of allergy by a recombinant multi-allergen vaccine with reduced IgE binding and preserved T cell epitopes. European Journal of Immunology 2005 Nov;35(11):3268-76.

[120] Treudler R, Simon JC. Overview of Component Resolved Diagnostics. Current Allergy and Asthma Reports 2013 Feb;13(1):110-7.

[121] Ebo DG, Hagendorens MM, Bridts CH, Schuerwegh AJ, De Clerck LS, Stevens WJ. In vitro allergy diagnosis: should we follow the flow? Clinical and Experimental Allergy 2004 Mar;34(3):332-9.

[122] Ebo DG, Sainte-Laudy J, Bridts CH, Mertens CH, Hagendorens MM, Schuerwegh AJ, et al. Flow-assisted allergy diagnosis: current applications and future perspectives. Allergy 2006 Sep;61(9):1028-39.

[123] Ebo DG, Faber M, Sabato V, Leysen J, Bridts CH, De Clerck LS. Component-resolved diagnosis of wasp (yellow jacket) venom allergy. Clinical and Experimental Allergy 2013 Feb;43(2):255-61.

[124] Kubelka V, Altmann F, Staudacher E, Tretter V, Marz L, Hard K, et al. Primary Structures of the N-Linked Carbohydrate Chains from Honeybee Venom Phospholipase-A2. European Journal of Biochemistry 1993 May 1;213(3):1193-204.

[125] Lai CC, Her GR. Analysis of phospholipase A(2) glycosylation patterns from venom of individual bees by capillary electrophoresis/electrospray ionization mass spectrometry using an ion trap mass spectrometer. Rapid Communications in Mass Spectrometry 2000;14(21):2012-8.

[126] Kubelka V, Altmann F, Marz L. The Asparagine-Linked Carbohydrate of Honeybee Venom Hyaluronidase. Glycoconjugate Journal 1995 Feb;12(1):77-83.

[127] Kolarich D, Leonard R, Hemmer W, Altmann F. The N-glycans of yellow jacket venom hyaluronidases and the protein sequence of its major isoform in *Vespula vulgaris*. FEBS Journal 2005 Oct;272(20):5182-90.

[128] Blank S, Neu C, Hasche D, Bantleon FI, Jakob T, Spillner E. Polistes species venom is devoid of carbohydrate-based cross-reactivity and allows interference-free diagnostics. Journal of Allergy and Clinical Immunology 2013 Apr;131(4):1239-42.

[129] Kaulfurst-Soboll H, Mertens M, Brehler R, von Schaewen A. Reduction of Cross-Reactive Carbohydrate Determinants in Plant Foodstuff: Elucidation of Clinical Relevance and Implications for Allergy Diagnosis. PLoS ONE 2011 Mar 14;6(3):e17800.

[130] Malandain H. IgE-reactive carbohydrate epitopes--classification, cross-reactivity, and clinical impact. European Annals of Allergy and Clinical Immunology 2005;37(4):122-8.

[131] Chung CH, Mirakhur B, Chan E, Le Q, Berlin J, Morse M, et al. Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. New England Journal of Medicine 2008 Mar 13;358(11):1109-17.

[132] Commins SP, Platts-Mills TAE. Allergenicity of Carbohydrates and Their Role in Anaphylactic Events. Current Allergy and Asthma Reports 2010 Jan;10(1):29-33.

[133] Koetz A. The Asian honey bee (*Apis cerana*) and its strains - with special focus on *Apis cerana* Java genotype. State of Queensland; 2013.

[134] Stapel SO, de Raadt JWL, van Toorenenbergen AW, de Groot H. Allergy to bumblebee venom. II. IgE cross-reactivity between bumblebee and honeybee venom. Allergy 1998 Aug;53(8):769-77.

[135] Huynh K, Wang M, van Egeraat C, Lopez M, Hovanec-Burns D, Banik U. Evaluation of bumblebee venom for the diagnosis of hymenoptera venom allergy. 2009.

[136] Hoffman DR. Ant venoms. Current Opinion in Allergy and Clinical Immunology 2010 Aug;10(4):342-6.

[137] Wurm Y, Wang J, Riba-Grognuz O, Corona M, Nygaard S, Hunt BG, et al. The genome of the fire ant *Solenopsis invicta*. Proceedings of the National Academy of Sciences of the United States of America 2011 Apr 5;108(14):5679-84.

[138] Bonasio R, Zhang GJ, Ye CY, Mutti NS, Fang XD, Qin N, et al. Genomic Comparison of the Ants *Camponotus floridanus* and *Harpegnathos saltator*. Science 2010 Aug 27;329(5995):1068-71.

[139] Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, et al. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. Proceedings of the National Academy of Sciences of the United States of America 2011 Apr 5;108(14):5667-72.

[140] Ebo DG. Component-resolved allergy diagnosis: a new era? Verhandelingen - Koninklijke Academie voor Geneeskunde van België 2011;73(3):163-79.

[141] Hofmann SC, Pfender N, Weckesser S, Blank S, Huss-Marp J, Spillner E, et al. Detection of IgE to recombinant Api m 1 and rVes v 5 is valuable but not sufficient to distinguish bee from wasp venom allergy Reply. Journal of Allergy and Clinical Immunology 2011 Jul;128(1):248.

[142] Valenta R, Vrtala S. Recombinant allergens for specific immunotherapy. Allergy 1999;54:43-4.

[143] Muller UR, Johansen N, Petersen AB, Fromberg-Nielsen J, Haeberli G. Hymenoptera venom allergy: analysis of double positivity to honey bee and Vespula venom by estimation of IgE antibodies to species-specific major allergens Api m1 and Ves v5. Allergy 2009 Apr;64(4):543-8.

[144] Sturm GJ, Hemmer W, Hawranek T, Lang R, Ollert M, Spillner E, et al. Detection of IgE to recombinant Api m 1 and rVes v 5 is valuable but not sufficient to distinguish bee from wasp venom allergy. Journal of Allergy and Clinical Immunology 2011 Jul;128(1):247-8.

[145] Ebo DG, Hagendorens MM, Bridts CH, De Clerck LS, Stevens WJ. Sensitization to cross-reactive carbohydrate determinants and the ubiquitous protein profilin: mimickers of allergy. Clinical and Experimental Allergy 2004 Jan;34(1):137-44.

[146] Jin C, Focke M, Leonard R, Jarisch R, Altmann F, Hemmer W. Reassessing the role of hyaluronidase in yellow jacket venom allergy. Journal of Allergy and Clinical Immunology 2010 Jan;125(1):184-90.

[147] Mittermann I, Zidarn M, Silar M, Markovic-Housley Z, Aberer W, Korosec P, et al. Recombinant allergen-based IgE testing to distinguish bee and wasp allergy. Journal of Allergy and Clinical Immunology 2010 Jun;125(6):1300-7.

[148] Hofmann SC, Pfender N, Weckesser S, Huss-Marp J, Jakob T. Added value of IgE detection to rApi m 1 and rVes v 5 in patients with Hymenoptera venom allergy. Journal of Allergy and Clinical Immunology 2011 Jan;127(1):265-7.

[149] Korosec P, Valenta R, Mittermann I, Celesnik N, Silar M, Zidarn M, et al. High sensitivity of CAP-FEIA rVes v 5 and rVes v 1 for diagnosis of Vespula venom allergy. Journal of Allergy and Clinical Immunology 2012 May;129(5):1406-8.

[150] Jakob T, Köhler J, Blank S, Huss-Marp J, Spillner E, Lidholmh J. Comparable IgE-reactivity to nApi m 1 and rApi m 1 in CCD negative bee venom allergic patients. Journal of Allergy and Clinical Immunology 2012;130:276-8.

[151] Vos B, Kohler J, Muller S, Soubeiran E, Stretz E, Przybilla B, et al. Improved diagnosis of wasp venom allergy by IgE detection to wasp venom extract spiked with rVes v 5. Allergy 2012 Nov;67:613.

[152] Gonzalez-Buitrago JM, Ferreira L, Isidoro-Garcia M, Sanz C, Lorente F, Davila I. Proteomic approaches for identifying new allergens and diagnosing allergic diseases. Clinica Chimica Acta 2007 Oct;385(1-2):21-7.

[153] Kemeny DM, Harries MG, Youlten LJF, Mackenziemills M, Lessof MH. Antibodies to Purified Bee Venom Proteins and Peptides .1. Development of A Highly Specific Rast for Bee Venom Antigens and Its Application to Bee Sting Allergy. Journal of Allergy and Clinical Immunology 1983;71(5):505-14.

[154] Rendic D, Wilson IBH, Paschinger K. The glycosylation capacity of insect cells. Croatica Chemica Acta 2008 Apr;81(1):7-21.

[155] Altmann F, Staudacher E, Wilson IBH, Marz L. Insect cells as hosts for the expression of recombinant glycoproteins. Glycoconjugate Journal 1999 Feb;16(2):109-23.

[156] Seismann H, Blank S, Braren I, Greunke K, Cifuentes L, Grunwald T, et al. Dissecting cross-reactivity in hymenoptera venom allergy by circumvention of alpha-1,3-core fucosylation. Molecular Immunology 2010 Jan;47(4):799-808.

[157] Sturm GJ, Bilo MB, Bonadonna P, Hemmer W, Caruso B, Bokanovic D, et al. Ves v 5 can establish the diagnosis in patients without detectable specific IgE to wasp venom and a possible north-south difference in Api m 1 sensitization in Europe. Journal of Allergy and Clinical Immunology 2012 Sep;130(3):817.

[158] Blank S, Bantleon FI, McIntyre M, Ollert M, Spillner E. The major royal jelly proteins 8 and 9 (Api m 11) are glycosylated components of *Apis mellifera* venom with allergenic potential beyond carbohydrate-based reactivity. Clin Exp Allergy 2012 Jun;42(6):976-85.

[159] Grunwald T, Bockisch B, Spillner E, Ring J, Bredehorst R, Ollert MW. Molecular cloning and expression in insect cells of honeybee venom allergen acid phosphatase (Api m 3). Journal of Allergy and Clinical Immunology 2006 Apr;117(4):848-54.

[160] Georgieva D, Greunke K, Genov N, Betzel C. 3-D Model of the bee venom acid phosphatase: Insights into allergenicity. Biochemical and Biophysical Research Communications 2009 Jan 23;378(4):711-5.

[161] Schenk MF, Gilissen LJWJ, Esselink GD, Smulders MJM. Seven different genes encode a diverse mixture of isoforms of Bet v I, the major birch pollen allergen. Bmc Genomics 2006 Jul 4;7.

[162] Wagner S, Radauer C, Bublin M, Mann-Sommergruber KH, Kopp T, Greisenegger EK, et al. Naturally occurring hypoallergenic Bet v 1 isoforms fail to induce IgE responses in individuals with birch pollen allergy. Journal of Allergy and Clinical Immunology 2008 Jan;121(1):246-52.

[163] Formesyn EM, Danneels EL, de GDC. Proteomics of the Venom of the Parasitoid *Nasonia vitripennis*. Parasitoid viruses - Symbionts and pathogens. 1 ed. 2012. p. 233-46.

The Laboratory of Zoophysiology has a strong connection with the Flemish beekeeping sector. It houses the Service Centre for Beekeepers and conducts the diagnosis of honeybee diseases in the Diagnostic Centre for Bee Diseases. It sustains many bee hives throughout the year, which also allows access to these fascinating insects for research purposes. In addition to the expertise in bee pathology and the search for the causes of winter losses in bee colonies, our group has shown strong interest in the venom composition of the honeybee and parasitoid wasp *Nasonia vitripennis*. The function of *Nasonia* venom in host-parasitoid releationships is being investigated, while studies of the honeybee venom have a biomedical finality. Our approach as insect physiologists reveals a remarkable biochemical complexity of the venom with putative immunological consequences. Our work is the basis for further medical investigations focusing on improvement of venom allergy diagnosis and treatment, and elucidation of the clinical consequences of bee stings.

This PhD has two general objectives. The first objective is to obtain in-depth insights in the venom composition of the honeybee (*A. mellifera*) and bumblebee (*B. terrestris*) by integrating genome, transcriptome and proteome information. Second, this work aims to advance knowledge about the immunological implications of the venom proteome by investigating the allergenic properties of immunologically uncharacterized venom compounds and by analyzing the immunological relevance of allergen protein heterogeneity. We conducted several experiments to achieve these objectives. The results are described in five consecutive chapters. In **chapter 1**, we want to identify novel honeybee venom compounds using liquid chromatography-mass spectrometry, an approach which overcomes the issues of gel-based proteomics. The second objective is to investigate if the Ag5-like sequence, previously found by mining the honeybee genome, is expressed by the honeybee venom glands. Finally, we try to confirm that the novel identified compounds and the Ag5-like compound are present in the venom by analyzing their IgG4-reactivity using sera of immune beekeepers. In **chapter 2** we explore the hidden honeybee venom proteome by

integrating a combinatorial peptide ligand library venom pre-treatment with FTMS, while in **chapter 3** the venom proteome of the European buff-tailed bumblebee, *B. terrestris*, is unraveled using an identical approach. Also genome information is used to obtain further insights in the venom composition of both species. The objective of **chapter 4** is to investigate the nature of Api m 10 protein heterogeneity and to explore its effect on IgE reactivity using sera of honeybee venom allergic patients. In **chapter 5**, we evaluate the allergenic potential of the honeybee venom C1q-like and PVF1 proteins by analyzing IgE reactivity and basophil activation.

Chapter 1

Extending the honeybee venome with the antimicrobial peptide apidaecin and a protein resembling wasp antigen 5

1.1 CONTRIBUTIONS

D. de Graaf and B. Devreese assisted with the study design. The proteomic analysis of honeybee worker venom was executed by M. Brunain and G. Van Driessche, while D. Cardoen and P. Verleyen performed the peptidomic analysis of the venom apparatus tissue. M. Brunain, E. M. Formesyn, C. Baillon and B. Demets applied RT-PCRs on venom gland tissue to confirm apidaecin expression and explore spatial and seasonal variation of Ag5-like gene expression. T. Wenseleers contributed by executing the phylogenetics analysis. During a 6 week internship of M. Van Vaerenbergh at the Institute of Biochemistry and Molecular Biology (Hamburg University, Germany), he was assisted by F. I. Bantleon and S. Blank for Ves v 5 insect cell expression. M. Van Vaerenbergh verified all data, performed the RT-PCR, cloning, expression and purification of the antigen 5-like protein, determined the IgG4 titers of beekeeper sera using ELISA and conducted the immunoblotting experiment. M. Van Vaerenbergh wrote the article and was assisted by the co-authors through the writing phase.

1.2 ABSTRACT

Honeybee venom is a complex mixture of toxic proteins and peptides. In this study we tried to extend our knowledge of the venom composition by two different approaches. First,

worker venom was analyzed by liquid chromatography-mass spectrometry and this revealed for the first time the antimicrobial peptide apidaecin in such samples. Its expression in the venom gland was confirmed by reverse transcription PCR and by a peptidomic analysis of the venom apparatus tissue. Second, genome mining revealed a list of proteins with resemblance to known insect allergens or venom toxins, one of which showed homology to proteins of the antigen 5 (Ag5)/Sol i 3 cluster. It was demonstrated that the honeybee Ag5-like gene is expressed by venom gland tissue of winter bees but not of summer bees. Besides this seasonal variation, it shows an interesting spatial expression pattern with additional production in the hypopharyngeal glands, the brains and the midgut. Finally, our immunoblot study revealed that both synthetic apidaecin and the Ag5-like recombinant from bacteria evoke no humoral activity in beekeepers. Also, no IgG4-based cross-reactivity was detected between the honeybee Ag5-like protein and its yellow jacket paralogue Ves v 5.

1.3 INTRODUCTION

Honeybees defend the hive against predators and external threats using venom which contains several toxic compounds that cause death in other insects or inflict pain in higher organisms. In man, early exposure to bee venom evokes IgG1, IgG2 and to a lesser extent IgG4 antibody responses, whereas long-term exposure often found in beekeepers drives the immunity to an IgG4 type of humoral response [1;2]. Allergy to a bee sting is mediated by IgE antibodies and, so far, 12 honeybee venom allergens have been listed by the International Union of Immunological Societies (IUIS; http://www.allergen.org/Allergen.aspx), representing most of the compounds that are immunologically meaningful.

Following the sequencing of the honeybee genome [3], venom protein maps became available with newly discovered proteins, some of which were subsequently studied in detail and assigned as new allergens [4-9]. Remarkably, the venom protein composition could also be further completed by whole venom gland tissue mass spectrometry, a study initially performed in order to understand why the toxic compounds are not self-destructive [10]. However, the previous venom and gland proteomic studies combined two-dimensional (2D) gel electrophoresis with MALDI-TOF/TOF (matrix-assisted laser desorption/ionization tandem time of flight) and/or liquid chromatography (LC)-MS/MS, and these gel-based

approaches have some disadvantages: very lowly abundant compounds are not visible on the gel and low molecular weight fractions are lacking because of their higher electrophoretic mobility which allows them to migrate out of the gel. In order to overcome these issues related to gel-based proteomics and to gain deeper insights in venom and venom gland proteomes/peptidomes we extended our search for the venom constituents with a mass spectrometric study of in liquid digested LC fractions from venom and venom gland tissue.

In addition, we focused on a remarkable peculiarity of honeybee venom: the lack of an antigen 5 (Ag5) homologue, an issue that was contested by a genome mining study that revealed multiple protein predictions with resemblance to the proteins of the Ag5/Sol i 3 cluster [3]. Ag5 is a common venom allergen of the vespid group that includes wasps, yellow-jackets and hornets of the genera *Vespula*, *Vespa*, *Dolichovespula* and *Polistes*. In fact, according to the IUIS allergen list Ag5 has been discovered in almost every species of the here above listed vespid genera and for some of them it seems to be the solely known venom allergen. Moreover, the ants of the genus *Solenopsis* (*Solenopsis invicta*, Sol i 3; *Solenopsis richteri*, Sol r 3; *Solenopsis saevissima*, Sol s 3) all have a major allergen that shows strong resemblance to vespid Ag5. Although immunologically characterized in detail, the function of Ag5 proteins within wasp and fire ant venom remains largely unexplored [11]. Proteins belonging to the Ag5/Sol i 3 cluster form a major and distinct clade of the CAP (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins) superfamily, whose members are found in a broad range of organisms spanning the entire animal kingdom [11]. Remarkably, Ag5 homologues have so far been discovered in the venoms of none of the hymenopteran species belonging to the Corbiculate bees, such as honeybees and bumblebees. In this second part, we focused on a honeybee Ag5-like protein (NCBI RefSeq: XP_001122516.2) showing the highest sequence similarity with the venom Ag5 of the yellow jacket, *Vespula vulgaris*. We relied on gene expression studies in order to verify whether this protein is produced by the honeybee venom glands and determined its phylogenetic relationship with other hymenopteran Ag5s.

Finally, the immunological significance of this Ag5-like protein and of new venom/venom gland compounds derived from the proteomic study was determined by Western blot using sera of highly exposed, immune beekeepers.

1.4 MATERIALS AND METHODS

1.4.1 Ethics statement

Blood sampling was approved by the local ethics committee (registration number: B67020072793) and all participants provided verbal informed consent. There was a verbal agreement with the participants that the samples would be used for no other purposes than the determination of the immune status against bee venom. The procedure was kept rather informal as the participants (beekeepers) were employed or in a way connected with the research center.

1.4.2 Animals, venom and tissue collection

Summer and winter worker honeybees (*Apis mellifera carnica*) were collected from the hives of the experimental apiaries of Ghent University and K.U. Leuven, which were reared using standard beekeeping methods. BFR honeybees were collected during the winter from a hive that was placed for several months in a climate room that simulates summer conditions (temperature fixed at 25°C, *ad libitum* sugar solution, plain water and pollen).

Pure honeybee worker venom was collected by 'manually milking' as described by Peiren *et al.* [9]. Venom glands and their reservoirs of 10 honeybees were dissected as described previously [10]. Honeybee worker brain, hemocytes, venom gland, hypopharyngeal gland, salivary gland, drone mucus glands, midgut, deviscerated abdomen and muscle tissue for RNA extraction were dissected as described by de Graaf *et al.* [5] and submerged in RNALater®.

1.4.3 Proteomics/peptidomics

1.4.3.1 Proteomic analysis of pure venom

Two milligrams of pure worker honeybee venom was dissolved in 200 µl of 0.1% TFA (buffer A) and separated on a Shimadzu RP-HPLC system consisting of an SCL-10Avp system controller, LC-ADvp pump, FCV-10Alvp low pressure gradient unit, SPD-10Avp UV-VIS detector and an FRC-10A fraction collector. Proteins were eluted from the Pathfinder 300 C18 AP column (Shimadzu) by a linear gradient from 0-100% buffer B containing 30% 0.1% TFA and 70% acetonitrile over a 35 minute period (0.7ml/min). Separate protein peaks were collected and dried by vacuum centrifugation. Consequently, 10 µg protein of each peak was dissolved in 10 µl of 50 mM ammonium bicarbonate. Further reduction, alkylation, tryptic

digestion and guanidination were executed according to the In-Solution Tryptic Digestion and Guanidination kit protocol (Thermo Scientific).

MALDI mass spectrometry was carried out on a 4700 MALDI TOF/TOF Analyzer (Applied Biosystems, Boston, MA, USA). One microliter of the guanidinated sample was mixed with 1 µl α-cyano-4-hydroxycinnamic acid (10 mg/ml) in 60% acetonitrile containing 0.1% TFA and 10% ethanol. All MALDI spectra were calibrated with 4700 Proteomics Analyzer Calibration Mixture (4700 Cal Mix, Applied Biosystems) and, prior to data collection, all instrumental parameters were tuned. Protein identification (Peptide Mass Fingerprinting) was performed by searching the extracted peaks against the SwissProt and in-house *Apis mellifera* database using the MASCOT search engine (Matrixscience, London, UK) [peptide mass tolerance: 0.100 Da, 2 missed cleavages, deamidation (NQ) and oxidation (M)]. Further confirmation of the identification of the proteins was done by selecting the highest peaks for MS/MS fragmentation spectrometry and using the above described Mascot search engine (peptide mass tolerance: 0.250 Da, 1 missed cleavage, MS/MS tolerance: 0.25 Da).

1.4.3.2 Peptidomic analysis of venom apparatus tissue

We made a peptide extract of 10 dissected poison sacks in 100 µl methanol/water/acetic acid (90/9/1, v/v/v). Upon thoroughly sonicating, the sample was centrifuged for 10 min at 14000 rpm at 4 °C. The supernatants was transferred and the resulting pellet was resuspended in 20 µl of methanol/water/acetic acid (90/9/1, v/v/v). Sonication and centrifugation steps were repeated and both supernatants were pooled, dried (vacuum centrifuge) and stored at -30 °C until further analysis. To analyze the sample, the dry extract was dissolved in 5% acetonitrile and 0.5% formic acid and separated by nanoLC with a Dionex UltiMate™ 3000 Dual LC System (Dionex) device, coupled online to a MicrOTOF-Q (Bruker Daltonics) mass spectrometer. We applied an acetonitrile gradient from 5% to 40% in 30min, followed by a gradient to 90% in 3 min and back to 5% in 10 min. As much peptides as possible were fragmented in a collision cell.

The data obtained by mass spectrometry were converted to an .mgf-file and used as the input on the in-house Mascot server for an MS/MS–ion search. We performed searches with variable modifications (amidation, pyroglutamate and oxidation of methionine) and with a peptide tolerance and MS/MS tolerance of 0.2 Da in the in-house *Apis* neuropeptide

precursor database. This database consists of the 39 neuropeptide precursors, including two precursors for apidaecin [12;13].

1.4.4 RNA isolation, cDNA synthesis, primer development and reverse transcription PCR

Tissue RNA isolation, cDNA synthesis, reverse transcription PCR and amplicon sequencing was done as described by de Graaf *et al.* [5]. DNA elongation in PCR was adapted to 1 min at 72 °C. For amplification of transcripts from various cDNA sources, primers were developed with a melting temperature ™ of approximately 60 °C, using the formula Tm = 2(A + T) + 4(G + C) °C. The different primer sets are listed in Table 1.1. Primer sequences for amplification of apidaecin were developed as used by Casteels-Josson *et al.* [14] (primers 5SB6-2 and 3(S)B6). Profilin (NM_001098167) primers to control for the presence of genomic DNA are

Table 1.1: Primer sets used in reverse transcription PCR: CDS: Primer set to amplify the coding sequence; gDNA_CTL: Primer set to control the presence of genomic DNA; cDNA_CTL: Primer set to control the presence of cDNA; CONS: Primer set to amplify a conserved domain; ORF: Primer set to amplify the entire open reading frame; MAT: Primer set to amplify the mature fragment (=ORF without the signal sequence).

Gene	Primer sequence	Amplicon length
Apidaecin (CDS)	5'-CCAACCTAGATCCGCCTACTCGACCT-3' 5'-TATTTCACGTGCTTCATATTCTTC-3'	multiple isoforms
Profilin (gDNA_CTL)	5'-GCGACAAGAGGGAAAGTACG-3' 5'-CGGTGGACAAAATTCTGGAG-3'	685 bp
Profilin (cDNA_CTL)	5'-GGCTTCGAAGTAAGTAAAGAGGA-3' 5'-AGTTTTTCAACGACCGATGC-3'	248 bp
PLA$_2$ (cDNA_CTL)	5'-ATGCAAGTCGTTCTCGGATC-3' 5'-ATACTTGCGAAGATCGAACCA-3'	501 bp
Ag5-like (CONS)	5'-CTGACCTGGGACGATGAACT-3' 5'-GCCTATTAAATAACTATTAGCCCAGAA-3'	252 bp
Ag5-like (ORF)	5'-ATGGCGCGGGAGGGAATAA-3' 5'-TTAACATCGAGTTCCCAGATAA-3'	672 bp
Ag5-like (MAT)	5'-CACCGATGTTATTTCCTGCATCGGC-3' 5'-TTAACATCGAGTTCCCAGATAA-3'	609 bp

developed within its first intron, while exon primers are used to control for the presence of cDNA. Honeybee venom phospholipase A_2 (PLA$_2$, NM_001011614.1) primers and primers for amplification of the predicted honeybee antigen 5-like sequence (Ag5-like, XM_001122516.2) are developed at the extreme 5' and 3' ends of the coding sequence. Additionally, primers were developed for amplifying the secreted mature Ag5–like sequence which lacks a signal sequence (determined using the SignalP 4.0 Server: http://www.cbs.dtu.dk/services/SignalP/) [15]. For directional cloning of the mature Ag5-like sequence in the prokaryotic expression vector, the forward primer was preceded by four bases (CACC). Also a set of primers in conserved regions of the Ag5-like sequence was determined by aligning a set of Ag5 homologous sequences.

1.4.5 Production and purification

1.4.5.1 Recombinant baculovirus expression

Recombinant yellow jacket Ves v 5 was produced in Sf9 insect cells as described by Seismann *et al.* [16]. The cellular supernatant containing the secreted recombinant protein was dialyzed to PBS pH 8.0 and supplied to a Hi-Trap column (Sigma-Aldrich) for His-tag purification at a flow rate of 1 ml/min (ÄKTAprime™ plus system, GE Healthcare Life Sciences). The column was washed by a four step-gradient with elution buffer (PBS pH8.0 containing 300 mM imidazole): 10 min gradient to 3% elution buffer followed by 10 min at constant 3% elution buffer, which was repeated with elevating concentrations of elution buffer (to 6%, 10% and 15%). The recombinant protein was eluted from the column with 100% elution buffer. Protein dialysis to PBS was executed by desalting (PD MidiTrap G-25, GE Healthcare) and sample purity was determined by Coomassie Brilliant Blue R-250 staining of an SDS-PAGE gel run under reducing (2x Laemmli sample buffer with 10% of β-mercaptoethanol) and denaturing conditions (sample at 100 °C for 5 min and SDS added to PAGE gel and running buffer). The protein concentration of the sample was estimated by comparing the staining intensity on a Coomassie stained SDS-PAGE gel with a dilution series of albumin standard.

1.4.5.2 Recombinant bacterial expression

Equine uterocalin and the mature honeybee Ag5-like sequence were cloned, sequenced and expressed following procedures described by de Graaf *et al.* [5]. For denaturing purification

of recombinant proteins by His-tag, cell pellets were dissolved in 8 ml of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazol and 8M ureum) and sonicated on ice with five ten-second pulses at high intensity. After centrifugation (13500 rpm, 4°C, 30 min) supernatant was supplied to the Profinity IMAC Ni-charged resin (Bio-RAD) which was equilibrated with lysis buffer. The resin was washed with 24 ml of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole and 8M ureum) and recombinant protein was eluted in elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazol and 8M ureum). After dialysis to lysis buffer (SnakeSkin dialysis tubing, 3.5 MWCO, Thermo Scientific) recombinant Ag5-like protein was further purified by a second identical purification step. Dialysis to PBS, sample purity and protein concentration determination was done as mentioned previously.

1.4.6 Sera

Sera were collected of 10 highly exposed beekeepers which have no allergic symptoms upon bee stings. Information about whether or not they have a history of yellow jacket stings was also available. In addition, we collected five negative control sera from persons that had never been stung by hymenopteran insects.

1.4.7 ELISA

Honeybee venom-, honeybee venom PLA_2- and melittin-specific serum IgG4 titers were determined for all sera using ELISA. Nunc MaxiSorp® flat bottom 96 well plates were coated with 150 µl of honeybee venom (2 µg/ml, Sigma-Aldrich), purified honeybee venom PLA_2 (4 µg/ml, Latoxan) and purified melittin (1 µg/ml, Latoxan) in coating buffer (100 mM bicarbonate/carbonate buffer, pH 9.6) at 4°C overnight. Subsequently, wells were washed three times with PBST, blocked with 50 mg/ml skimmed milk powder in PBS at room temperature for 2 hours and washing was repeated. For each serum sample a two-fold dilution series from 1:40 to 1:20480 was performed using blocking buffer. Plates were incubated for 45 minutes at 37°C and washed three times before bound IgG4 was detected with 150 µl of HRP-conjugated mouse anti-human IgG4 (Southern Biotech) diluted in 1:10000 in blocking buffer (1 hour at 37°C). Wells were washed three times with PBST and 200 µl of substrate solution (SIGMAGAST OPD, Sigma-Aldrich) was added to each well. After 30 minutes, the reaction was stopped with 100 µl of stop solution (3M HCl) and the plates were read at 490 nm. Antibody titer was defined as the highest dilution with a reading above

the mean of the negative controls plus 1.96 SDs.

1.4.8 Western blotting

Twelve microgram of each protein fraction was separated by 15% SDS-PAGE under reducing and denaturing conditions in a discontinuous system (Bio-Rad) and blotted to polyvinylidene difluoride membrane. After blotting, the membrane was cut into strips, blocked for 2 hours and incubated overnight at 4 °C with 2 ml of diluted serum (1/16 diluted in blocking solution). Subsequently, strips were washed three times with blocking solution and incubated in 2 ml of 1:1000 diluted HRP-conjugated mouse anti-human IgG4 antibody (Southern Biotech) in blocking solution. Finally, blots were washed 3 times with PBST and once with PBS before DAB staining. Anti-His staining of His-tagged recombinant proteins was done as described before [4;5]. Ponceau S staining was used for determination of blotting success of non-His-tagged proteins. PBS buffer and 1 μg of recombinant equine uterocalin was spotted to serve as negative controls [17].

1.4.9 Phylogenetic analysis

Sequences related to our honeybee Ag5-like sequence in other Hymenoptera were retrieved based on a protein blast against all non-redundant protein and predicted protein sequences in NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) using default search parameters. Low quality or incomplete sequences, or sequences that showed unusually high sequence divergence, were removed. Subsequently, sequences were aligned using MUSCLE [18], after which a neighbor-joining phylogenetic tree was estimated using the Jones-Taylor-Thornton ([19]) model (+G). A discrete Gamma distribution with three categories was used to model differences in the substitution rate among sites. Mean evolutionary rates in these categories were estimated at 0.34, 0.84, 1.82 substitutions per site and the shape parameter of the gamma distribution was estimated at 1.9. In all calculations, positions with less than 50% site coverage were eliminated. This resulted in a total of 51 sequences of 208 amino acid positions each in the final dataset. The reliability of the phylogenetic placement of the sequences was assessed using the bootstrap method using 500 bootstrap replicates. A tentative classification of sequences into orthologue groups was made based on the repeated appearance in the tree of sequences from species with known phylogenetic placement and fully sequenced genomes (e.g. *Nasonia*, *Apis*). All evolutionary analyses were

conducted using MEGA5 [20;20;20;20;21]). Maximum likelihood or Bayesian trees, obtained using Mr. Bayes, resulted in very similar phylogenetic patterns as the ones obtained using a neighbor-joining approach (results not shown).

1.5 RESULTS

1.5.1 Proteomic analysis of pure worker venom

The chromatogram of the RP-HPLC separated worker honeybee venom is shown in Figure 1.1. MALDI-TOF/TOF analysis of twenty-six separated protein peaks resulted in the identification of nine known venom proteins as well as of the antimicrobial peptide apidaecin which up to now had not been detected in honeybee venom. The results of the mass spectrometric identifications are summarized in Table 1.2.
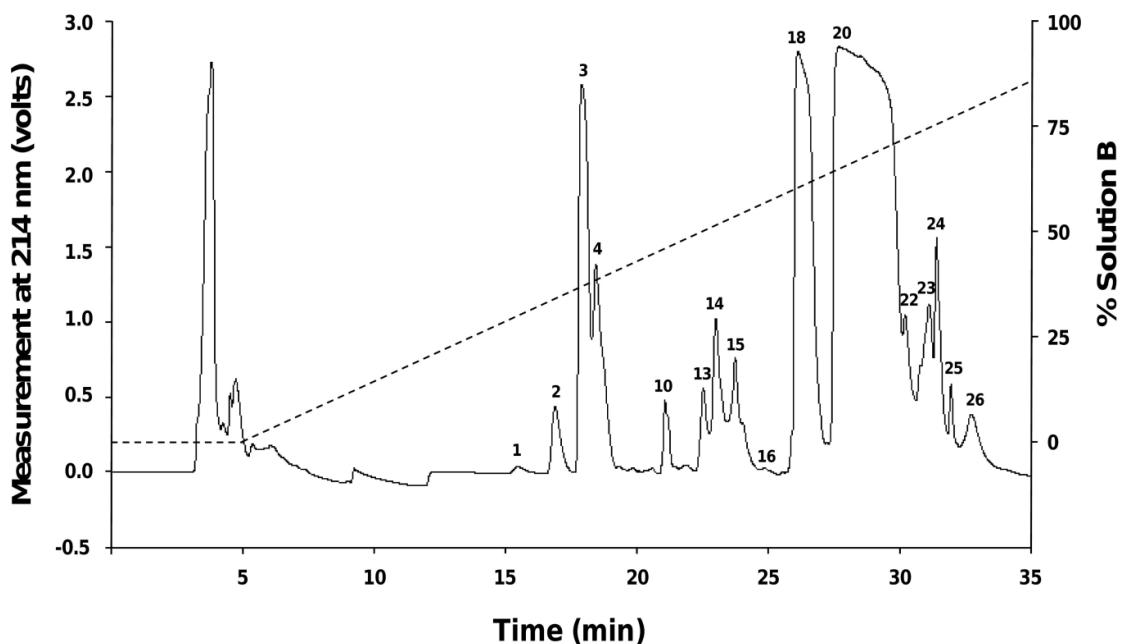
Figure 1.1: RP-HPLC chromatogram of 2 mg of separated pure honeybee venom. The linear gradient of 0-100% buffer B over a 35 minute period is represented by the dotted line. More information about the evaluated peaks, which are numbered on the chromatogram, can be found in Table 1.2.

1.5.2 Peptidomic analysis of venom apparatus tissue

The GNNRPVYIPQPRPPHP peptide was also found in our analysis of the venom apparatus tissue peptidome (Figure S1.1), which suggests the expression of apidaecin by the venom gland. Additional peptides of melittin, phospholipase A_2, PDGF/VEGF-like protein and

secapin were also found, all representing compounds that are known to occur in honeybee venom and/or venom apparatus tissue [9;10].

1.5.3 RT-PCR confirmation of apidaecin expression by honeybee worker venom gland tissue

Reverse transcription PCR (RT-PCR) on honeybee worker venom gland tissue using primers at the extreme ends of the coding sequence of prepro-apidaecin generated a pool of amplicons with different sizes (Figure 1.2). An identical band pattern was seen in RT-PCR of hemocyte derived cDNA and confirmed earlier observations of Casteels-Josson *et al.* [14]. Moreover, in the latter study it was proven that all generated PCR fragments contained genuine apidaecin sequences. Sequencing of the smallest amplicon band revealed its apidaecin precursor identity and confirmed the apidaecin expression by the venom gland.
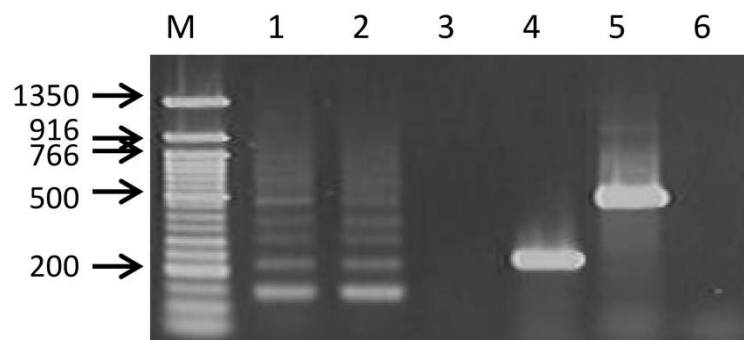
Figure 1.2: Reverse transcription PCR confirmation of apidaecin expression by honeybee worker venom gland tissue. 1= venom gland cDNA with apidaecin coding sequence primers, 2= hemocyte cDNA with apidaecin coding sequence primers, 3= venom gland gDNA control with profilin intron primers, 4= venom gland cDNA control with profilin exon primers, 5= venom gland cDNA control based on the honeybee venom PLA$_2$ sequence, 6= no template control with PLA$_2$ primer set, M= 50 bp DNA marker (New England Biolabs). Base pair lengths of major marker bands are shown.

1.5.4 Spatial and seasonal variation of Ag5-like gene expression

RT-PCR on venom gland tissue from winter bees with primers developed in conserved regions of the Ag5-like gene prediction generated an amplicon of the expected length (252 bp, result not shown). In contrast, RT-PCR on different tissues of summer bees using the same conserved primer set demonstrated the expression of the Ag5-like gene in the

Table 1.2. Protein identification on the HPLC peaks of honeybee venom. The molecular weight (MW), pI (isoelectric point) and sequence coverage of the mature protein (without signal peptide) are represented between brackets. [1] Based on the found peptides no differentiation of the apidaecin precursor isoform was possible. Data are presented for apidaecin 14 (GI:58585168). Other apidaecin precursors containing the found peptide sequences were present in the database: apidaecin 22 (GI:58585226), apidaecin 73 (GI:4539289).

Name	Acc. N°	MW in Da	pI	Peptide mass	Peptide sequences	Sequence coverage	Score	Peak
Apamin	gi|58585166|	5,220 (1,928)	8.77 (8.83)	726.33 987.48	RCQQH APETALCAR	34 (82)	86	3, 4
Mast cell-degranulating peptide	gi|58585162|	5,777 (2,477)	9.87 (10.28)	1158.64 1314.72	HVIKPHICR RHVIKPHICR	20 (45)	79	3, 4
Apidaecin [1]	gi|58585168|	19,368 (17,359)	11.29 (11.30)	1837.97 1994.07	GNNRPVYIPQPRPPHP GNNRPVYIPQPRPPHPR	9 (11)	17	10, 12, 13
Secapin	gi|58585180|	8,674 (2,868)	9.51 (10.05)	971.54	YIIDVPPR	10 (32)	44	13, 14, 15
Venom allergen Api m 6	gi|94400907|	10.015 (7,598)	10.06 (9.89)	1091.40 1108.48 1184.57 1293.59	ICAPGCVCR CPSNEIFSR FGGFGGFGGLGGR GKCPSNEIFSR	35 (45)	143	13, 15
Icarapin	gi|60115688|	24,773 (22,697)	4.51 (4.40)	1003.51 1072.53 1282.67 1581.91 1606.74	EQMAGILSR EQGVVNWNK IPEQGVVNWNK KNDTVLVLPSIER SVESVEDFDNEIPK	21 (23)	113	16, 22, 23
Phospholipase A$_2$	gi|58585172|	19,058 (15.149)	7.55 (8.07)	918.30 1034.46 1125.56 1193.69	HTDACCR CLHYTVDK VYQWFDLR HGLTNTASHTR	43 (53)	256	17, 19

				1253.56	LEHPVTGCGER			
				1663.64	THDMCPDVMSAGESK			
				1837.73	LSCDCDDKFYDCLK			
Melittin	gi|58585154|	7,580 (5,119)	4.69 (4.51)	1510.91 1667.01	VLTTGLPALISWIK VLTTGLPALISWIKR	31	97	20, 22, 23, 24, 25
Hyaluronidase	gi|58585182|	44,232 (40.47)	8,82 (8.82)	798.45 1191.62 1228.53	HLQVFR DHLINQIPDK EHPFWDDQR	6 (7)	77	23
Venom acid phosphatase	gi|61656214|	45,360 (43,905)	5.63 (5.83)	888.52 1015.62 1152.59 1379.74 1752.80 1982.99 2068.00	QINVIFR KLYGGPLLR EYQLGQFLR IVYYLGIPSEAR DPYLYYDFYPLER LQQWNEDLNWQPIATK FVDESANNLSIEELDFVK	28 (29)	85	26

hypopharyngeal glands, the brains and the midgut only, but not in the venom glands and the other tissues (hemocytes, salivary glands, drone mucus glands, deviscerated abdomen and muscle tissue). The Ag5-like gene seemed to be expressed most abundantly in the brains (results not shown).

To confirm these results, RT-PCRs were carried out on venom gland tissue of winter, summer and bee flight room (BFR) honeybees using primers developed at the 5'- and 3'-terminal ends of the predicted coding sequence. In contrast to winter bee venom glands, no Ag5-like gene expression was demonstrated in summer and BFR bee venom glands. Additionally, the Ag5-like gene was shown to be expressed abundantly in brains of winter and BFR bees. The same results were obtained with primers for amplification of the Ag5-like fragment without signal peptide (Figure 1.3). The mature winter venom gland fragment was cloned and sequenced (NCBI accession number JX310326), which confirmed its Ag5 similarity. Six nucleotide substitutions were found between the cloned fragment and the predicted NCBI sequence (XM_001122516.2), but none of them influenced the amino acid sequence (sequence alignments in Figure S1.2).
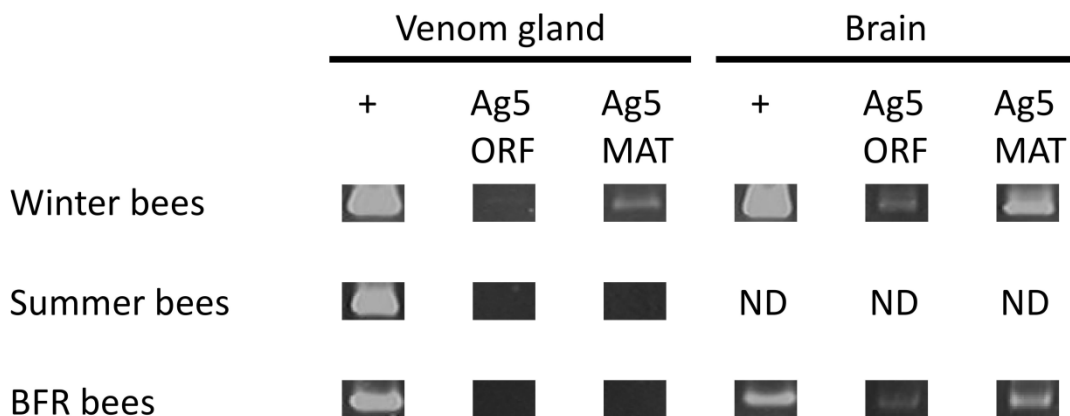
Figure 1.3: Seasonal variation of Ag5-like gene expression in venom gland and brain tissue of honeybee workers. Expression patterns of the Ag5-like gene were determined by reverse transcription PCR. In winter honeybees the Ag5-like gene is expressed by venom glands and brain tissue. Summer and BFR venom glands lack Ag5-like gene expression, while its expression was demonstrated in brain tissue of BFR bees. Ag5-like gene expression in brain tissue of summer bees was not determined (ND). ORF= open reading frame based on the Ag5-like prediction (XM_001122516.2). MAT= mature Ag5-like prediction without signal sequence. += profilin cDNA control.

1.5.5 Ag5 sequence analysis and phylogenetics

Sequence alignment of hymenopteran Ag5-like sequences shows that Ag5 has a complex evolutionary history, with frequent gene duplications and losses. Within the Hymenoptera alone, at least six orthologous groups of sequences can be distinguished (Figure 1.4, groups A-F). Among these, the studied *A. mellifera* Ag5-like sequence (Figure 1.4, highlighted) is orthologous to other predicted Ag5-like sequences from the bumblebees *Bombus terrestris* and *B. impatiens*, the leafcutter bee *Megachile rotundata* and the jewel wasp *Nasonia vitripennis* (Figure 1.4, group D). The sequence, however, is clearly paralogous to previously reported Ag5 allergen sequences in Vespidae wasps (*Vespula, Vespa, Dolichovespula, Polistes, Polybia* and *Rhynchium*) and ants (e.g. *Solenopsis*) (Figure 1.4 group A).

1.5.6 Immunoblotting

We were able to produce recombinant honeybee Ag5-like protein and equine uterocalin (irrelevant protein) in *E. coli* and Ves v 5 in Sf9 insect cells and purify them by affinity chromatography (Figure S1.3). The apidaecin peptide sequence GNNRPVYIPQPRPPHPRL was produced synthetically.

ELISA revealed very high titers of IgG4-antibodies specific for honeybee venom, PLA$_2$ and melittin in all beekeepers' sera (results in Table S1.1).

Western blots showed lacking IgG4 recognition of purified bacterial recombinant Ag5-like protein and synthetically produced apidaecin in all sera (Figure 1.5). In addition, several sera of beekeepers with a history of yellow jacket stings (sera 2-8 and 10) recognize the recombinant Ves v 5 by IgG4 (strips 5-8 and 10). Thus, no cross-reactivity was observed between the Ag5-like protein from bacteria and Ves v 5 from insect cells. Anti-His staining revealed multiple bands probably resulting from dimerization and/or degradation of the recombinants. One of the negative controls (strip 12) contains IgG4-antibodies responsive to the blocking milk proteins.

1.6 DISCUSSION

First, the present study aimed to unravel the complex honeybee venom mixture by a gel-free proteomic analysis. Our preceding approaches were based on 2D gel-separation of venom proteins followed by a mass spectrometric analysis of excised spots [5;9]. Although those
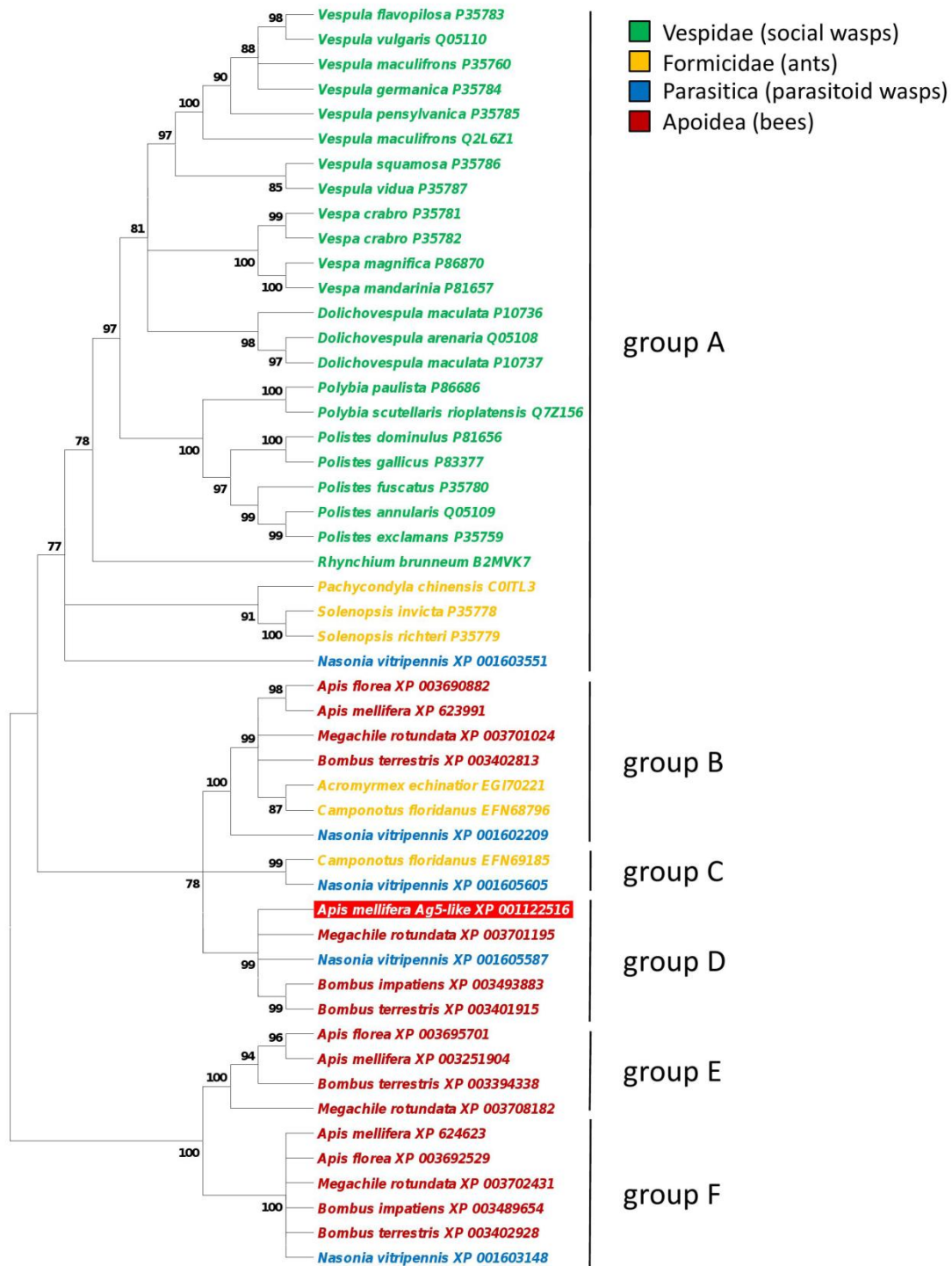
Figure 1.4: Neighbor-joining phylogenetic tree of hymenopteran Ag5s. Phylogenetic placement of the *Apis mellifera* Ag5 like protein sequence (highlighted) in comparison to other known or predicted hymenopteran Ag5 like proteins, as indicated based on a neighbor joining analysis with a Jones-Taylor-Thornton (1992) (G+) substitution model. Tentative orthologue groups are indicated with letters A-F. The numbers at the nodes indicate bootstrap support. Branches with less than 70% bootstrap support were collapsed.

analyses successfully identified four new venom compounds, that technology is limited to detect low molecular weight (<10kDa) and lowly abundant compounds. The present liquid-based venom proteome analysis confirmed the presence of multiple honeybee venom compounds, but also revealed the presence of three additional peptides which were not found in our preceding gel-based analyses: while apamin and mast cell degranulating peptide (MCD) were already discovered before [22-24], the 18 amino acid peptide apidaecin had never been described in venom samples.

For venom research, a venom sample collected by manual milking is preferred over samples collected by electrostimulation, as the latter may contain contaminants derived from saliva or digestive tract fluids (de Graaf *et al.*, 2009). Moreover, the protein composition of a venom sample collected by manual milking may closely resemble that of venom injected during a natural honeybee sting: in addition to the release of venom proteins which are produced by the venom glands, also proteins originating from other tissues such as the sting apparatus cell lining, stinger lancets and/or stinger lubricant, which has been hypothesized to be generated by the Dufour gland [25], may be released. As such, we were uncertain about the tissue origin of the apidaecin peptide detected in the venom sample. This issue was resolved by the detection of apidaecin transcripts (by RT-PCR) and an apidaecin peptide (by peptidomics) in the venom gland tissue, which indicates that this peptide is produced by the venom glands.

Apidaecins were firstly discovered by Casteels *et al.* [26] in honeybee lymph upon bacterial infection and were described to be expressed by hemocytes [14]. They are small, proline-rich antibacterial peptides which are generated by processing of single precursor proteins. This multipeptide precursor structure allows to amplify the insects' immune response upon bacterial challenge [14]. Moreover, its genomic structure enables the development of a pathogen-specific response by splice variation [27]. Different isoforms are described which all derive from a single prepro-protein. The peptides detected in the venom and venom glands correspond to the apidaecin isoforms Ia or Ib [28] and may also play important antimicrobial roles. As antimicrobial peptide expression by barrier epithelial cell linings of multiple tissues seems to be a general feature of host defense in multicellular organisms [29], apidaecin expression by venom apparatus epithelial cells may protect the individual honeybee against invading pathogens. Alternatively, its presence in the venom may also have a function in the social immunity of the hive. Indeed, as honeybee
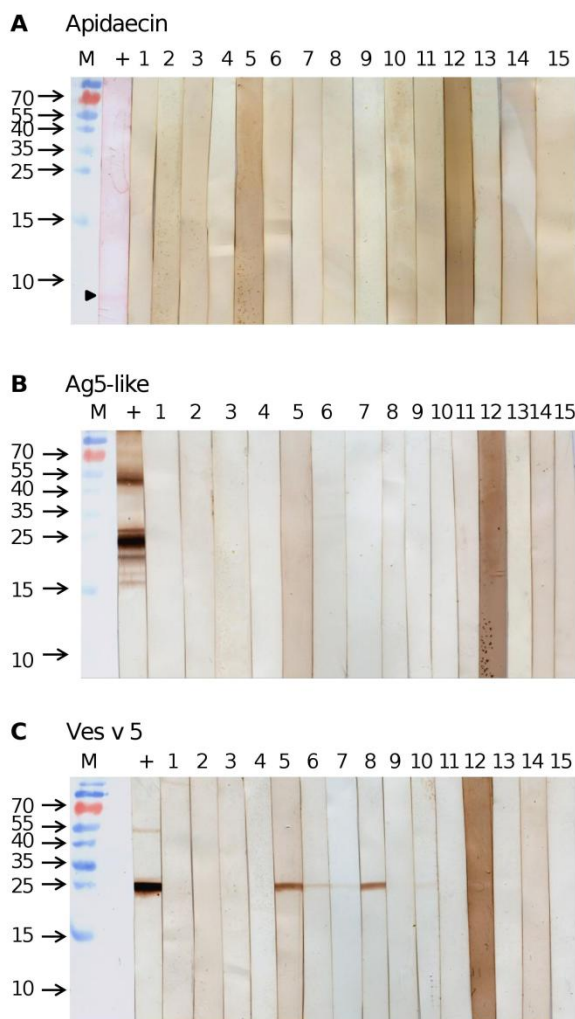
Figure 1.5: IgG4 responses to apidaecin, honeybee Ag5-like protein and yellow jacket Ves v 5 in beekeepers. Pure proteins were separated by SDS-PAGE under reducing and denaturing conditions and transferred to PVDF. Membranes were incubated with sera of beekeepers (strip 1-10) and negative control sera of subjects never stung by hymenopterans (strip 11-15), followed by enzyme-linked anti-human IgG4. M= PageRuler™ Prestained Protein ladder (Fermentas). Molecular weights (kDa) of the marker bands are shown. Positive controls (+) are executed by Ponceau (apidaecin band indicated by an arrowhead) and anti-His staining (Ag5-like protein and Ves v 5).

venom is present on the cuticle of adult bees and on comb wax, it has been suggested that it may act as a social antiseptic device [22;30]. Unlike the reported lytic antibacterial activity of other venom peptides such as melittin and possibly also MCDP [22;31], apidaecin kills bacteria through a bacteriostatic process. It is predominantly active against many Gram-negative bacteria by special antibacterial mechanisms [31]. Consequently, apidaecin may be one of the peptides playing an important role in protection of the hive against these pathogenic bacteria.

Second, this study focuses on the identification of an Ag5-like gene transcript expressed by the honeybee venom glands. Ag5s are important and highly abundant venom allergens within the Vespidae and Formicidae families. Remarkably, an Ag5 protein has never been detected in venom proteome analyses of honeybees or any other member of the Apidae family. Based on an NCBI prediction, we were able to clone, sequence and recombinantly produce an Ag5-like sequence expressed by the venom gland of winter honeybees. However, it seems that venom gland tissue of summer and BFR honeybees do not express this protein, whereas its expression is maintained in the brain tissue. So far, most proteomic studies focused on venom of summer bees, which may explain why it remained undetected. Because of the putative presence of a signal peptide, we hypothesize that the honeybee Ag5-like protein is secreted by the venom gland of winter bees. A proteomic study focusing on the venom of winter bees should further confirm this hypothesis.

Additionally, we demonstrated Ag5-like expression in the hypopharyngeal glands and the midgut of summer bees, while expression is lacking in hemocytes, salivary gland, mucus glands (drones), devisrated abdomen and muscle tissue. Midgut expression has also been reported in *Drosophila* [11]. On the other hand, honeybee Ag5-like protein is not expressed in salivary glands, while Ag5s have been found in the saliva of blood feeding Diptera such as ticks, sand flies, stable flies and mosquitoes [11]. Unfortunately, the function of any of the Ag5s remains unknown [11], which makes it difficult to explain this spatial and seasonal variation pattern in the honeybee.

Third, our analysis revealed the lack of IgG4 recognition of both apidaecin and honeybee Ag5-like protein by the beekeepers' sera. Beekeepers are regularly stung in summer time, which is known to cause a strong venom-specific IgG4 response [2]. We are unable to conclude whether this lack in humoral response against both compounds is the result of low immunogenicity, low abundance in the venom and/or low exposure. In case of the antimicrobial peptide apidaecin a low immunogenicity can be explained by its short length [32]. For the honeybee Ag5-like protein, its restricted expression in winter time certainly lowers the exposure to this venom compound significantly, as beekeepers are then hardly stung. However, we cannot exclude that the immunogenicity of the natural protein differs from that of the tested recombinant due to a possible incorrect folding of the latter. An incorrect folding of bacterial recombinants may be caused by the lack of post-translational

modifications such as glycosylations and disulfide bridges [33]. The absence of glycosylations cannot be responsible for an incorrect folding of the Ag5-like recombinant as glycosylation sites are absent (inferred by NetNGlyc 1.0 and NetOGlyc 3.0). In contrast, four disulfide bridges are predicted to be present (inferred by DISULFIND). As the bacterial recombinant was produced in the cytoplasm, it may lack these disulfide bridges [33]. Moreover, for the immunoblot experiment, electrophoretic separation has been conducted under reducing and denaturing conditions, which generally disrupts the protein conformation. As such, the lack of IgG4 recognition of the recombinant Ag5-like protein may be the result of the loss of relevant discontinuous B-cell epitopes, in contrast to the continuous epitopes which have been preserved. However, conformational epitope renaturation during or after transfer of the protein to the Western blot membrane has been described [34]. Further experiments should clarify if the honeybee venom Ag5-like protein truly lacks immunoreactivity. Also, a preceding study showed that the conformation of the Ag5 from wasp venom, Ves v 5, differs between the bacterial recombinant and the natural protein, and that refolding strategies were needed to obtain its full immunoreactivity [35]. Therefore, refolding strategies may also help to obtain a correctly folded Ag5-like bacterial recombinant. This can be confirmed by comparing the solubility, electrophoretic behavior, disulfide content and circular dichroism-spectrum between natural and recombinant Ag5-like protein. Besides, yeast (*Pichia pastoris*) or insect (Sf9) cell expression systems were shown to be good alternatives for production of a correctly folded Ves v 5 [16;36]. Next, immunoreactivity should be analyzed under non-reducing and non-denaturing conditions, for example by ELISA.

Also, no IgG4 cross-reactivity between the Ag5-like protein and Ves v 5 was detected. Although two expression systems with differential capacities to perform post-translational modifications were used to produce these recombinants, we believe this may not have influenced the outcome of our immunoblot experiment. As also Ves v 5 lacks glycosylation sites, both proteins were produced without carbohydrate groups. In contrast, while the bacterial system probably has not foreseen the Ag5-like recombinant with the correct disulfide bridges, Ves v 5 produced in the baculovirus system is secreted and likely contains appropriate disulfide bridges. However, this difference has been neutralized in our immunoblot experiment by the electrophoretic separation of both proteins under reducing and denaturing conditions. As such, we hypothesize that the low sequence identity between both compounds (25% sequence identity, Figure 1.6) plays a more significant role. In addition, the honeybee

Ag5-like sequence shows low sequence identity (max 27%) with other hymenopteran venom Ag5 allergens, a remarkable characteristic which was resolved by our phylogenetic analysis. It appears that hymenopteran Ag5s have a complex evolutionary history with frequent gene duplications and losses, and that the honeybee Ag5-like protein is clearly paralogous to the group of ant and wasp venom allergens (Figure 1.4, resp. groups D and A). Sequence identity is generally much higher between the orthologous group of venom allergens, which may be responsible for the observed IgE cross-reactivities between venom Ag5s of the *Vespula* genus and even between *Vespula* and *Dolichovespula* Ag5s and between *Vespula* and *Vespa* Ag5s [37]. Most likely, IgE-level cross-reactivity between the honeybee Ag5-like protein and these other hymenopteran venom Ag5s may also be lacking due to this low sequence identity.

```
VES_v_5    MEISGLVYLIIIVTIIDLPYGKANNYCKIKCLKGGVHTACKYGSLKPNCGNKVVVSYGLT 60
Ag5_like   MAREGIIRLTLIFIFVSTGS--------------------ADVISCIGKALLKSDQVS 38
           *  .*:: * :*. ::.                        ..: .  *: :: *  ::

VES_v_5    KQEKQDILKEHNDFRQKIARGLETRGNPGPQPPAKNMKNLVWNDELAYVAQVWANQCQYG 120
Ag5_like   CQDKQLILDEHNSLREKISFG-EIQG----MPSAANMRELTWDDELADMAQNWATKCAEI 93
            *:** **.***.:*:**: * * :*     *.* **::*.*:**** :** **.:*

VES_v_5    HDTCRDVAKYQVGQNVALTGSTAAKYDDPVKLVKMWEDEVKDYNPKKKFSGNDFLKTGHY 180
Ag5_like   RDPFRHIRGFYVEQNMLKTISSTQFNDLPD-----WRQTIYNWFNEVQYKGG----ISHY 144
           :*. *.:  : * **:  * *::   * *      *.: : ::  : ::.*.    .**

VES_v_5    TQMVWANTKEVGCGSIKYIQEKW-HKHYLVCNYGPSGNFMNEELYQTK----------- 227
Ag5_like   NQIFWANSYLIGCGFTYYFDAGHGYVKNYVCNYAPGDNVDGHLPYEQGQPACSNYKMTSS 204
           .*:.***:  :***   *::     : :  ****.*..*. ..  *:

VES_v_5    ------------------
Ag5_like   SRFRGLCSAGTNNYLGTRC 223
```

Figure 1.6: Protein sequence alignment of yellow jacket Ves v 5 (Ves_v_5; Q05110.1) and detected honeybee Ag5-like (XP_001122516.2) by ClustalW. They have a sequence identity of 25%.

Whereas traditional diagnostic tools rely on whole venom preparations, the so-called component resolved diagnosis (CRD) allows to determine the patients' allergen recognition profile. Originally aimed at adapting the immunotherapy to the patients-specific profile, this approach also allows to determine the culprit species, a problem that often raises because the patients fail to identify or name the hymenopteran species that stung and because presently used diagnostic tests based on whole venom often reveal a false double positivity to multiple species due to their similar cross-reactive allergens and cross-reactive carbohydrate determinants (CCD's) [38]. CRD using species-specific allergens may solve this issue. In many European countries, the European honeybee (*A. mellifera*) and yellow jackets (*V. vulgaris*) are

the most prevalent stinging insects. As such, several research studies have recently focused on their differential allergy diagnosis by CRD [16;39;40]. One of the important differential allergens is Ag5 which is a major venom allergen of the yellow jacket, while so far no honeybee venom Ag5 homologue had been described. Our present finding on the occurrence of a paralogue of Ves v 5 in the bee venom gland and on its peculiar expression restricted to winter time is an important observation. Due to the limited contact between humans and winter bees, we hypothesize that sera of honeybee venom allergic patients lack specific IgE antibodies to the honeybee Ag5-like protein. Moreover, cross-reactivities with wasp and ant venom Ag5s may not be present due to a low sequence identity. As such, our findings are so far in favor of a differential diagnosis of sting allergy by CRD.

1.7 ACKNOWLEDGEMENTS

1.8 ADDENDUM

Supplementary figures and tables can be found on the included CD-ROM or can be requested by e-mail from matthiasvanvaerenbergh@hotmail.com and Dirk.deGraaf@UGent.be.

Figure S1.1: Peptide mass fingerprint and MS/MS fragmentation spectrum of the identified apidaecin peptide in venom gland tissue.

Figure S1.2: Nucleotide and protein sequence alignments of the cloned honeybee Ag5-like sequence with the NCBI prediction (by ClustalW).

Figure S1.3: Coomassie blue staining of SDS-PAGE separated synthetic apidaecin peptide and purified recombinants uterocalin, Ves v 5 and honeybee Ag5-like protein.

Table S1.1: Information about beekeeper sera.

1.9 REFERENCES

[1] Larche M, Akdis CA, Valenta R. Immunological mechanisms of allergen-specific immunotherapy. Nat Rev Immunol 2006 Oct;6(10):761-71.
[2] Muller UR. Bee venom allergy in beekeepers and their family members. Curr Opin Allergy Clin Immunol 2005 Aug;5(4):343-7.
[3] Weinstock GM, Robinson GE, Gibbs RA, Worley KC, Evans JD, Maleszka R, et al. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 2006 Oct 26;443(7114):931-49.
[4] Peiren N, de Graaf DC, Brunain M, Bridts CH, Ebo DG, Stevens WJ, et al. Molecular cloning and expression of icarapin, a novel IgE-binding bee venom protein. FEBS Lett 2006 Sep 4;580(20):4895-9.
[5] de Graaf DC, Brunain M, Scharlaken B, Peiren N, Devreese B, Ebo DG, et al. Two novel proteins expressed by the venom glands of *Apis mellifera* and *Nasonia vitripennis* share an ancient C1q-like domain. Insect Mol Biol 2010 Feb;19 Suppl 1:1-10.
[6] Peiren N, de Graaf DC, Evans JD, Jacobs FJ. Genomic and transcriptional analysis of protein heterogeneity of the honeybee venom allergen Api m 6. Insect Mol Biol 2006 Oct;15(5):577-81.
[7] Hoffman DR. Hymenoptera venom allergens. Clin Rev Allergy Immunol 2006 Apr;30(2):109-28.
[8] Blank S, Seismann H, Bockisch B, Braren I, Cifuentes L, McIntyre M, et al. Identification, recombinant expression, and characterization of the 100 kDa high molecular weight Hymenoptera venom allergens Api m 5 and Ves v 3. J Immunol 2010 May 1;184(9):5403-13.
[9] Peiren N, Vanrobaeys F, de Graaf DC, Devreese B, Van BJ, Jacobs FJ. The protein composition of honeybee venom reconsidered by a proteomic approach. Biochim Biophys Acta 2005 Aug 31;1752(1):1-5.
[10] Peiren N, de Graaf DC, Vanrobaeys F, Danneels EL, Devreese B, Van BJ, et al. Proteomic analysis of the honey bee worker venom gland focusing on the mechanisms of protection against tissue damage. Toxicon 2008 Jul;52(1):72-83.
[11] Gibbs GM, Roelants K, O'Bryan MK. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins--roles in reproduction, cancer, and immune defense. Endocr Rev 2008 Dec;29(7):865-97.
[12] Audsley N, Weaver RJ. Analysis of peptides in the brain and corpora cardiaca-corpora allata of the honey bee, *Apis mellifera* using MALDI-TOF mass spectrometry. Peptides 2006;27(3):512-20.
[13] Hummon AB, Richmond TA, Verleyen P, Baggerman G, Huybrechts J, Ewing MA, et al. From the Genome to the Proteome: Uncovering Peptides in the Apis Brain. Science 2006 Oct 27;314(5799):647-9.
[14] Casteels-Josson K, Capaci T, Casteels P, Tempst P. Apidaecin multipeptide precursor structure: a putative mechanism for amplification of the insect antibacterial response. EMBO J 1993 Apr;12(4):1569-78.

[15] Emanuelsson O, Brunak S, von HG, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2007;2(4):953-71.

[16] Seismann H, Blank S, Cifuentes L, Braren I, Bredehorst R, Grunwald T, et al. Recombinant phospholipase A1 (Ves v 1) from yellow jacket venom for improved diagnosis of hymenoptera venom hypersensitivity. Clin Mol Allergy 2010;8:7.

[17] Suire S, Stewart F, Beauchamp J, Kennedy MW. Uterocalin, a lipocalin provisioning the preattachment equine conceptus: fatty acid and retinol binding properties, and structural characterization. Biochem J 2001 Jun 1;356(Pt 2):369-76.

[18] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. 2004 Mar.

[19] Jones DT. The Rapid Generation of Mutation Data Matrices from Protein Sequences. 1992 Jun.

[20] Tamura K. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. 2011 Oct.

[21] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 2011 Oct;28(10):2731-9.

[22] Baracchi D, Turillazzi S. Differences in venom and cuticular peptides in individuals of *Apis mellifera* (Hymenoptera: Apidae) determined by MALDI-TOF MS. J Insect Physiol 2010 Apr;56(4):366-75.

[23] Matysiak J, Schmelzer CE, Neubert RH, Kokot ZJ. Characterization of honeybee venom by MALDI-TOF and nanoESI-QqTOF mass spectrometry. J Pharm Biomed Anal 2011 Jan 25;54(2):273-8.

[24] Stöcklin R, Favreau P. Proteomics of venom peptides. In: Ménez A, editor. Perspectives in Molecuar Toxinology. Wiley; 2002. p. 106-23.

[25] Martin SJ, Dils V, Billen J. Morphology of the Dufour gland within the honey bee sting gland complex. Apidologie 2005 Oct;36(4):543-6.

[26] Casteels P, Ampe C, Jacobs F, Vaeck M, Tempst P. Apidaecins: antibacterial peptides from honeybees. EMBO J 1989 Aug;8(8):2387-91.

[27] Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, et al. Immune pathways and defence mechanisms in honey bees *Apis mellifera*. Insect Mol Biol 2006 Oct;15(5):645-56.

[28] Casteels P, Romagnolo J, Castle M, Casteels-Josson K, Erdjument-Bromage H, Tempst P. Biodiversity of apidaecin-type peptide antibiotics. Prospects of manipulating the antibacterial spectrum and combating acquired resistance. J Biol Chem 1994 Oct 21;269(42):26107-15.

[29] Tzou P, Ohresser S, Ferrandon D, Capovilla M, Reichhart JM, Lemaitre B, et al. Tissue-specific inducible expression of antimicrobial peptide genes in *Drosophila* surface epithelia. Immunity 2000 Nov;13(5):737-48.

[30] Baracchi D, Francese S, Turillazzi S. Beyond the antipredatory defence: honey bee venom function as a component of social immunity. Toxicon 2011 Nov;58(6-7):550-7.

[31] Li WF, Ma GX, Zhou XX. Apidaecin-type peptides: biodiversity, structure-function relationships and mode of action. Peptides 2006 Sep;27(9):2350-9.

[32] Chiarella P, Edelmann B, Fazio VM, Sawyer AM, de Marco A. Antigenic features of protein carriers commonly used in immunisation trials. Biotechnology Letters 2010 Sep;32(9):1215-21.

[33] Sahdev S, Khattar SK, Saini KS. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem 2008 Jan;307(1-2):249-64.

[34] Zhou YH, Chen ZC, Purcell RH, Emerson SU. Positive reactions on Western blots do not necessarily indicate the epitopes on antigens are continuous. Immunology and Cell Biology 2007 Jan;85(1):73-8.

[35] Suck R, Weber B, Kahlert H, Hagen S, Cromwell O, Fiebig H. Purification and immunobiochemical characterization of folding variants of the recombinant major wasp allergen Ves v 5 (Antigen 5). International Archives of Allergy and Immunology 2000 Apr;121(4):284-91.

[36] Monsalve RI, Lu G, King TP. Expressions of recombinant venom allergen, antigen 5 of yellowjacket (*Vespula vulgaris*) and paper wasp (Polistes annularis), in bacteria or yeast. Protein Expression and Purification 1999 Aug;16(3):410-6.

[37] Henriksen A, King TP, Mirza O, Monsalve RI, Meno K, Ipsen H, et al. Major venom allergen of yellow jackets, Ves v 5: structural characterization of a pathogenesis-related protein superfamily. Proteins 2001 Dec 1;45(4):438-48.

[38]   de Graaf DC, 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 Mar 6;72(2):145-54.

[39]   Mittermann I, Zidarn M, Silar M, Markovic-Housley Z, Aberer W, Korosec P, et al. Recombinant allergen-based IgE testing to distinguish bee and wasp allergy. J Allergy Clin Immunol 2010 Jun;125(6):1300-7.

[40]   Müller UR, Johansen N, Petersen AB, Fromberg-Nielsen J, Haeberli G. Hymenoptera venom allergy: analysis of double positivity to honey bee and Vespula venom by estimation of IgE antibodies to species-specific major allergens Api m1 and Ves v5. Allergy 2009 Apr;64(4):543-8.

Chapter 2

Exploring the hidden honeybee (*Apis mellifera*) venom proteome by integrating a combinatorial peptide ligand library approach with FTMS

The work presented in Chapter 2 was adapted from the following manuscripts:

1) *M. Van Vaerenbergh, G. Debyser, B. Devreese, D. C. de Graaf. Exploring the hidden honeybee (Apis mellifera) venom proteome by integrating a combinatorial peptide ligand library approach with FTMS. Journal of Proteomics, in press. This work will be published as a companion paper of the unpublished manuscript of the Honeybee Genome Sequencing and Analysis Consortium, reporting on the re-sequencing of the honeybee genome:*

2) *Honeybee Genome Sequencing and Analysis Consortium. Finding the missing honey bee genes: lessons learned from a genome upgrade. Unpublished work.*

2.1 CONTRIBUTIONS

D. de Graaf and B. Devreese assisted with the study design. M. Van Vaerenbergh executed all experiments. G. Debyser provided technical assistance for the LC-ESI-LTQ-FT-ICR-MS experiments and setting up Mascot searches. L. De Smet assisted during cutting out gel slices from the SDS-PAGE gels. K. Morreel calibrated the FT-ICR and LTQ mass analyzers. M. Van Vaerenbergh performed all data analysis and gene annotation.

M. Van Vaerenbergh wrote the manuscript reporting on the identification of novel honeybee venom compounds. This manuscript is accepted for publication in Journal of Proteomics and will be published as a companion paper of the main genome paper from the Honeybee Genome Sequencing and Analysis Consortium, which will publish the data of a re-sequencing of the honeybee genome. Publication of this companion paper is put on hold by

Journal of Proteomics until the main genome paper will be released. M. Van Vaerenbergh also wrote a section about the annotation of venom genes, the contribution of improved gene predictions to the identification of new venom proteins, and genome mining to discover the tertiapin gene, which will be included in the main genome paper manuscript. All authors of the companion paper will also be included in the author list of the main genome paper, known as the Honeybee Genome Sequencing and Analysis Consortium. The co-authors assisted throughout the writing phase of the companion paper and section of the main genome paper.

2.2 ABSTRACT

At present, 30 compounds have been described in the venom of the honeybee, and 16 of them were confirmed by mass spectrometry. Previous studies typically combined 2-D PAGE with MALDI-TOF/TOF MS, a technology which now appears to lack sensitivity to detect additional venom compounds. Here, we report an in-depth study of the honeybee venom proteome using a combinatorial peptide ligand library sample pretreatment to enrich for minor components followed by shotgun LC-FT-ICR MS analysis. This strategy revealed an unexpectedly rich venom composition: in total 102 proteins and peptides were found, with 83 of them never described in bee venom samples before. Based on their predicted function and subcellular location, the proteins could be divided into two groups. A group of 33 putative toxins is proposed to contribute to venom activity by exerting toxic functions or by playing a role in social immunity. The other group, considered as venom trace molecules, appears to be secreted for their functions in the extracellular space, or are unintentionally secreted by the venom gland cells due to insufficient protein recycling or co-secretion with other compounds. In conclusion, our approach allowed to explore the hidden honeybee venom proteome and extended the list of potential venom allergens.

2.3 GRAPHICAL ABSTRACT
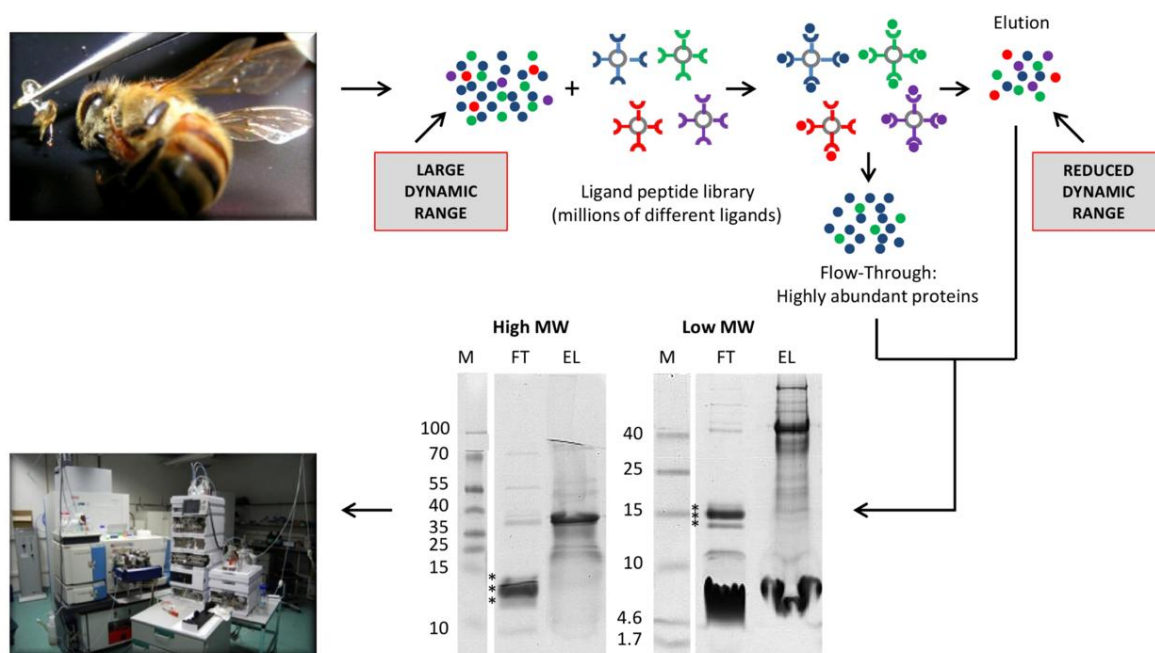
2.4 INTRODUCTION

Honeybee venom is composed of a mixture of biogenic amines, peptides and proteins. The venom causes local tissue damage, which induces death in other insects and pain and inflammation in higher organisms [1]. Worker bees use their sting apparatus in order to defend the colony and their food stock. Moreover, as it was recently demonstrated that in the bee hive venom peptides are smeared on the body surface of females and on wax combs, an additional function of bee venom in social immunity has been hypothesized [2].

Early efforts to unravel the bee venom composition date back to the work of Langer in 1897 [3]. Today, up to 30 venom proteins and peptides are described [4-22]. However, some of those lack a proper characterization. Several venom constituents isolated in the 1970's and 80's by chromatographic means are described by their enzymatic activity or amino acid composition, but amino acid sequences are often lacking, e.g. for minimine [22], cardiopep [5] and adolapin [6]. The development of proteomic methods, and later of the *Apis mellifera* genome [23], has boosted the detection of new bee venom compounds significantly [7;9;11;19;20].

Preceding studies investigating the honeybee venom proteome often combined 2-DE with MALDI-TOF/TOF MS [19;21]. However, this method lacks dynamic range and sensitivity to

allow the detection of lowly abundant compounds. In addition, low molecular weight fractions remain largely unexplored due to poor resolution in classical SDS-PAGE separation. Therefore, previous studies suggested the existence of yet unknown venom compounds [19;21]. For example, a large part of a typical honeybee 2D-gel is obscured due to highly abundant compounds, such as phospholipase A_2 (PLA$_2$; 10–12% of the venom dry weight [24]), that mask the detection of lowly abundant compounds with similar molecular weight and p*I*.

Several studies have shown that the use of a combinatorial peptide ligand library (CPLL) can significantly improve the coverage of proteomic analyses as this allows to access many lowly abundant compounds in complex proteomes [25]. The method has been used in studies of the 'deep' venom proteome of two snake species, namely the Western diamondback rattlesnake (*Crotalus atrox*) [25] and the African puff adder (*Bitis arietans*) [26;27] and led to the discovery of a large number of proteins previously undetected in these proteomes. In this work, we now adopted CPLL that has so far not been used in Hymenoptera venomics.

The modular arrangement of MALDI and ESI ionization with different types of mass analyzers has resulted in a wide variety of mass spectrometric instrumentation [28]. Many of them have been used in honeybee venom research, except FTMS, although this equipment provides the highest performance in mass resolution and mass accuracy [29]. In the context of venom research, only few research groups used this technology (Orbitrap or FT-ICR) for protein identifications within the entire venom [30-34]. The discovery of a high number of new venom constituents in these studies points to a highly effective technology for identification purposes in these complex protein mixtures.

Since 2006, the honeybee genome became available [23], which provided significant benefits for protein identifications as the mass spectra can be searched against the available protein predictions. However, the first generated genome sequence was noted to have a bimodal GC content that affected the quality of the assembly in some regions and the annotation had fewer genes in the initial gene set (OGSv1.0) than would have been expected based on other insect genomes sequenced since then [35]. Therefore, while the previous genome sequence was obtained by Sanger sequencing and a whole-genome-shotgun model, the honeybee genome was recently re-sequenced using next-generation sequencing which allows a much deeper sequence coverage. This resulted in an improved genome assembly (Amel_4.5), which is more contiguous and complete, and a new gene annotation set

(OGSv3.2), which includes ~5000 more protein-coding genes, increasing the gene set by about 50% [35]. Therefore, searching generated venom mass spectra against this improved dataset may identify a new set of venom proteins.

The present study aimed an in-depth analysis of the honeybee venom proteome by merging CPLL sample pretreatment and nanoLC FT-ICR MS/MS. The CPLL flow-through and elution samples were separated using 1D-SDS-PAGE. Then, proteins and peptides were identified by a complete slice-by-slice LC-ESI-LTQ-FT-ICR MS/MS analysis of tryptic peptides. Such a sample decomplexation/fractionation before mass spectrometry is the best approach for maximum protein coverage [33]. Functions of the identified compounds were predicted using bioinformatics. All venom genes were annotated on the improved honeybee genome assembly [35] and the contribution of the improved gene predictions to the identification of novel venom proteins was determined.

2.5 MATERIALS AND METHODS

2.5.1 Venom collection

Mid July 2011, adult worker honeybees (*Apis mellifera carnica*) were collected at the hive entrance. Pure venom was collected as previously described [19]. Venom of 150 honeybees was pooled to a protein concentration of 69.54 mg/ml, as determined by Bradford protein assay (Thermo Scientific Pierce, Hudson, NH, USA).

2.5.2 Protein enrichment

The dynamic range of protein concentrations in the honeybee venom sample was compressed by a CPLL approach (ProteoMiner protein enrichment small-capacity kit, Bio-Rad Laboratories, Hercules, CA, USA). This experiment was performed according to the instructions of the manufacturer. In brief, 200 µl of venom sample was added to the beads for 2 hours at room temperature on a rotational shaker. Subsequently, the non-binding fraction was collected (=flow-through). Non-specific binding components were removed by 3 rounds of washing (150 mM NaCl, 10 mM NaH_2PO_4, pH 7.4). Bound proteins were eluted with 3 steps of 20 µl elution buffer (8 M urea, 2% CHAPS, 5% acetic acid) and pooled.

2.5.3 1D-SDS-PAGE

Approximately 100 µg of both flow-through and eluted protein fractions were loaded on a 10% Tris-glycine-SDS-PAGE gel and separation was carried out at 140V. In addition, 25 µg of flow-through proteins and 50 µg of elution proteins were separated at 100V on a 16.5% Tris-tricine-SDS-PAGE gel. Gel separation was carried out on a Mini protean 3 system (Bio-Rad Laboratories, Hercules, CA, USA) and was continued until the blue bromophenol front reached the bottom of the gels. Tricine gel fixation was performed in 0.3% TCA for 30 min. Both gels were stained with Coomassie Brilliant Blue G250 and the background was destained with 30% MeOH. Proteins were reduced in-gel by adding 10 mM DTT/25 mM NH_4HCO_3 (56°C for 45 min) and alkylated in 55 mM iodoacetamide/25 mM NH_4HCO_3 (RT for 45 min). Subsequently, the gel was washed in 25 mM NH_4HCO_3. All flow-through and elution protein bands larger than 40 kDa were cut out of the 10% glycine gel, while those smaller than 40 kDa were cut out of the 16.5% tricine gel. Also gel parts without any visible protein bands were excised and analyzed. Residual Coomassie staining was removed by washing the gel pieces in 150 µl of 200 mM NH_4HCO_3/50% ACN for 30 min at 37°C. Gel pieces were dried in a speedvac (Thermo Savant, Holbrook, NY, USA).

2.5.4 In-gel digest

An in-gel tryptic digest was performed by adding 12 µl of trypsin solution (0.002 µg/µl in 50 mM NH_4HCO_3; sequencing grade modified trypsin, Promega, Madison, WI, USA) to each gel piece. After overnight incubation at 37°C, the solution with hydrophilic tryptic peptides was collected. Hydrophobic tryptic peptides were extracted from the gel by two subsequent incubation steps (15 min at 30 °C) with respectively 60 and 40 µl of 60% ACN/0.1% formic acid. Hydrophilic and hydrophobic peptides of each gel piece were pooled, dried by speedvac and dissolved in 15 µl 2% ACN/0.1% formic acid.

2.5.5 LC-ESI-LTQ-FT-ICR-MS

Five µl of the tryptic peptide fractions were analyzed, with the exception of fractions obtained from heavily stained gel bands for which only 2 µl was injected. LC-ESI-FT MS analysis was performed as described in previous research [32]. Subsequently, raw LC-MS/MS data were analyzed using the Mascot v2.3 search engine (Matrix Science, London, UK). MS/MS data were searched against the Amel4.5 NCBI Refseq (available at

ftp://ftp.ncbi.nih.gov/genomes/Apis_mellifera/protein/; database contains 10570 sequences and 5504336 residuals) and Augustus9 (available at http://www.hgsc.bcm.tmc.edu/ftp-archive/Amellifera/Amel_4.5GenePredictions/Augustus/; database contains 11560 sequences and 7698817 residuals) protein prediction databases [35]. An automatic decoy database search was performed to enable false discovery rate (FDR) determination. The significance threshold was adapted to 0.001 to reach a FDR<1% for the identity threshold of both database searches. Searches were executed with carbamidomethylation of cysteines as a fixed modification and oxidation of methionines as a variable modification. One tryptic miscleavage was permitted and peptide mass tolerance and MS/MS tolerance were set to 10 ppm and 0.3 Da, respectively. Precursor peptide charge state was set to 2+ and 3+.

2.5.6 Criteria for positive identifications

Setting the significance threshold at p<0.001 led to a FDR of 0.32% for the Amel4.5 NCBI Refseq search and 0.93% for the Augustus9 search. We defined positive protein identifications as queries detected by at least two unique, bold and red (significant and top ranking) peptides from the Mascot output with an ion score ≥30. In addition, the discovery of small peptides was enabled by allowing queries with a sequence coverage higher than 10% due to the detection of only one, bold and red (significant and top ranking) peptide with an ion score ≥30. All protein identifications were merged in one list and all double identifiers were removed.

2.5.7 Sequence analysis

As not all sequences are correct in prediction datasets, we tried to determine the correct protein sequence for each identification. First, this was done by searching for available EST data by Blast searches against the Amel_4.5 scaffolds on Beebase [36], which shows available honeybee ESTs mapped on the genome. Second, UniProt blast searches [37] were performed to find homologues in well-annotated species such as *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens*. Sequence identity of the honeybee venom predictions and their homologues was evaluated using the ClustalW software (standard parameters; [38]). The combination of honeybee EST evidence and homology-based evidence was used to determine the correct protein sequences. A database containing all correct(ed) protein sequences was constructed, which was used for performing bioinformatic analyses further described in this section.

The presence of an N-terminal secretion signal peptide was verified using the SignalP4.0 server [39]. Sequences were truncated to a length of 200 amino acids and the D-cutoff for signalP-TM networks was set to 0.350. Next, to determine a putative function for protein predictions, protein signatures were searched by InterProScan [40] and peptidase and protease inhibitor families were searched by Blast MEROPS [41]. GO-terms were assigned using Blast2GO v.2.6.0 [42]: honeybee venom proteins were subjected to a BlastP against the Swiss-Prot database using an expect value of E= 10^{-3}. Mapping and GO-annotation were performed using default parameters. Proteins existing in exosomes were searched in the exosome protein database ExoCarta (http://www.exocarta.org/). Also, similar proteins present in venoms of other species were searched by a stand-alone Blast search of the Apis venom sequence database against a constructed database containing all venom proteins present in GenBank, but lacking predicted venom sequences, *Apis mellifera* sequences and patent sequences.

2.5.8 Annotation of venom genes and contribution of improved gene prediction datasets to the identification of new venom proteins

The correct sequence of every identified honeybee venom compound was determined by the combination of honeybee EST evidence, homology-based evidence (see section 2.4.7) and peptide information. Venom genes were manually annotated on the improved honeybee genome assembly using Apollo for the *A. mellifera* assembly Amel_4.5 [35].

To determine the contribution of the improved gene prediction set to the identification of novel venom compounds, the venom mass spectra were searched against the newest gene set (OGSv3.2) and the initial official gene set (OGSv1.0) [35]. Search parameters and FDRs were identical to those described higher (see section 2.4.5).

2.6 RESULTS AND DISCUSSION

2.6.1 Identification of honeybee venom proteins

Our in-depth analysis revealed an unexpectedly rich composition of the honeybee worker venom. The detection of 705 unique tryptic peptides provides biological evidence for 102 venom proteins and peptides (Table S2.1 and S2.2). This list includes 19 compounds found in preceding honeybee venom proteome analyses and 6 additional compounds described in a

study of the honeybee venom gland proteome (Tables 2.1 and 2.2). Interestingly, we also present the sequence data of three venom enzymes, yet only identified by enzymatic activity tests in studies published more than 30 years ago: the glucosidase 2, β-galactosidase and group XV PLA$_2$ sequences may represent the enzymes catalyzing the α-glucosidase [4], β-galactosidase [15] and lysophospholipase [16;17] enzymatic reactions, respectively. In addition, this study detected 83 new venom compounds, which was enabled by the combined use of the CPLL technology, the high performance mass spectrometric instrument (LC-ESI-FT-ICR-MS) and the improved honeybee gene prediction sets [35]. Indeed, the CPLL pretreatment clearly decreased the dynamic range of protein concentrations in the venom sample (Figure 2.1). PLA$_2$ and melittin contents, comprising 10-12% and 50% of the total venom dry weight respectively [24], were diminished in the elution fraction while also a broad range of bands appeared which are not visible upon separation of an untreated venom sample. Besides, while previous honeybee venom mass spectrometry studies focused on a specific molecular weight range ([7]: 950-4000Da; [11]: 750-15000Da; [19-21]: SDS-PAGE gel lacks low molecular weight fraction), this study extended its search towards the complete molecular weight range.

Despite the high number of identifications, some of the previously reported honeybee venom compounds are missing in this study. The absence of the antigen5-like wasp venom paralogue is not surprising as it has only been described as a venom gland transcript in winter bees [18], while this study focused on venom of summer bees. Also, four compounds described long ago (cardiopep [5], minimine [22], adolapin [6] and β-acetylaminodeoxyglucosidase [15]) are missing, as their sequence information has so far not been determined. In addition, small peptides, such as tertiapin [13], procamine [12], apidaecin [18] and mast cell degranulating peptide (MCDP) [11] are difficult to detect using our approach, although a MCDP tryptic peptide (HVIKPHICR) with an ion score of 26 was detected at less stringent search parameters (p<0.01 and FDR of 4.05%). Finally, two high molecular weight proteins, hexamerin [20] and vitellogenin [43], may be lacking due to venom sample variation (different collection method, spatial and/or seasonal venom variation) or technological variation (liquid versus gel-based proteomics).
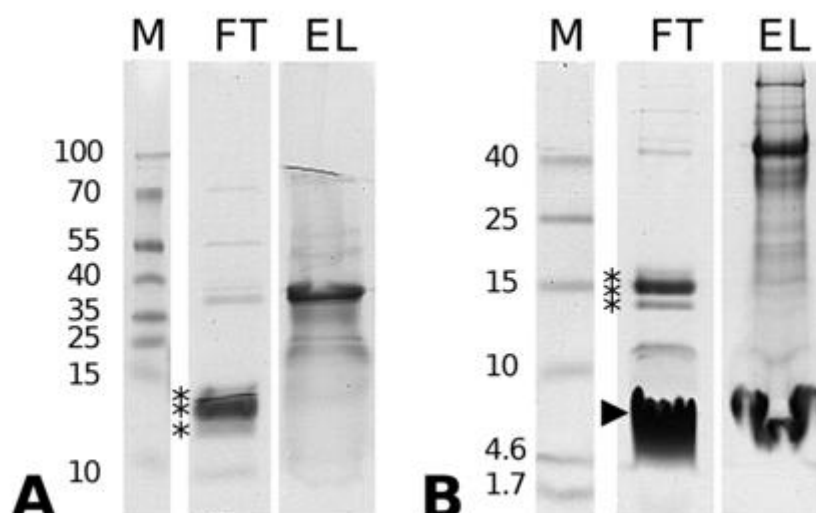
Figure 2.1: Electrophoretic separation of a combinatorial peptide ligand library (CPLL)-treated honeybee venom sample. CPLL flow-through (=FT) and elution (=EL) samples are separated on a 10% Tris-glycine-SDS-PAGE gel (A) and a 16.5% Tris-tricine-SDS-PAGE gel (B). Molecular weight regions which are known to contain high amounts of PLA$_2$ (*) and melittin (▶) are indicated. Molecular weights (in kDa) of the markers (Thermo Scientific, Rockford, IL, USA) are indicated in the figure: A) PageRuler Prestained Protein Ladder; B) Spectra Multicolor Low Range Protein Ladder.

2.6.2 Categorization of venom proteins

As lowly abundant compounds are enriched by the CPLL pretreatment, we expected that the extended venom protein list would contain many compounds which probably have no function once they are injected into the victim. These so-called venom trace molecules only have a local function in the venom duct or reservoir or are released by leakage of the gland tissue [1]. In contrast, toxins are typically highly abundant and are actively secreted by the venom glands to contribute to the venom defense or social immunity function. As such, we categorized the detected compounds in those two groups (Table 2.1 and Table 2.2) based on their predicted subcellular location and protein function. As the CPLL treatment shifts relative protein abundances, we were unable to use protein abundance as a distinguishing parameter.

The subcellular location of each identified compound was predicted by several parameters. Many compounds (86/102) contain an N-terminal secretion signal peptide, which allows to target proteins to the secretory pathway [endoplasmic reticulum (ER), Golgi complex, lysosomes]. Generally, proteins lacking this signal peptide are not actively secreted and were therefore assigned to the group of venom trace molecules (Table 2.2). The subcellular location

of homologues of well annotated species (fruitfly, mouse and human) and the 'cellular component' GO-terms were chosen as additional parameters (Table S2.3). In contrast to putative toxins which have secreted homologues and assigned GO-terms linked to the extracellular space, venom trace molecules are often found in intracellular compartments of the secretory pathway. Also, multiple compounds contain a C-terminal ER-retention signal (XDEL) which is indicative for ER-retained proteins (Table 2.2).

To derive a function for all identified compounds, we used different levels of information, i.e. known functions of homologues, the 'molecular function' and 'biological process' GO-terms, and the predicted functional domains (Table S2.3). Only compounds with a (putative) function in defense or social immunity were assigned to the list of putative toxins (Table S2.4). In addition, we searched for venom homologues of other species, found by venom gland transcriptome and venom proteome studies (Tables 2.1 and 2.2 and Table S2.5).

This approach allowed us to identify 33 putative toxins and 58 trace molecules, which are listed according to their function (Table 2.1) and subcellular location (Table 2.2), respectively. At present, no function could be attributed to eleven compounds. They were categorized separately (Table 2.3) because they lack functional domains and/or similar annotated sequences.

2.6.2.1 Putative toxins

This study confirmed the presence of multiple toxins found in preceding honeybee venom analyses. Toxic functions have been proposed for most of them: phospholipase A_2-1, melittin, apamin, hyaluronidase, major royal jelly proteins (MRJPs) [44], dipeptidyl peptidase IV [45], Api m 6 [46] and CUB serine protease [47]. For others, such as icarapin and secapin, the function remains elusive, but we hypothesize that they have a toxic function as they have been detected with less sensitive technologies [9;11;18;19] and may therefore be moderately/highly abundant. Additionally, for some well-known but functionally non-characterized honeybee venom compounds we propose a function based on similar venom compounds from other species: first, in snake venom, acid phosphatase has been suggested to play a role in liberating purines (mainly adenosine). It acts as a multitoxin and potentiates venom-induced hypotension and paralysis [48]. Second, the platelet-derived growth factor may act similarly to snake venom VEGF-like molecules, which are the most potent vascular

Table 2.1. All discovered putative toxins are classified according to their function, and GenBank/Augustus accession numbers (Acc N°) and allergen names (Allergen) are shown (derived from http://www.allergen.org/). All of them contain a secretion signal peptide (derived by SignalP). Their finding in preceding honeybee venom (HBV) and venom gland (HBV GL) proteomic studies is marked with "X". Also proteins which exist in the exosome protein database ExoCarta (derived from http://www.exocarta.org/) are indicated with "X" (Exosome). In gray, the results are shown of stand alone blasts against a venom sequence database, which reveals the existence of similar venom proteins and venom gland transcripts (Evidence) of other species (Species). Type of evidence: P= venom protein; T=venom gland transcript; EA= enzymatic activity; U= unknown. Identified putative venom toxins

Name	Acc. N°	Allergen	HBV	HBV GL	Exosome	Species	Evidence	Ref.
						Similar venom compounds		
Esterases								
Phospholipase A$_2$-1	gi|58585172	Api m 1	x	x				
Phospholipase A$_2$-2	gi|110758297							
Group XV phospholipase A$_2$	gi|328791555		x					
Acid phosphatase 1	gi|301601654	Api m 3	x	x	x	*Nasonia vitripennis*	P	[32]
Acid phosphatase 2	gi|328790726				x	*Pteromalus puparum*	T+EA	[53]
Acid phosphatase 3	gi|110768981				x	*Pteromalus puparum*	T+EA	[53]
5'-nucleotidase	gi|66523706				x	*Gloydius blomhoffi*	P	[54]
Carboxylesterase	gi|187281550	Api m 8	x		x	*Nasonia vitripennis*	P	[32]
Proteases and peptidases								
CLIP serine protease	gi|66507455					*Bombus ignitus*	P	[52]
CUB serine protease 1	gi|58585116	Api m 7	x			*Apis cerana*	U	gi|146395065
CUB serine protease 2	gi|48101366					*Nasonia vitripennis*	T	[32]
Putative trypsin	au9.g8903.t1					*Nasonia vitripennis*	P	[32]
Serine protease snake	gi|328783264							
Dipeptidyl peptidase IV	gi|187281543	Api m 5	x		x	*Vespula vulgaris*	P	[45]
Serine carboxypeptidase	gi|226533687	Api m 9	x	x	x	*Crotalus adamanteus*	T	[55]
Prolylcarboxypeptidase	gi|328778095				x			
Metalloprotease	gi|110748908				x	*Eulophus pennicornis*	T	[56]

Table 2.1. Continued

Name	Acc. N°	Allergen	HBV	HBV GL	Exosome	Similar venom compounds		
						Species	Evidence	Ref.
Protease inhibitors								
Api m 6	gi|94400907	Api m 6	x					
Serpin 1	gi|328793022				x	*Crotalus adamanteus*	T	[55]
Serpin 2	gi|328791596				x	*Crotalus adamanteus*	T	[55]
Serpin 3	gi|328780925				x	*Crotalus adamanteus*	T	[55]
Carbohydrate metabolism								
Hyaluronidase	gi|58585182	Api m 2	x	x		*Apis cerana*	T	[57]
N-sulfoglucosamine sulfohydrolase	gi|328793712					*Crotalus adamanteus*	T	[55]
Endochitinase	gi|66511507					*Nasonia vitripennis*	T	[32]
Growth factors								
Platelet-derived growth factor	gi|328789531		x	x	x	*Bitis gabonica gabonica*	P	[58]
Imaginal disc growth factor 4	gi|66514614			x		*Chelonus inanitus*	P	[59]
Major royal jelly proteins								
MRJP8	gi|58585070	Api m 11	x	x		*Chelonus inanitus*	P	[59]
MRJP9	gi|67010041	Api m 11	x			*Chelonus inanitus*	P	[59]
Peptides								
Melittin	gi|58585154	Api m 4	x			*Vespula maculifrons*	T	[60]
Apamin	gi|58585166		x					
Secapin	gi|58585180		x			*Vespa velutina nigrithorax*	U	gi|33321084
Other toxins								
C-type lectin	gi|328792562							
Icarapin	gi|60115688	Api m 10	x	x		*Apis cerana*	T	[61]

permeability factors known and which can facilitate venom spreading [49]. Two other enzymes, carboxylesterase and serine carboxypeptidase could play a role in degradation of insect neurotransmitters [50] and a wide range of proteins [51], respectively.

In addition, our study identified 17 new putative toxins. They belong to the classes of esterases, proteases, protease inhibitors, carbohydrate-degrading enzymes, growth factors and C-type lectins. Their putative functions are presented in Table S2.4. All newly detected toxins may allow spreading of the venom and/or cause tissue damage. Despite the diminished amounts of the melittin peptide in the CPLL elution (Figure 2.1), no new small peptides were discovered. As such, besides melittin (and also MCDP and apidaecin), no additional antimicrobial peptides playing a role in social immunity [2;7] were found.

Interestingly, our study revealed that honeybee venom contains multiple toxins belonging to the same protein class: five S1 serine endopeptidases, three acid phosphatases, three serpins and two group III PLA_2s were identified. Their combination with unique binding domains, as is seen in the group of serine proteases (no domain/CUB/CLIP), may allow a similar catalytic activity, but each directed towards a specific target. This activity may even be directed towards species-specific targets, as similar toxins could have evolved because of the biochemical arms race with specific attacker species belonging to the distinct classes of arthropods (wasp and robber bees stealing honey) and vertebrates (birds, mice). Moreover, the CLIP serine protease showed to be similar to a bumblebee (*B. ignitus*) CLIP serine protease, which was demonstrated to play a distinct role in insects and mammals [52]. Alternatively, similar proteins may perform variable catalytic functions due to sequence differences in the active site, as is seen in the newly detected acid phosphatases (APH2 and APH3) which contain an amino acid substitution in the active site septapeptide compared to APH1 (RHGXKXP → RHGXRXP). Consequently, although multiple proteins belong to the same protein class, they may act upon a wide range of targets and/or exert different functions, which broadens the panel of toxins and allows the honeybees to efficiently defend the hive.

Multiple honeybee venom toxins show resemblance to toxins found in other hymenopteran venoms. As honeybees are closely related to bumblebees and wasps, and as their venoms share the same functions, homologues are likely to share similar functions. In contrast, venom homologues of parasitoid wasps may serve different functions as their venom is not used for defense, but influences the arthropod host's immunity, physiology, mobility, reproductive capacity and behavior to keep them alive and serve as food for their offspring

[32]. Also multiple snake venom analogs (growth factors [49], hyaluronidases [62], C-type lectins [63], serine proteases [64], metalloproteases [65;66], acid phosphatases [48], nucleotidases [54;67] and PLA_2s [68]) were found, supporting evidence for convergent evolution.

2.6.2.2 Venom trace molecules

Preceding honeybee venom proteome analyses presumably found mostly abundant compounds using technologies with a lower sensitivity. Therefore, the C1q-like protein was so far the only honeybee venom trace molecule found in a mass spectrometric study [21]. de Graaf and coworkers already indicated that further digging in the venom proteome would yield additional lowly abundant venom trace molecules [1]. Indeed, more than half of the proteins (58/102) identified in this study appears to belong to this category. Their putative functions are briefly summarized in Table S2.4.

We believe that some of them are actively secreted as they may have a local function in the venom duct or reservoir, playing a role in maturation (e.g. peptidylglycine α-hydroxylating monooxygenase) or stabilization (e.g. heat shock proteins) of the secretes. Others are secreted by diverse tissues for exerting essential functions in the extracellular space (e.g. immunity-related proteins and apolipophorins). However, most of the venom trace molecules play roles in secretory pathway processes, such as protein folding, degradation and post-translational modification, N-glycan maturation and degradation, and sphingolipid metabolism. As several compounds contain an ER-retention signal, their presence in the venom may be explained by their unintentional release due to an inefficient retrieval and retrograde transport within the secretory pathway of the highly active secreting venom gland tissue. Other secretory pathway proteins may remain bound to toxins during their transfer through the secretory pathway and may be unintentionally co-secreted. Finally, the release of large secretory pathway-localized multiprotein complexes (e.g. the BiP complex) may contribute to the high number of detected trace molecules. The identification of few plasma membrane compounds in the venom may be explained by the release of their often large extracellular domains.

Remarkably, the venom contains two mitochondrial compounds and 16 compounds which lack a secretion signal peptide. First, both the phospholipid hydroperoxide glutathione peroxidase (GTPX) and kynurenine-oxoglutarate transaminase 1 (KAT) show the highest

Table 2.2. All discovered venom trace molecules are classified according to their subcellular localization. GenBank/Augustus accession numbers (Acc N°) are shown. The presence of a secretion signal peptide (derived by SignalP), ER-retention signal peptide (ER) and their finding in preceding honeybee venom (HBV) and venom gland (HBV GL) proteomic studies is marked with "X". Also proteins which exist in the exosome protein database ExoCarta (derived from http://www.exocarta.org/) are indicated with "X" (EX). In gray, the results are shown of stand alone blasts against a venom sequence database, which reveals the existence of similar venom proteins and venom gland transcripts (Ev.) of other species (Species). Type of evidence: P= venom protein; T=venom gland transcript; U= unknown.

Name	Acc. N°	SignalP	EX	ER	HBV	HBV GL	Similar venom compound Species	Ev.	Ref.
Secreted proteins									
C1q-like protein	gi|221325614	x	x		x				
Lysozyme c-1	gi|328779578	x	x						
Peptidoglycan-recognition protein SA	gi|254910928	x	x						
Transferrin	gi|58585086	x	x			x			
Modular serine protease	gi|328780689	x					*Nasonia vitripennis*	P	[32]
Cathepsin F	gi|328788558	x					*Crotalus adamanteus*	T	[55]
Cathepsin K	au9.g225.t1	x					*Mesobuthus eupeus*	U	gi|148970410
Peptidylglycine α-hydroxylating monooxygenase	gi|328787622	x	x				*Crotalus adamanteus*	T	[55]
Apolipophorins	gi|328780886	x	x						
	gi|328780884								
Dorsal-ventral patterning protein Sog	gi|328791019	x							
Laminin subunit γ-1	gi|328776171	x	x						
Endoplasmic reticulum									
Glucosidase 2, subunit α	gi|66500170	x	x		x		*Crotalus adamanteus*	T	[55]
Glucosidase 2, subunit β	gi|328789473	x	x	x					
Calreticulin	gi|66545506	x	x	x			*Nasonia vitripennis*	T	[32]
UDP-glucose:glycoprotein glucosyltransferase	gi|328786702	x	x	x					
Protein disulfide-isomerase A3	gi|66546657	x	x	x		x	*Crotalus adamanteus*	T	[55]
Peptidyl-prolyl cis-trans isomerase B	gi|335892796	x	x			x	*Crotalus adamanteus*	T	[55]
Hsc70-3	gi|229892214	x	x	x		x	*Crotalus adamanteus*	T	[55]
Hypoxia up-regulated protein 1 (Hsp70)	gi|328784616	x	x	x			*Crotalus adamanteus*	T	[55]

Table 2.2. Continued

Name	Acc. N°	SignalP	EX	ER	HBV	HBV GL	Species	Ev.	Ref.
							Similar venom compound		
Endoplasmic reticulum continued									
Endoplasmin (Hsp90)	gi|110758921	x	x	x			*Crotalus adamanteus*	T	[55]
dnaJ/Hsp40/ERdj3	gi|328790510	x	x				*Crotalus adamanteus*	T	[55]
Endoplasmic reticulum lectin 1	gi|328787701	x					*Crotalus adamanteus*	T	[55]
Protein disulfide-isomerase	gi|328790461	x		x			*Crotalus adamanteus*	T	[55]
Endoplasmic reticulum resident protein 29	gi|110751310	x		x			*Crotalus adamanteus*	T	[55]
Endoplasmic reticulum resident protein 44	gi|328777360	x	x	x			*Crotalus adamanteus*	T	[55]
Calumenin	gi|66509518	x		x					
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	gi|328784759	x	x						
ERGIC-53	gi|328785297	x	x				*Crotalus adamanteus*	T	[55]
Golgi complex									
α-Mannosidase 2	gi|66514147	x							
Mannosyl-oligosaccharide α-1,2-mannosidase isoform A	gi|328782017		x				*Crotalus adamanteus*	T	[55]
Mannosyl-oligosaccharide α-1,2-mannosidase isoform B	gi|328781530		x				*Crotalus adamanteus*	T	[55]
Glycoprotein 3-α-L-fucosyltransferase A	gi|328785366	x	x						
Metallocarboxypeptidase	gi|328785691	x	x				*Crotalus adamanteus*	T	[55]
Nucleobindin	gi|328789222	x	x				*Crotalus adamanteus*	T	[55]
Calsyntenin-1	gi|328778301	x	x				*Crotalus adamanteus*	T	[55]
Lactosylceramide 4-α-galactosyltransferase	gi|328793424	x							
Lysosomes									
α-L-fucosidase	gi|328793281	x	x				*Crotalus adamanteus*	T	[55]
β-galactosidase	gi|110764149	x	x		x				
Aspartic protease (cathepsin D)	gi|66560290	x	x				*Crotalus adamanteus*	T	[55]
γ-interferon-inducible lysosomal thiol reductase	gi|328785531	x					*Microctonus hyperodae*	T	[69]
Glucosylceramidase	gi|66511554	x	x						
Proactivator polypeptide	gi|328782499	x	x			x	*Crotalus adamanteus*	T	[55]

Table 2.2. Continued

Name	Acc. N°	SignalP	Ex	ER	HBV	HBV GL	Similar venom compound Species	Ev.	Ref.
Plasma membrane proteins									
Renin receptor	gi|328776891	x	x				*Crotalus adamanteus*	T	[55]
V-ATPase subunit S1	gi|328783393	x					*Crotalus adamanteus*	T	[55]
Multiple inositol polyphosphate phosphatase	gi|328778827	x	x				*Nasonia vitripennis*	P	[32]
Mitochondrial proteins									
Phospholipid hydroperoxide glutathione peroxidase	gi|110756698	x	x				*Bufo gargarizans*	U	gi|89515096
Kynurenine-oxoglutarate transaminase	gi|328789112						*Crotalus adamanteus*	T	[55]
Proteins without signal peptide									
Hsp70Ab	gi|229892265		x				*Crotalus adamanteus*	T	[55]
Hsc70-4	gi|229892210		x				*Crotalus adamanteus*	T	[55]
Histone H4	gi|328789054		x				*Pelinobius muticus*	T	[70]
Ras-related protein Rab-1A	gi|328784309		x				*Crotalus adamanteus*	T	[55]
Ras-related protein Rab-11A	gi|328778735		x				*Crotalus adamanteus*	T	[55]
Actin related protein 1	gi|297591985		x				*Crotalus adamanteus*	T	[55]
Tubulin β-1 chain	gi|48095525		x				*Crotalus adamanteus*	T	[55]
Tubulin β-chain-like	gi|110762983		x				*Crotalus adamanteus*	T	[55]
Moesin/ezrin/radixin homolog 1	gi|328784401		x				*Crotalus adamanteus*	T	[55]
Ubiquitin-60S ribosomal protein L40	gi|110756311		x				*Ophiophagus hannah*	T	[71]
Elongation factor 1-alpha	gi|58585198		x				*Crotalus adamanteus*	T	[55]

similarity to the *Apis cerana cerana* [72] and *Aedes aegypti* [73] mitochondrial homologues, respectively. However, other GTPXs and KATs have been found in different subcellular compartments. As such, more research is needed to reveal the exact subcellular location of these honeybee proteins. Second, most of the compounds without a signal peptide are in a way connected to the secretory pathway, which might explain their presence in the venom. In contrast, the cytoskeletal proteins are typical cytoplasmic compounds, suggesting that they are released by a different process. Although apoptosis of cells of the venom apparatus cell lining may contribute to their presence within the venom, we would expect to find a higher number of cytoplasmic proteins and even some nuclear compounds. However, all detected cytoskeletal proteins have been found in exosomes, which are small membrane vesicles of endocytic origin that are secreted in various extracellular fluids, including in the venoms of the snake *Gloydius blomhoffii blomhoffii* [74] and two solitary wasps [75]. As such, we propose that the selective, active process of exosome-mediated secretion [76] is responsible for their release. Moreover, as about 64% (58/91) of the identified proteins can be found in exosomes (Table 2.1 and 2.2), it seems that this represents an important way of secretion in the honeybee venom gland.

Table 2.3. Detected venom proteins with an unknown function. GenBank/Augustus accession numbers (Acc. N°) are shown and the presence of a signal peptide (SignalP) and ER-retention signal (ER) is indicated with 'X'.

Name	Acc. N°	SignalP	ER
Unknown function protein 1	gi|110748765	x	
Unknown function protein 2	gi|328778365	x	
Unknown function protein 3	gi|48132776	x	
Unknown function protein 4	gi|66546405	x	
Unknown function protein 5	gi|328781452	x	
Unknown function protein 6	gi|66518899	x	
Unknown function protein 7	gi|328783315		x
Unknown function protein 8	gi|328778397	x	
Unknown function protein 9	gi|328782411	x	
Unknown function protein 10	gi|110776960	x	
Unknown function protein 11	gi|328780111		

2.6.3 Annotation of venom genes and contribution of improved gene predictions to the identification of new venom proteins

The honeybee genome, published in 2006, was noted to have a bimodal GC content that affected the quality of the assembly in some regions and the annotation had fewer genes in the initial gene set (OGSv1.0) than would have been expected based on other species sequenced since. With the advent of next-generation sequencing technologies, sequencing genomes has changed. Therefore, recently an improved genome assembly and gene annotation set (OGSv3.2) for the honeybee has been generated [35]. We could show that the OGSv3.2 gene set, which contains about 5000 new genes, delivers a significant contribution to our venom proteome research. Searching the venom mass spectra against both the OGSv1.0 and OGSv3.2 gene sets revealed that the improved OGSv3.2 gene set enabled the detection of 21 additional peptides supporting 9 new venom protein identifications. Besides, extra tryptic peptides were discovered for 7 venom proteins as a result of improved gene predictions (Table S2.6).

The reduced sequencing cost of second generation sequencing methods also implies the generation of much more transcript sequences than ever before. These transcript data, while short and difficult to assemble into complete transcripts are very useful as evidence supporting gene model prediction and annotation [35]. Also most honeybee venom genes are fully (76.5%) or partially (19.6%) covered by EST evidence. The combination of EST and proteome data allowed to determine their correct gene sequence and all 102 venom genes were manually annotated on the improved honeybee genome assembly (Table S2.7).

The tertiapin peptide, which has been described to be present in the venom already many years ago [13], was not found in the present proteomic analysis. However, no genomic or transcriptomic evidence for this peptide has been described. We solved this issue as we discovered the tertiapin gene by genome mining. The genome improvement project supplies both a gene prediction (GB40695, NCBI Gene ID: 100576769) and EST evidence (Genbank:HP466647.1). The gene is positioned on chromosome 12, next to the apamin and mast cell degranulating peptide venom genes. The three genes are arranged tandemly which may point to a joint control of transcription [77].

2.6.4 Consequences for honeybee venom allergy

Systemic allergic reactions after a honeybee sting have been reported to occur in 0.8-5% of

the general population and may be life-threatening [1]. Nowadays, already 12 honeybee venom allergens are immunologically characterized (Table 2.1). This mass spectrometric study analyzed venom of *A. mellifera carnica*, a subspecies which is the second most popular among beekeepers. Moreover, venom of adult worker honeybees was collected during the summer, when they actively forage. As the general human population is mainly stung by these foraging individuals, this proteomic study offers an extended list of potential new venom allergens.

2.7 CONCLUSIONS

The combination of improved gene prediction sets, the CPLL approach and FT-ICR MS/MS allowed us to explore the hidden honeybee venome. In total, 102 compounds were detected, which were categorized according to their putative function and/or subcellular localization. The 33 putative toxins belong to the classes of esterases, proteases, protease inhibitors, carbohydrate-degrading enzymes, growth factors, MRJPs and antimicrobial peptides. Their (predicted) biological function provides insights into the venom toxicity. In addition, our highly sensitive approach yielded a long list of lowly abundant venom trace molecules. As preceding studies described eleven honeybee venom compounds which remained undetected in this analysis, the honeybee venome is now largely extended to 113 compounds. Finally, this study offers a long list of potential new venom allergens.

2.8 ACKNOWLEDGEMENTS

2.9 ADDENDUM

Supplementary tables can be found on the included CD-ROM or can be requested by e-mail from matthiasvanvaerenbergh@hotmail.com and Dirk.deGraaf@UGent.be.

Table S2.1: This table presents all significant proteins found by searching the generated MS/MS spectra against the Amel_4.5 protein NCBI Refseq and Augustus 9 databases

(Database search). Protein names (Name), accession numbers (Acc. N°), Mascot scores (Score) and molecular weight (MW) are shown. Also the percentage of the protein sequence covered by assigned peptides (Seq. cov.), the number of assigned peptides (# ass. peptides) and the number of unique assigned peptides (# unique ass. peptides) is given.

Table S2.2: Peptide information for each identified protein, including all identified peptide sequences (pep_seq), peptide variable modifications (pep_mod), the experimental mass of charge observed (m/z) and the charge of the precursor ion (z), the peptide Mascot score (pep_score) and its associated probability value (p-value).

Table S2.3: Table showing the enzyme codes, GO-terms (molecular function, cellular compounds and biological process) and InterProScan functional domain codes (inferred by BLAST2GO) for all identified honeybee venom proteins.

Table S2.4: A brief description of the putative function of all new identified venom compounds is presented in the worksheet 'Functions', while corresponding references are shown in the worksheet 'References'.

Table S2.5: Similar compounds found in venom proteome/venom gland transcriptome studies of other species were searched by stand alone Blast searches with the identified *Apis mellifera* venom queries. The accession number (Acc. N°) and name (Name) of the queries and Blast results are shown. Also, the species in which these similar proteins were discovered, the type of evidence for their identification (P= venom protein; T=venom gland transcript; EA= enzymatic activity; U= unknown) and the references of the published studies are presented. Additionally, the blast parameters are included.

Table S2.6: Searching the venom mass spectra against both the OGSv1.0 and OGSv3.2 gene sets revealed that the improved OGSv3.2 gene set allowed the detection of 21 additional peptides supporting 9 new venom protein identifications (grey). Besides, additional tryptic peptides were discovered for 7 venom proteins as a result of improved gene predictions (green).

Table S2.7 shows all manually annotated honeybee venom genes. Annotation data such as species-specific name, gene ID of the initiating sequence and the gene coordinates (scaffold, start and end position, positive or negative strand) are presented. Genes were annotated using Apollo for the *A. mellifera* assembly Amel_4.5. Also the existence of EST evidence in BeeBase (X= full EST, partial= partial EST, /= no EST) is indicated.

2.10 REFERENCES

[1]   de Graaf DC, 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 Mar 6;72(2):145-54.

[2]   Baracchi D, Francese S, Turillazzi S. Beyond the antipredatory defence: honey bee venom function as a component of social immunity. Toxicon 2011 Nov;58(6-7):550-7.

[3]   Zalat S, Nabil Z, Hussein A, Rahka M. Biochemical and haematological studies of some solitary and social bee venoms. Egyptian Journal of Biology 1999;1:57-71.

[4]   Shkenderov S, Ivanova I, grigorova K. An acid monophosphatase and alpha-glucosidase enzymes newly isolated from bee venom. Toxicon Suppl I 1979;17:1969-70.

[5]   Vick JA, Shipman WH, Brooks R, Jr. Beta adrenergic and anti-arrhythmic effects of cardiopep, a newly isolated substance from whole bee venom. Toxicon 1974 Mar;12(2):139-44.

[6]   Shkenderov S, Koburova K. Adolapin--a newly isolated analgetic and anti-inflammatory polypeptide from bee venom. Toxicon 1982;20(1):317-21.

[7]   Baracchi D, Turillazzi S. Differences in venom and cuticular peptides in individuals of *Apis mellifera* (Hymenoptera: Apidae) determined by MALDI-TOF MS. J Insect Physiol 2010 Apr;56(4):366-75.

[8]   Blank S, Bantleon FI, McIntyre M, Ollert M, Spillner E. The major royal jelly proteins 8 and 9 (Api m 11) are glycosylated components of *Apis mellifera* venom with allergenic potential beyond carbohydrate-based reactivity. Clin Exp Allergy 2012 Jun;42(6):976-85.

[9]   Stöcklin R, Favreau P. Proteomics of venom peptides. In: Ménez A, editor. Perspectives in molecular toxinology.Chichester: John Wiley & Sons Ltd; 2002. p. 107-23.

[10] Winningham KM, Fitch CD, Schmidt M, Hoffman DR. Hymenoptera venom protease allergens. J Allergy Clin Immunol 2004 Oct;114(4):928-33.

[11] Matysiak J, Schmelzer CE, Neubert RH, Kokot ZJ. Characterization of honeybee venom by MALDI-TOF and nanoESI-QqTOF mass spectrometry. J Pharm Biomed Anal 2011 Jan 25;54(2):273-8.

[12] Peck ML, O'Connor R. Procamine and other basic peptides in the venom of the honeybee (*Apis mellifera*). J Agric Food Chem 1974 Jan;22(1):51-3.

[13] Ovchinnikov YA, Miroshnikov AI, Kudelin AB, Kostina MB, Boikov VA, Gotgilf IM. Structure and presynaptic activity of tertiapin-neurotoxin from bee venom *Apis mellifera*. Bioorganicheskaya Khimiya 1980;6(3):359-65.

[14] Shkenderov S. A protease inhibitor in bee venom. Identification, partial purification and some properties. FEBS Lett 1973 Jul 15;33(3):343-7.

[15] Hsiang HK, Elliott WB. Differences in honey bee (*Apis mellifera*) venom obtained by venom sac extraction and electrical milking. Toxicon 1975 Mar;13(2):145-8.

[16] Ivanova I, Shkenderov S. A newly isolated enzyme with lysophospholipase activity from bee venom. Toxicon 1982;20(1):333-5.

[17] Doery HM, Pearson JE. Phospholipase B in snake venoms and bee venom. Biochem J 1964 Sep;92(3):599-602.

[18] Van Vaerenbergh M, Cardoen D, Formesyn E, Brunain M, Van Driessche G, Blank S, et al. Extending the honeybee venome with the antimicrobial peptide apidaecin and a protein resembing wasp antigen 5. Insect Molecular Biology 2013;22(2):199-210.

[19] Peiren N, Vanrobaeys F, de Graaf DC, Devreese B, Van BJ, Jacobs FJ. The protein composition of honeybee venom reconsidered by a proteomic approach. Biochim Biophys Acta 2005 Aug 31;1752(1):1-5.

[20] Hoffman DR. Hymenoptera venom allergens. Clin Rev Allergy Immunol 2006 Apr;30(2):109-28.

[21] de Graaf DC, Brunain M, Scharlaken B, Peiren N, Devreese B, Ebo DG, et al. Two novel proteins expressed by the venom glands of *Apis mellifera* and Nasonia vitripennis share an ancient C1q-like domain. Insect Mol Biol 2010 Feb;19 Suppl 1:1-10.

[22] Lowy PH, Sarmiento L, Mitchell HK. Polypeptides minimine and melittin from bee venom: effects on *Drosophila*. Arch Biochem Biophys 1971 Jul;145(1):338-43.

[23] Weinstock GM, Robinson GE, Gibbs RA, Worley KC, Evans JD, Maleszka R, et al. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 2006 Oct 26;443(7114):931-49.

[24] Hider RC. Honeybee venom: a rich source of pharmacologically active peptides. Endeavour 1988;12(2):60-5.

[25] Calvete JJ, Fasoli E, Sanz L, Boschetti E, Righetti PG. Exploring the venom proteome of the western diamondback rattlesnake, *Crotalus atrox*, via snake venomics and combinatorial peptide ligand library approaches. J Proteome Res 2009 Jun;8(6):3055-67.

[26] Fasoli E, Sanz L, Wagstaff S, Harrison RA, Righetti PG, Calvete JJ. Exploring the venom proteome of the African puff adder, *Bitis arietans*, using a combinatorial peptide ligand library approach at different pHs. J Proteomics 2010 Mar 10;73(5):932-42.

[27] Fasoli E, Farinazzo A, Sun CJ, Kravchuk AV, Guerrier L, Fortis F, et al. Interaction among proteins and peptide libraries in proteome analysis: pH involvement for a larger capture of species. J Proteomics 2010 Feb 10;73(4):733-42.

[28] Schuchardt S, Sickmann A. Protein identification using mass spectrometry: a method overview. EXS 2007;97:141-70.

[29] Baykut G, Jertz R, Witt M. Matrix-assisted laser desorption/ionization fourier transform ion cyclotron resonance mass spectrometry with pulsed in-source collision gas and in-source ion accumulation. Rapid Commun Mass Spectrom 2000;14(14):1238-47.

[30] Quinton L, Le Caer JP, Phan G, Ligny-Lemaire C, Bourdais-Jomaron J, Ducancel F, et al. Characterization of toxins within crude venoms by combined use of Fourier transform mass spectrometry and cloning. Anal Chem 2005 Oct 15;77(20):6630-9.

[31] Quinton L, Le Caer JP, Vinh J, Gilles N, Chamot-Rooke J. Fourier transform mass spectrometry: a powerful tool for toxin analysis. Toxicon 2006 May;47(6):715-26.

[32] de Graaf DC, Aerts M, Brunain M, Desjardins CA, Jacobs FJ, Werren JH, et al. Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies. Insect Mol Biol 2010 Feb;19 Suppl 1:11-26.

[33] Fox JW, Ma L, Nelson K, Sherman NE, Serrano SM. Comparison of indirect and direct approaches using ion-trap and Fourier transform ion cyclotron resonance mass spectrometry for exploring viperid venom proteomes. Toxicon 2006 May;47(6):700-14.

[34] Trachsel C, Siegemund D, Kampfer U, Kopp LS, Buhr C, Grossmann J, et al. Multicomponent venom of the spider Cupiennius salei: a bioanalytical investigation applying different strategies. FEBS Journal 2012 Aug;279(15):2683-94.

[35] Honey bee Genome Sequencing and Analysis Consortium. Finding the missing honey bee genes: lessons learned from a genome upgrade. Unpublished.

[36] Munoz-Torres MC, Reese JT, Childers CP, Bennett AK, Sundaram JP, Childs KL, et al. Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera. Nucleic Acids Res 2011 Jan;39(Database issue):D658-D662.

[37] The Uniprot Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Research 2012 Jan;40(D1):D71-D75.

[38] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007 Nov 1;23(21):2947-8.

[39] Petersen TN, Brunak S, von HG, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 2011;8(10):785-6.

[40] Zdobnov EM, Apweiler R. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics 2001 Sep;17(9):847-8.

[41] Rawlings ND, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 2012 Jan;40(Database issue):D343-D350.

[42] Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005 Sep 15;21(18):3674-6.

[43] Blank S, Seismann H, McIntyre M, Ollert M, Wolf S, Bantleon FI, et al. Vitellogenins Are New High Molecular Weight Components and Allergens (Api m 12 and Ves v 6) of *Apis mellifera* and *Vespula vulgaris* Venom. PLoS ONE 2013 Apr 23;8(4):e62009.

[44] Peiren N, de Graaf DC, Vanrobaeys F, Danneels EL, Devreese B, Van BJ, et al. Proteomic analysis of the honey bee worker venom gland focusing on the mechanisms of protection against tissue damage. Toxicon 2008 Jul;52(1):72-83.

[45] Blank S, Seismann H, Bockisch B, Braren I, Cifuentes L, McIntyre M, et al. Identification, recombinant expression, and characterization of the 100 kDa high molecular weight Hymenoptera venom allergens Api m 5 and Ves v 3. J Immunol 2010 May 1;184(9):5403-13.

[46] Peiren N, de Graaf DC, Evans JD, Jacobs FJ. Genomic and transcriptional analysis of protein heterogeneity of the honeybee venom allergen Api m 6. Insect Mol Biol 2006 Oct;15(5):577-81.

[47] Georgieva D, Greunke K, Betzel C. Three-dimensional model of the honeybee venom allergen Api m 7: structural and functional insights. Mol Biosyst 2010 Jun;6(6):1056-60.

[48] Dhananjaya BL, D'Souza CJM. The Pharmacological Role of Phosphatases (Acid and Alkaline Phosphomonoesterases) in Snake Venoms related to Release of Purines - a Multitoxin. Basic & Clinical Pharmacology & Toxicology 2011 Feb;108(2):79-83.

[49] Francischetti IM, My-Pham V, Harrison J, Garfield MK, Ribeiro JM. *Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. Gene 2004 Aug 4;337:55-69.

[50] Yan S, Cui F, Qiao C. Structure, function and applications of carboxylesterases from insects for insecticide resistance. Protein Pept Lett 2009;16(10):1181-8.

[51] Mittapalli O, Wise IL, Shukle RH. Characterization of a serine carboxypeptidase in the salivary glands and fat body of the orange wheat blossom midge, *Sitodiplosis mosellana* (Diptera: Cecidomyiidae). Insect Biochem Mol Biol 2006 Feb;36(2):154-60.

[52] Choo YM, Lee KS, Yoon HJ, Kim BY, Sohn MR, Roh JY, et al. Dual function of a bee venom serine protease: prophenoloxidase-activating factor in arthropods and fibrin(ogen)olytic enzyme in mammals. PLoS ONE 2010;5(5):e10393.

[53] Zhu JY, Ye GY, Hu C. Molecular cloning and characterization of acid phosphatase in venom of the endoparasitoid wasp *Pteromalus puparum* (Hymenoptera: Pteromalidae). Toxicon 2008 Jun 15;51(8):1391-9.

[54] Ogawa Y, Murayama N, Yanoshita R. Molecular cloning and characterization of ecto-5'-nucleotidase from the venoms of *Gloydius blomhoffi*. Toxicon 2009 Sep 15;54(4):408-12.

[55] Rokyta DR, Lemmon AR, Margres MJ, Aronow K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 2012;13:312.

[56] Price DR, Bell HA, Hinchliffe G, Fitches E, Weaver R, Gatehouse JA. A venom metalloproteinase from the parasitic wasp *Eulophus pennicornis* is toxic towards its host, tomato moth (Lacanobia oleracae). Insect Mol Biol 2009 Apr;18(2):195-202.

[57] Shen LR, Zhang CX, Cheng JA. Cloning and sequencing of gene encoding hyaluronidase from the venom of *Apis cerana cerana*. enotomolia sinica 2001;8:353-60.

[58] Calvete JJ, Marcinkiewicz C, Sanz L. Snake venomics of *Bitis gabonica gabonica*. Protein family composition, subunit organization of venom toxins, and characterization of dimeric disintegrins bitisgabonin-1 and bitisgabonin-2. J Proteome Res 2007 Jan;6(1):326-36.

[59] Vincent B, Kaeslin M, Roth T, Heller M, Poulain J, Cousserans F, et al. The venom composition of the parasitic wasp *Chelonus inanitus* resolved by combined expressed sequence tags analysis and proteomic approach. BMC Genomics 2010;11:693.

[60]   Shi WJ, Zhang SF, Zhang CX, Cheng JA. Cloning and comparative analysis of the venom prepromelittin genes from four wasp species. Yi Chuan Xue Bao 2003 Jun;30(6):555-9.

[61]   Wong KL, Li H, Wong KK, Jiang T, Shaw PC. Location and reduction of icarapin antigenicity by site specific coupling to polyethylene glycol. Protein Pept Lett 2012 Feb;19(2):238-43.

[62]   Fox JW. A brief review of the scientific history of several lesser-known snake venom proteins: L-amino acid oxidases, hyaluronidases and phosphodiesterases. Toxicon 2013 Feb;62:75-82.

[63]   Clemetson KJ. Snaclecs (snake C-type lectins) that inhibit or activate platelets by binding to receptors. Toxicon 2010 Dec 15;56(7):1236-46.

[64]   Serrano SMT. The long road of research on snake venom serine proteinases. Toxicon 2013 Feb;62:19-26.

[65]   Calvete JJ. The continuing saga of snake venom disintegrins. Toxicon 2013 Feb;62:40-9.

[66]   Markland FS, Swenson S. Snake venom metalloproteinases. Toxicon 2013 Feb;62:3-18.

[67]   Dhananjaya BL, D'Souza CJM. The pharmacological role of nucleotidases in snake venoms. Cell Biochemistry and Function 2010 Apr;28(3):171-7.

[68]   Gutierrez JM, Lomonte B. Phospholipases A(2): Unveiling the secrets of a functionally versatile group of snake venom toxins. Toxicon 2013 Feb;62:27-39.

[69]   Crawford AM, Brauning R, Smolenski G, Ferguson C, Barton D, Wheeler TT, et al. The constituents of *Microctonus* sp. parasitoid venoms. Insect Mol Biol 2008 Jun;17(3):313-24.

[70]   Diego-Garcia E, Peigneur S, Waelkens E, Debaveye S, Tytgat J. Venom components from *Citharischius crawshayi* spider (Family Theraphosidae): exploring transcriptome, venomics, and function. Cell Mol Life Sci 2010 Aug;67(16):2799-813.

[71]   Wei SS, Zhang YX, Lee WH, Zhang Y. Molecular cloning and comparison of ubiquitin fusion protein and ribosomal protein L30 from *Ophiophagus hannah*. Zoological Research 2005;26(4):397-403.

[72]   Wang M, Kang M, Guo X, Xu B. Identification and characterization of two phospholipid hydroperoxide glutathione peroxidase genes from *Apis cerana cerana*. Comp Biochem Physiol C Toxicol Pharmacol 2010 Jun;152(1):75-83.

[73]   Han Q, Gao YG, Robinson H, Ding H, Wilson S, Li J. Crystal structures of *Aedes aegypti* kynurenine aminotransferase. FEBS J 2005 May;272(9):2198-206.

[74]   Ogawa Y, Kanai-Azuma M, Akimoto Y, Kawakami H, Yanoshita R. Exosome-like vesicles in *Gloydius blomhoffii blomhoffii* venom. Toxicon 2008 May;51(6):984-93.

[75]   Baek JH, Lee SH. Identification and characterization of venom proteins of two solitary wasps, *Eumenes pomiformis* and *Orancistrocerus drewseni*. Toxicon 2010 Sep 15;56(4):554-62.

[76]   Denzer K, Kleijmeer MJ, Heijnen HF, Stoorvogel W, Geuze HJ. Exosome: from internal vesicle of the multivesicular body to intercellular signaling device. J Cell Sci 2000 Oct;113 Pt 19:3365-74.

[77]   Gmachl M, Kreil G. The Precursors of the Bee Venom Constituents Apamin and Mcd Peptide Are Encoded by 2 Genes in Tandem Which Share the Same 3'-Exon. Journal of Biological Chemistry 1995 May 26;270(21):12704-8.

Chapter 3

Unraveling the venom proteome of the bumblebee (*Bombus terrestris*) by integrating a combinatorial peptide ligand library approach with FT-ICR MS

The work presented in Chapter 3 was adapted from the following manuscripts:

1) *M. Van Vaerenbergh, G. Debyser, G. Smagghe, B. Devreese, D. C. de Graaf. Unraveling the venom proteome of the bumblebee (Bombus terrestris) by integrating a combinatorial peptide ligand library approach with FT-ICR MS. Toxicon, in press. This work will be published as a companion paper of the unpublished manuscript of the International Bumblebee Genomics Consortium, reporting on the sequencing of the genome of two bumblebee species:*

2) *International Bumblebee Genomics Consortium. Two bumblebee genomes show the route to advanced social living. Unpublished work.*

3.1 CONTRIBUTIONS

D. de Graaf and B. Devreese assisted with the study design. M. Van Vaerenbergh executed all experiments. G. Debyser gave technical assistance for the LC-ESI-LTQ-FT-ICR-MS experiments and setting up Mascot searches. L. De Smet assisted during cutting out gel slices from the SDS-PAGE gels. K. Morreel calibrated the FT-ICR and LTQ mass analyzers. M. Van Vaerenbergh performed all data analysis and gene annotation. Eckart Stolle determined the syntenic regions in the honeybee and bumblebee genomes.

M. Van Vaerenbergh wrote the manuscript reporting on the identification of novel bumblebee venom compounds. This manuscript is submitted to Toxicon and will be published as a companion paper of the main genome paper from the International Bumblebee Genomics Consortium, which will publish the data of the sequencing of the genomes of *B. terrestris* and

B. impatiens. Publication of this companion paper is put on hold by Toxicon until the main genome paper will be released. M. Van Vaerenbergh also wrote a section about the annotation of *B. terrestris* venom genes, the search for honeybee venom homologues in the bumblebee genomes and homologous venom genes in the genomes of *B. terrestris* and *B. impatiens*, which will be included in the main genome paper manuscript. All authors of the companion paper will also be included in the author list of the main genome paper, known as the International Bumblebee Genomics Consortium. The co-authors assisted throughout the writing phase of the companion paper and section of the main genome paper.

3.2 ABSTRACT

Within the Apidae, the largest family of bees with over 5600 described species, the honeybee is the sole species with a well studied venom proteome. So far, only little research has focused on bumblebee venom. Recently, the genome sequence of the European large earth bumblebee (*Bombus terrestris*) became available and this allowed the first in-depth proteomic analysis of its venom composition. We identified 57 compounds, with 52 of them never described in bumblebee venom. Remarkably, 72% of the detected compounds were found to have a honeybee venom homologue, which reflects the similar defensive function of both venoms and the high degree of homology between both genomes. However, both venoms contain a selection of species-specific toxins, revealing distinct damaging effects that may have evolved in response to species-specific attackers. Further, this study extends the list of potential venom allergens. The availability of both the honeybee and bumblebee venom proteome may help to develop a strategy that solves the current issue of false double sensitivity in allergy diagnosis, which is caused by cross-reactivity between both venoms. A correct diagnosis is important as it is recommended to perform an immunotherapy with venom of the culprit species.

3.3 GRAPHICAL ABSTRACT
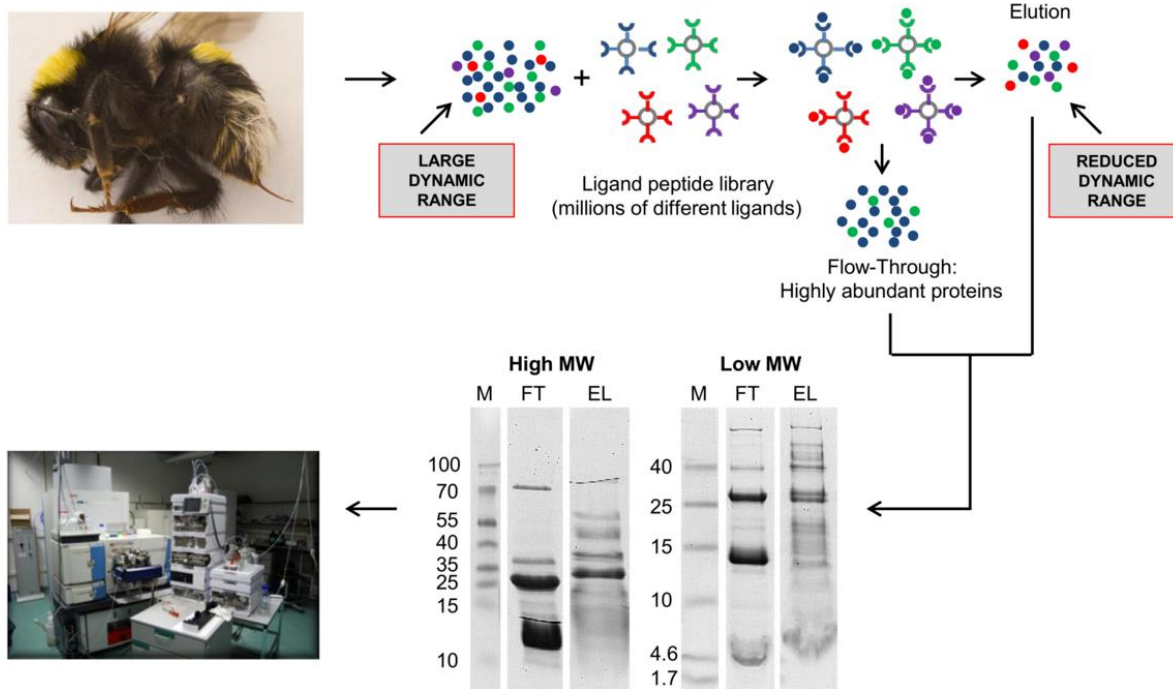
3.4 INTRODUCTION

Several hymenopteran stinging species are known to cause allergic reactions. As bumblebees are not aggressive, the risk of being stung by a bumblebee is very small. However, due to the use of bumblebees as pollinators of greenhouse plants, the prevalence of bumblebee venom allergy increased, especially in greenhouse workers [1-3]. Furthermore, a significant immunological cross-reactivity between bumblebee and honeybee venom is caused by the presence of cross-reactive IgE-antibodies which recognize similar protein and carbohydrate epitopes. Hence, concurrent sensitization can be found in many patients [4].

Knowledge of the venom composition of multiple hymenopteran species may contribute to an improved allergy diagnosis and treatment by immunotherapy. Venom immunotherapy (VIT) is preferably executed using venom of the culprit species, but sometimes the decision which life-saving immunotherapy should be started is difficult to make. Indeed, patients often fail to identify the stinging insect species and the modern whole-venom-based immunodiagnostics not always brings relief due to cross-reactivity or double sensitivity [4]. Component-resolved diagnosis using differential species-specific venom allergens may solve this issue by the detection of species-specific IgE antibodies in patient's

serum [5]. However, well studied venom proteomes and allergen repertoires are a prerequisite for the development of such diagnostic tools.

Within the Apidae, the largest family of bees with over 5600 described species [6], the honeybee is the sole species with a well-studied venom protein composition [7]. So far, only few research studies have focused on bumblebee venom. Using enzymatic activity tests, both highly abundant proteins [phospholipase A_2 (PLA$_2$) and casein hydrolyzing protease] and minor components (hyaluronidase and acid phosphatase) were detected in the venom of the European *large earth* bumblebee, *Bombus terrestris* [2]. The PLA$_2$ and casein hydrolyzing protease are recognized by IgE antibodies and are known as the allergens Bom t 1 and Bom t 4, respectively [2]. Recently, a venom serine protease [8] and a Kunitz-type serine protease inhibitor [9] were identified, which affect the victim's hemostatic system via the venom serine protease inhibitor-mediated antifibrinolytic activity and venom serine protease-mediated fibrin(ogen)olytic activities [9]. In addition, two venom bombolitins have been described, which constitute the highest abundant compounds in bumblebee venom ([10]; GenBank: ADY75782.1). Venom proteomes of few other bumblebee species (*B. ignitus*, *B. pennsylvanicus*, *B. lapidarius, B. hypocrita sapporoensis* and *B. ardens ardens*) have been investigated, identifying PLA$_2$s [2;11], serine proteases [2;12;13], bombolitins [14-16], a mast cell degranulating peptide [17] and a Kunitz-type serine protease inhibitor [18;19].

Most preceding research relied on single compound-oriented, time-consuming and low sensitive techniques such as bio-assays and chemical sequencing via Edman degradation, identifying only a few, primarily highly abundant, compounds. Novel mass spectrometry-based studies often apply bottom-up shotgun approaches. So far, *B. lapidarius* is the only bumblebee species from which MS data on its venom proteins are available. ESI-MS resulted in the detection of 24 compounds and the three major compounds were identified as three bombolitins using a combination of tandem MS with Edman degradation [16]. However, this study was hampered by the lack of a well annotated genome which is a prerequisite for further completing the list of venom proteins [20], as the mass spectra can then be searched against the available protein predictions. We previously reported that the use of high performing mass spectrometry technologies can result in a comprehensive identification of the venom components, as was shown for the parasitoid wasp *Nasonia vitripennis* and the honeybee *Apis mellifera* [7;21]. We have demonstrated that additional lowly abundant

compounds were accessed by reducing the dynamic range of protein concentrations in venom samples by incubation with a combinatorial hexapeptide ligand library (CPLL) [7]. The recent genome publication of the bumblebee *B. terrestris* [22] paved the way for a similar approach for analyzing the venom composition of this species.

The present study explored the worker venom proteome of the large earth bumblebee (*B. terrestris*) by high resolution mass spectrometry analysis (LC-ESI-LTQ-FT-ICR MS/MS). As PAGE separation of *B. terrestris* venom proteins revealed several highly abundant venom compounds [20] which may hamper the detection of lowly abundant compounds in a mass spectrometry experiment, CPLL pretreatment of the venom sample has been conducted. Venom compounds were identified by searching MS/MS spectra against *B. terrestris* genome protein prediction sets [22]. Functions of the identified compounds were predicted using bioinformatics. Finally, all venom genes were annotated on the *B. terrestris* genome assembly [22].

3.5 MATERIALS AND METHODS

3.5.1 Venom collection

Worker bumblebees (*Bombus terrestris*) were collected from three commercially available nests (Biobest Co., Belgium). Pure venom was collected as was previously described for honeybees [23]. Venom of 75 bumblebees was pooled to a protein concentration of 69.54 mg/ml, determined by Bradford protein assay (Thermo Scientific Pierce, Hudson, NH, USA).

3.5.2 Mass spectrometric analysis

Unless otherwise indicated, all experiments were conducted as described in a preceding honeybee venom proteome study [7]. The dynamic range of protein concentrations in the bumblebee venom sample was compressed by a CPLL approach (ProteoMiner protein enrichment small-capacity kit, Bio-Rad Laboratories, Hercules, CA, USA). Approximately 100 µg and 70 µg of the CPLL-flow-through and eluted protein fractions, respectively, were separated on a 10% Tris-glycine-SDS-PAGE gel. In addition, 25 µg of CPLL flow-through proteins and 30 µg of CPLL elution proteins were separated on a 16.5% Tris-tricine-SDS-PAGE gel. After Coomassie staining, proteins were in-gel reduced and alkylated. All flow-through and elution protein bands larger than 40 kDa were cut out of the 10% glycine gel, while those smaller than 40 kDa

were cut out of the 16.5% tricine gel. Also gel parts without any visible protein bands were excised and analyzed. After in-gel tryptic digestion, tryptic peptides were extracted from the gel and analyzed by LC-ESI-LTQ-FT-ICR MS/MS. MS/MS data were searched using the Mascot v2.3 database search engine (Matrix Science, London, UK) against two protein databases of the newly released Bter_1.0 genome [22]: the Augustus5 (https://www.hgsc.bcm.edu/ftp-archive/Bterrestris/Bumblebee_B_ter_GenePredictions/AUGUSTUS/; database contains 30976 sequences and 34153612 residuals) and the Bter_1.0 NCBI Refseq database (ftp://ftp.ncbi.nih.gov/genomes/Bombus_terrestris/protein/; database contains 10577 sequences and 5951597 residuals), added with characterized *B. terrestris* venom sequences: phospholipase A$_2$ (Genbank: P82971.1), serine protease (GenBank: ADY75780.1), bombolitin 1 (GenBank: ADY75781.1) and bombolitin 2 (GenBank: ADY75782.1). To determine false discovery rates (FDR) for the identity threshold, an automatic decoy database search was conducted. Setting the significance threshold at 0.0016 led to a FDR of 0.99% for the Augustus5 search and 0.53% for the Bter_1.0 NCBI Refseq search. Searches were executed with carbamidomethylation of cysteines as a fixed modification and oxidation of methionines as a variable modification. One tryptic miscleavage was permitted and peptide mass tolerance and MS/MS tolerance were set to 10 ppm and 0.3 Da, respectively. Precursor peptide charge state was set to 2+ and 3+.

3.5.3 Criteria for positive identifications

We defined positive protein identifications as queries detected by at least two unique, bold and red (significant and top ranking) peptides from the Mascot output with an ion score ≥30. In addition, the discovery of small peptides was enabled by allowing queries with a sequence coverage higher than 10% due to the detection of only one, bold and red (significant and top ranking) peptide with an ion score ≥30. All protein identifications were merged in one list and all double identifiers were removed.

3.5.4 Sequence analysis and gene annotation

A bioinformatics analysis of identified compounds was executed. As not all sequences are correct in prediction datasets, we tried to determine the correct protein sequence for each identification by searching for *B. terrestris* EST data and homologues of well-annotated species (*Drosophila melanogaster*, *Mus musculus* and *Homo sapiens*). *B. terrestris* EST sequences were

searched by Blasts of each prediction against the Bter_1.0 genome scaffolds on Beebase [24], while homologues were searched against the Uniprot database using the Blast algorithm [25]. A database containing all correct(ed) sequences was constructed and was used for performing bioinformatics analyses: the presence of secretion signal peptides, protein domains, protease and protease inhibitor families and GO-terms, and the existence of similar proteins in exosomes and venoms of other species were determined as previously described [7]. All venom genes were manually annotated on the *B. terrestris* genome assembly using Apollo for the *B. terrestris* assembly Bter_1.0 [22].

3.6 RESULTS AND DISCUSSION

3.6.1 Identification and categorization of venom proteins

This study presents the first in-depth proteomic analysis of the venom of workers of the large earth bumblebee. It revealed 519 unique tryptic peptides (Table S3.1 and S3.2) providing biological evidence for 57 venom proteins and peptides. All venom genes were manually annotated on the *B. terrestris* genome assembly. *B. terrestris* EST data, generated by next-generation sequencing methods by the International Bumblebee Genomics Consortium, support most annotations: 71.9% of the venom proteins have full EST evidence, while 22.8% have partial EST evidence (Table S3.3) [22]. Forty-one of the detected proteins show high sequence similarity to one of the compounds previously identified from honeybee venom (Table 3.1). The 16 other venom compounds seem to be specific to bumblebee (Table 3.2).

3.6.1.1 Bumblebee venom proteins with similarity to honeybee venom proteins

In a preceding study, honeybee venom compounds were categorized in the groups of putative toxins and venom trace molecules by prediction of their biological function and subcellular location [7]. Due to the high sequence similarity between bumblebee and honeybee venom homologues, we hypothesize that they share similar functions and subcellular locations. The identified putative toxins mainly contribute to the defense or social immunity function of the complete venom. They belong to the classes of esterases, proteases, protease inhibitors, carbohydrate-degrading enzymes, growth factors, major royal jelly proteins (MRJP) and antimicrobial peptides. As lowly abundant compounds are enriched by the CPLL pretreatment (Figure 3.1), a subgroup of identified proteins was found which probably have no function

once they are injected into the victim. These venom trace molecules possibly execute a local function in the venom duct or reservoir or are released by leakage of the gland tissue. They were listed according to their subcellular location (Table 3.1).
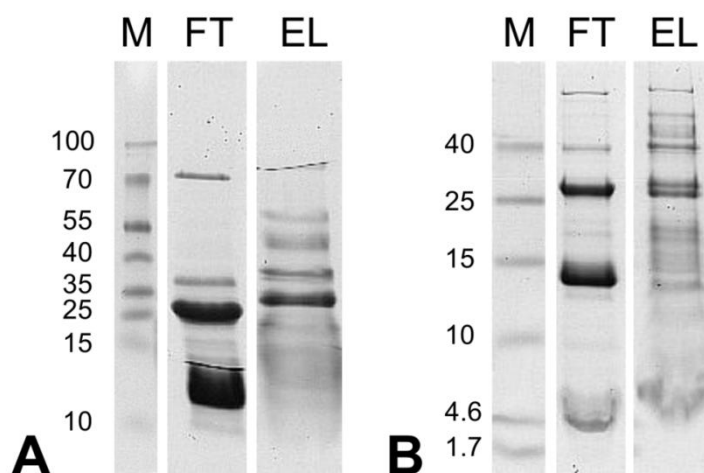
Figure 3.1: Electrophoretic separation of a combinatorial peptide ligand library (CPLL)-treated bumblebee venom sample. CPLL flow-through (=FT) and elution (=EL) samples are separated on a 10% Tris-glycine-SDS-PAGE gel (A) and a 16.5% Tris-tricine-SDS-PAGE gel (B). Molecular weights (kDa) of the marker proteins (M; Thermo Scientific, Rockford, IL, USA) are indicated in the figure: A) PageRuler Prestained Protein Ladder; B) Spectra Multicolor Low Range Protein Ladder.

Five of the identified toxins were already described in preceding *B. terrestris* venom studies. First, bombolitin 1 is a small 18 amino acid peptide with antimicrobial activity [10]. It is a highly abundant venom compound and shows structural and biological properties similar to honeybee venom melittin. In addition, this study presents full sequence data of four bumblebee venom enzymes, which were hitherto only poorly characterized. The first two described components, hyaluronidase and acid phosphatase were only identified by enzymatic activity tests [2], while proteomic identifications and sequence information were lacking. Additionally, we also provide corrected sequences of the *B. terrestris* venom allergens Bom t 1 (PLA$_2$) and Bom t 4 (casein hydrolyzing protease). The PLA$_2$-1 (au5.g6472.t1) derived from the genome sequence shows very high sequence identity (91%) to the previously available Bom t 1 sequence ([2]; UniProt: P82971.1). BeeBase tblastn searches of both sequences against the *B. terrestris* genome returned the same gene as the best Blast hit. However, a higher Blast score was shown with the Augustus5 PLA$_2$-1 prediction and also EST data (GenBank: FN616117.1) support the new sequence. Indeed, the Bom t 1 sequence was

exclusively determined by chemical sequence analysis using Edman degradation of its proteolytic peptide fragments [2], which may be prone to errors. Second, for Bom t 4, only the 20 N-terminal amino acid sequence was available until present ([2]; UniProt: P0CH88.1). We found that this fragment is exclusively present within the serine protease 1 prediction. As such, this FTMS analysis also revealed the complete Bom t 4 sequence.

For some (classes of) venom toxins, differences between honeybee and bumblebee venom were noticed. First, in addition to PLA$_2$-1, a second bumblebee venom PLA$_2$ (PLA$_2$-2) was found for the first time. Both PLA$_2$s share 61% sequence identity and are arranged in tandem in the genome [22]. Therefore, they probably result from a gene duplication event. In contrast, the two group III PLA$_2$s found in honeybee venom [7] share only 45% sequence identity and are positioned on different chromosomes [26]. Moreover, both bumblebee PLA$_2$s appear to show highest sequence similarity to the honeybee PLA$_2$-1 allergen (Table 3.1).

Second, our proteomic analysis revealed the presence of multiple serine proteases. Five of these, serine proteases 1 to 5, show high mutual similarity and all are similar to the honeybee venom CLIP serine protease (Table 3.1). Additionally, a sixth serine protease was described to be present in *B. terrestris* venom [8], but remained undetected in our analysis. This may be caused by venom sample variation (different collection method, spatial and/or seasonal venom variation) or due to its removal during washing procedures of the protein enrichment protocol. The six protease genes appear to be positioned in tandem within the *B. terrestris* genome [22] and may have evolved by gene duplication from a common ancestor. In contrast, the venom CLIP serine protease is the only protease gene present within the syntenic region of the honeybee genome [26]. Remarkably, only two of the bumblebee proteases, serine protease 3 and 6, retained the CLIP domain.

Third, compared to honeybee venom, a lower number of acid phosphatases, CUB serine proteases and MRJPs were found within the bumblebee venom. In the honeybee genome, nine MRJP genes reside in a gene cluster, but only MRJP 8 and 9 are detected in the venom [7;23;27]. The *B. terrestris* genomic information contains only a single-copy MRJP gene, and our study demonstrates that this compound is one of the venom constituents. Due to its expression by the hypopharyngeal glands, it has been suggested to play a role in food digestion or modification [28]. Its presence within honeybee and bumblebee venom may also suggest a toxic function.

Table 3.1. All identified bumblebee venom compounds with sequence similarity to honeybee venom components are shown. Sequences were categorized in a group of putative toxins and venom trace molecules, and further according to their putative function and subcellular location. GenBank/Augustus accession numbers (Acc N°) and allergen names (Allergen, derived from http://www.allergen.org/) are presented and their finding in preceding *Bombus terrestris* venom proteomic studies (BtV) is marked with "X". In gray, the corresponding honeybee venom sequences with the highest sequence similarity (Sim. in %) and their allergen names can be found.

Name	Acc. N°	Allergen	BtV	Sim. (%)	*Apis mellifera* venom protein Name	Acc. N°	Allergen
A. Putative Toxins							
Esterases							
Phospholipase A$_2$-1	au5.g6472.t1	Bom t 1	x	45	Phospholipase A$_2$-1	gi|58585172	Api m 1
Phospholipase A$_2$-2	gi|340723911			48	Phospholipase A$_2$-1	gi|58585172	Api m 1
Acid phosphatase	au5.g1511.t1		x	67	Acid phosphatase 1	gi|301601654	Api m 3
Carboxylesterase	gi|340712251			69	Carboxylesterase	gi|187281550	Api m 8
Proteases and peptidases							
Serine protease 1	gi|340713088	Bom t 4	x	58	CLIP serine protease	gi|66507455	
Serine protease 2	gi|340713090			58	CLIP serine protease	gi|66507455	
(CLIP) serine protease 3	gi|340713092			56	CLIP serine protease	gi|66507455	
Serine protease 4	gi|340713094			53	CLIP serine protease	gi|66507455	
Serine protease 5	gi|340713099			53	CLIP serine protease	gi|66507455	
CUB serine protease	gi|340728251			49	CUB serine protease 1	gi|58585116	Api m 7
Dipeptidyl peptidase IV	gi|340721615			75	Dipeptidyl peptidase IV	gi|187281543	Api m 5
Serine carboxypeptidase	gi|340723441			80	Serine carboxypeptidase	gi|226533687	Api m 9
Prolylcarboxypeptidase	gi|340710015			81	Prolylcarboxypeptidase	gi|328778095	
Protease inhibitors							
Serpin 1	gi|340721561			61	Serpin 1	gi|328793022	
Serpin 2	gi|340708853			65	Serpin 2	gi|328791596	
Serpin 3	gi|340728533			54	Serpin 1	gi|328793022	

Table 3.1. Continued

Name	Acc. N°	Allergen	BtV	*Apis mellifera* venom protein			
				Sim. (%)	Name	Acc. N°	Allergen
Carbohydrate metabolism							
Hyaluronidase	gi|340724556		x	75	Hyaluronidase	gi|58585182	Api m 2
N-sulfoglucosamine sulfohydrolase	gi|340721513			89	N-sulfoglucosamine sulfohydrolase	gi|328793712	
Growth factors							
Platelet-derived growth factor	gi|340710875			47	Platelet-derived growth factor	gi|328789531	
Major royal jelly proteins							
Major royal jelly protein	gi|340716434			55	MRJP9	gi|67010041	Api m 11
Peptides							
Bombolitin 1	gi|325071353		x	39	Melittin	gi|58585154	Api m 4
Other toxins							
Icarapin	gi|340715455			68	Icarapin	gi|60115688	Api m 10
Antigen 5-like protein	gi|340727156			49	Antigen 5-like	gi|328784851	
B. Trace molecules							
Secreted proteins							
Peptidylglycine α-hydroxylating monooxygenase	gi|340712968			85	Peptidylglycine α-hydroxylating monooxygenase	gi|328787622	
Cathepsin F	au5.g2529.t1			80	Cathepsin F	gi|328788558	
Lysozyme c-1	gi|340723421			78	Lysozyme c-1	gi|328779578	
Apolipophorins	gi|340717708			69	Apolipophorins	gi|328780884 gi|328780886	
Vitellogenin	au5.g1998.t1			51	Vitellogenin	gi|58585104	Api m 12
Endoplasmic reticulum							
Peptidyl-prolyl cis-trans isomerase B	gi|340716651			90	Peptidyl-prolyl cis-trans isomerase B	gi|335892796	
Glucosidase 2, subunit α	gi|340709031			89	Glucosidase 2, subunit α	gi|66500170	
Calreticulin	gi|340729835			93	Calreticulin	gi|66545506	

Table 3.1. Continued

Name	Acc. N°	Allergen	BtV	Sim. (%)	*Apis mellifera* venom protein		
					Name	Acc. N°	Allergen
Golgi complex							
Glycoprotein 3-α-L-fucosyltransferase A	gi|340712422			92	Glycoprotein 3-α-L-fucosyltransferase A	gi|328785366	
α-Mannosidase 2	gi|340729800			97	α-Mannosidase 2	gi|66514147	
Lysosomes							
Proactivator polypeptide	gi|340712613			76	Proactivator polypeptide	gi|328782499	
α-L-fucosidase	gi|340716134			78	α-L-fucosidase	gi|328793281	
Glucosylceramidase	gi|340729189			62	Glucosylceramidase	gi|66511554	
Aspartic protease (cathepsin D)	gi|340729556			87	Aspartic protease (cathepsin D)	gi|66560290	
Plasma membrane proteins							
Renin receptor	gi|340718491			87	Renin receptor	gi|328776891	
Multiple inositol polyphosphate phosphatase	gi|340725135			30	Multiple inositol polyphosphate phosphatase	gi|239787860	
Mitochondrial proteins							
Phospholipid hydroperoxide glutathione peroxidase	gi|340714042			77	Phospholipid hydroperoxide glutathione peroxidase	gi|110756698	
Proteins without signal peptide							
Actin	gi|340711865			97	Actin related protein 1	gi|297591985	

Fourth, our analysis revealed a low number of antimicrobial peptides within bumblebee venom. Only bombolitin 1 contains antimicrobial properties [10]. A second *B. terrestris* venom bombolitin is available in GenBank (ADY75782.1) and a higher number of venom bombolitins has been described in other bumblebee species [14;16]. However, in contrast to the *B. impatiens* genome which contains two in tandem positioned bombolitin genes, no additional bombolitin genes could be found within the *B. terrestris* genome [22]. This may be the result of an assembly issue, possibly caused by the existence of multiple closely related genes in tandem positioned in the genome. Furthermore, no homologues of the honeybee venom antimicrobial peptides apidaecin [29] and mast cell degranulating peptide (MCDP) [7;30] were found in the bumblebee venom by mass spectrometric means, although, based on sequence identity and/or presence within syntenic regions, putative homologues genes [apidaecin (GeneID: 100649867); MCDP (GeneID: 100644816 and 100644936)] were found to be present in the genome. The same approach allowed to identify four additional honeybee venom homologous genes in the *B. terrestris* genome, although these were also absent in our bumblebee venom proteome analysis (Table S3.4) [22]. In contrast, apamin and tertiapin, two neurotoxic honeybee venom peptides [29;30], were not found in the determined syntenic region [22]. Therefore, these genes may be absent from the bumblebee genome. Alternatively, as MCDP, apamin and tertiapin are closely related genes which are positioned in tandem within the honeybee genome, the syntenic region in the *B. terrestris* genome assembly may have been misassembled [22].

Finally, several proteins described to be present in honeybee venom were missing in the study which applied an identical technological approach for unraveling the honeybee venom proteome [7]. The current study identified two bumblebee venom homologues of these missing honeybee venom compounds. First, the absence of the antigen 5-like (Ag5-like) protein in the venom of summer honeybees was not surprising as it is probably exclusively expressed by the honeybee venom glands during the winter months [31]. Also, highly abundant paralogues have been identified in wasp and ant venoms [31]. In contrast to bumblebees and wasps, honeybee workers stay alive during the winter, which may result in differences of the venom composition of summer and winter bees. Unfortunately, the function of the venom Ag5s is unknown, which makes this variation in expression difficult to interpret. Second, vitellogenin is described as a high molecular weight honeybee venom allergen [32] and was now also found in bumblebee venom.

3.6.1.2 Bumblebee-specific venom compounds

We also categorized the 16 bumblebee-specific compounds in the groups of putative toxins and venom trace molecules (Table 3.2), which was done by the same approach as described for categorization of the identified honeybee venom compounds [7]. This led to the identification of six putative toxins and eight trace molecules. A brief description of their putative functions can be found in Table S3.3. Results of GO-term searches and of searches for venom homologues from other species can be found in Tables S3.4 and S3.5, respectively.

Both the classical secretory pathway and the exosome-based secretory pathway appeared to be important for protein secretion in the honeybee venom glands [7]. The absence of an N-terminal secretion signal peptide in three of the identified bumblebee venom compounds indicates that their secretion does not involve the classical secretory pathway. While actin and catalase can be secreted by exosomes ([7] and Table 3.2), glucose dehydrogenase (GLD) was not found in the current version of ExoCarta, an exosome content database. Also the *Nasonia vitripennis* GLD venom homologue lacks a signal peptide.

At present, the attribution of a function of two compounds is impossible. Basic sequence analysis of unknown function protein 1 (UFP1) has revealed no predicted functional domains or GO-terms. However, it is related to a fruit fly protein (UniProt: Q7KVT8; 42% sequence identity), which is required for its development beyond first instar. Additionally, it shows low similarity (20% sequence identity) to one of the major proteins of the venom reservoir of the parasitoid wasp *Microctonus hyperodae*, which is suggested to be a tyrosine kinase [33]. No functional domains, GO-terms or venom homologues were found to elucidate the biological function of UFP2.

3.6.2 Comparison of the honeybee and bumblebee venom composition

Honeybees and bumblebees both belong to the Apidae family but have diverged about 100 million years ago [34]. The implementation of an identical proteomic approach [7] allowed to compare the venom composition of both species. These venoms appear to be similar, as 72% of the detected bumblebee venom compounds proved to have a honeybee venom homologue. Also, a similar number of putative toxins was found (29 bumblebee toxins versus 33 honeybee toxins) and most belong to identical functional classes (Table 3.1). Moreover, honeybee venom homologues exist for 70% (23/29) of the detected bumblebee toxins. The presence of toxins with similar activities may be explained by the high degree of homology between both

Table 3.2. The identified bumblebee-specific compounds were categorized in a group of putative toxins and venom trace molecules, and further according to their putative function and subcellular location. The presence of a secretion signal peptide (derived by SignalP) and their presence in exosomes (Ex, derived from http://www.exocarta.org/) are indicated with "X". In gray, the results are shown of stand alone Blasts against a venom sequence database, which reveals the existence of similar venom proteins and venom gland transcripts (Ev.) of other species (Species). Type of evidence: P= venom protein; T=venom gland transcript.

| Name | Acc. N° | SignalP | Ex | Similar venom compound | | |
				Species	Ev.	Ref.
A. Putative toxins						
Chitinase	gi|340721438	x				
Glucose dehydrogenase	gi|340714301			*Nasonia vitripennis*	P	[20]
Plancitoxin	gi|340711594	x		*Crotalus adamanteus*	T	[35]
Serine protease K12H4.7	gi|340715980	x				
Metalloproteinase inhibitor	gi|340725794	x				
C3 and PZP-like α-2-macroglobulin domain-containing protein	gi|340712509	x	x	*Crotalus adamanteus*	T	[35]
B. Trace molecules						
Secreted proteins						
Trehalase	gi|340724978	x	x	*Pimpla hypochondriaca*	T	[36]
Prolyl 3-hydroxylase	gi|340708955	x	x			
Secretory pathway proteins						
Aldose 1-epimerase	gi|340722542	x	x			
Glutaminyl-peptide cyclotransferase	gi|340716459	x	x	*Hottentotta judaicus*	T	[37]
α-N-acetylgalactosaminidase	gi|340722731	x		*Crotalus adamanteus*	T	[35]
β-hexosaminidase subunit β	gi|340718728	x				
β-mannosidase	gi|340713893	x				
Catalase	gi|340714922		x			
Proteins with an unknown function						
Unknown function protein 1	gi|340712339	x	/	*Microctonus hyperodae*	P	[33]
Unknown function protein 2	gi|340721301	x	/			

genomes [34] and the need for similar defensive actions of both venoms. Honeybee and bumblebee stings cause local tissue damage, which induces death in other insects and pain and inflammation in higher organisms [20]. Their venoms contain typical (Hymenoptera) venom constituents, such as phospholipases A_2, proteases, acid phosphatases, hyaluronidases, protease inhibitors, growth factors and MRJPs, which contribute to the venom's toxic actions [7]. Further, these venoms appear to have 18 venom trace molecules in common. Some have a relevant function in the venom duct or reservoir for maturation of secretory proteins (e.g. peptidylglycine α-hydroxylating monooxygenase), while others are secreted by a broad range of tissues as they have an essential function within the extracellular space (e.g. immune system proteins, anti-oxidant enzymes, apolipophorin) [7]. However, most shared trace molecules are typical secretory pathway proteins, which may be unintentionally released due to an inefficient recycling in the venom gland tissue.

Both venoms also contain some species-specific characteristics. The bumblebee venom seems to have a lower complexity due to a smaller number of identified compounds: 57 bumblebee compounds versus 102 honeybee compounds. However, this difference is mainly evoked by a different number of venom trace molecules and proteins with an unknown function. Additionally, as mentioned above, variation between both groups of toxins can be noticed and both venoms contain a selection of species-specific toxins, which may point to distinct damaging effects. Therefore, these species may have undergone evolutionary adaptations in response to species-specific attackers. Indeed, although no data have been reported for *B. terrestris*, intraperitoneal injection of venom of the bumblebee *B. impatiens* (LD50= 7.2 mg/kg) in mice showed that it is about half as lethal to vertebrates as honeybee venom (3.5 mg/kg) [28]. Moreover, honeybees and bumblebees differ in their defensive behavior. Compared to honeybees, bumblebees rarely sting humans, which supports the lower activity of bumblebee venom towards vertebrates. Differences in protein abundance among homologues may further increase functional diversity of both venoms. Unfortunately, this parameter information could not be analyzed as it is influenced by the protein enrichment strategy that removes mainly abundant proteins in a non-linear fashion.

The application of venom on the body surface as a way of protection against pathogens has been suggested for multiple Hymenopteran species [38]. Remarkably, in contrast to honeybee venom, only few bumblebee venom antimicrobial peptides were found. However,

the social structure of honeybee and bumblebee colonies differs. In contrast to bumblebees which build a new colony every year, honeybees inhabit the hive for multiple years and also regulate the hive temperature during winter months. Additionally, bumblebees form colonies which are much less extensive than those of honeybees. Therefore, bumblebees may be less exposed to pathogens, which may explain the reduced need for a large repertoire of antimicrobial peptides. However, as small peptides are difficult to detect using our approach, future studies are needed to confirm this hypothesis.

3.6.3 Consequences for Hymenoptera venom allergy diagnosis

Although several stinging Apidae species can cause allergic reactions, only the venom of the honeybee has been properly characterized. This resulted in the identification of 12 allergens (derived from http://www.allergen.org/). In contrast, besides the allergens from *B. terrestris*, only two allergens of one other bumblebee species, the North-American bumblebee *B. pennsylvanicus*, have been characterized. Although these allergens from both species are similar, they contain different IgE reactive epitopes [2]. Also others reported that differences in IgE-binding between venoms of both species exist [39]. Further research is needed to determine whether this diversity in venom composition is also true for the other 250 bumblebee species [40]. At this moment, our high performance approach to analyze the venom composition is applicable to only one other bumblebee species. Besides *B. terrestris*, *B. impatiens* is the only species with an available genome sequence [22]. Using BLAST searches of each of the 57 identified *B. terrestris* venom proteins against the *B. impatiens* Refseq database, highly similar *B. impatiens* protein sequences were identified (Table S3.8) [22].

The unraveled venom proteomes of the honeybee and large earth bumblebee may contribute to an improved Hymenoptera venom allergy diagnosis and treatment. Hymenopteran venoms contain similar proteins and carbohydrates, which results in cross-reactivity. Several studies reported a high degree of cross-reactivity between honeybee and bumblebee venom [41]. This is supported by this study with the identification of many homologues with often high sequence identity (Table 3.1). Moreover, except for Api m 6, all characterized honeybee venom allergens appear to have a bumblebee venom homologue (Table 3.1), which represent putative bumblebee venom allergens.

Due to the reported high cross-reactivity between both venoms, for many years honeybee venom was used to treat patients sensitized to bumblebee venom [41]. However, treatment

by VIT in some patients failed [41], indicating the existence of species-specific venom epitopes. Indeed, this study shows that many honeybee- and bumblebee-specific venom compounds can be found. Additionally, in some cases sequence identity between homologues is limited, leading to distinct linear and/or conformational epitopes. For example, very little of the protein surface of the PLA_2 allergens from honeybee (Api m 1) and bumblebee (Bom t 1) venom is conserved [2]. Therefore, further research should characterize the relevant allergens. The use of species-specific venom allergens in component-resolved diagnosis may solve the issue of false double sensitivity in diagnostic tests, which is caused by cross-reactivity. A correct diagnosis is important as it is recommended that VIT is performed using venom of the culprit species.

3.7 CONCLUSIONS

This study unraveled the venom proteome of the bumblebee, *Bombus terrestris*, by integrating a combinatorial peptide ligand library approach with FTMS. In total, 57 venom compounds were found, which could be categorized according to their putative functions. In preceding research, honeybee venom has been analyzed by an identical approach. Many honeybee and bumblebee venom homologues were found, which may be explained by their similar defensive function and the high degree of homology between both genomes. Besides, this bumblebee species is increasingly used for pollination in greenhouses. Therefore, greenhouse workers are more exposed to bumblebee stings and often develop venom allergy. This study presents a list of potential new venom allergens. The availability of both the honeybee and bumblebee venom proteome may allow to develop a strategy that solves the current issue of false double sensitivity in allergy diagnosis, which is caused by cross-reactivity between both venoms.

3.8 ACKNOWLEDGEMENTS

3.9 ADDENDUM

Supplementary tables can be found on the included CD-ROM or can be requested by e-mail from matthiasvanvaerenbergh@hotmail.com and Dirk.deGraaf@UGent.be.

Table S3.1: This table presents all significant proteins found by searching the generated MS/MS spectra against the Bter_1.0 protein NCBI Refseq and Augustus5 databases (Database search). Protein names (Name), accession numbers (Acc. N°), Mascot scores (Score) and molecular weight (MW) are shown. Also the percentage of protein sequence covered by assigned peptides (Seq. cov.), the number of assigned peptides (# ass. peptides) and the number of unique assigned peptides (# unique ass. peptides) is given.

Table S3.2: Peptide information for each protein identification, including all identified peptide sequences (pep_seq), peptide variable modifications (pep_mod), the observed experimental mass-to-charge (m/z) and the charge of the precursor ion (z), the peptide mascot score (pep_score) and its associated probability value (p-value).

Table S3.3: This table presents the EST data and genome positions of all 57 annotated genes found in the *B. terrestris* venom proteome analysis. Genes were annotated using Apollo for the *B. terrestris* assembly Bter_1.0. EST evidence in Beebase (X= full EST, partial= partial EST, /= no EST), gene ID of the initiating sequence and genome positions (scaffold, start and end position, positive or negative strand) are shown.

Table S3.4: Although being described in preceding *B. terrestris* or *A. mellifera* venom research, several proteins were not found in the present bumblebee venom proteome analysis. Table S4 presents the list with genome positions (scaffold, start and end position, positive or negative strand) of these annotated (potential) venom genes.

Table S3.5: A brief description of the putative function of the identified bumblebee-specific venom compounds is presented in the worksheet 'Functions', while corresponding references are shown in the worksheet 'References'.

Table S3.6: Table showing the enzyme codes, GO-terms (molecular function, cellular compounds and biological process) and InterProScan functional domain codes (inferred by BLAST2GO) for all identified bumblebee venom proteins.

Table S3.7: Similar compounds found in venom proteome/venom gland transcriptome studies of other species were searched by stand alone Blasts with the identified *Bombus terrestris* venom queries. The accession number (Acc. N°) and name (Name) of the queries and Blast results are shown. Also, the species in which these similar proteins were discovered, the type of evidence for their identification (P= venom protein; T=venom gland transcript; EA= enzymatic activity; U= unknown) and the references of the published studies are presented. Additionally, the Blast parameters are included.

Table S3.8: Result of BLASTS of identified *B. terrestris* venom proteins against the *B. impatiens* Refseq database. Also, *B. impatiens* genome positions are presented for the six venom serine protease genes, showing that these are positioned at the ends of scaffolds.

3.10 REFERENCES

[1] Bucher C, Korner P, Wuthrich B. Allergy to bumblebee venom. Curr Opin Allergy Clin Immunol 2001 Aug;1(4):361-5.

[2] Hoffman DR, El-Choufani SE, Smith MM, de GH. Occupational allergy to bumblebees: allergens of *Bombus terrestris*. J Allergy Clin Immunol 2001 Nov;108(5):855-60.

[3] de Groot H. Allergy to bumblebees. Curr Opin Allergy Clin Immunol 2006 Aug;6(4):294-7.

[4] Bilo BM, Rueff F, Mosbech H, Bonifazi F, Oude-Elberink JN. Diagnosis of Hymenoptera venom allergy. Allergy 2005 Nov;60(11):1339-49.

[5] Muller UR, Johansen N, Petersen AB, Fromberg-Nielsen J, Haeberli G. Hymenoptera venom allergy: analysis of double positivity to honey bee and Vespula venom by estimation of IgE antibodies to species-specific major allergens Api m1 and Ves v5. Allergy 2009 Apr;64(4):543-8.

[6] Cardinal S, Straka J, Danforth BN. Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. Proceedings of the National Academy of Sciences of the United States of America 2010 Sep 14;107(37):16207-11.

[7] Van Vaerenbergh M, Debyser G, Devreese B, de Graaf DC. Exploring the hidden honeybee (*Apis mellifera*) venom proteome by integrating a combinatorial peptide ligand library approach with FTMS. Journal of Proteomics. In press 2013.

[8] Qiu Y, Choo YM, Yoon HJ, Jia JM, Cui Z, Wang D, et al. Fibrin(ogen)olytic activity of bumblebee venom serine protease. Toxicology and Applied Pharmacology 2011 Sep 1;255(2):207-13.

[9] Qiu Y, Lee KS, Choo YM, Kong DX, Yoon HJ, Jin BR. Molecular cloning and antifibrinolytic activity of a serine protease inhibitor from bumblebee (*Bombus terrestris*) venom. Toxicon 2013 Mar 1;63:1-6.

[10] Qiu Y, Choo YM, Yoon HJ, Jin BR. Molecular cloning and antibacterial activity of bombolitin isolated from the venom of a bumblebee, *Bombus terrestris*. Journal of Asia-Pacific Entomology 2012 Mar;15(1):21-5.

[11] Xin Y, Choo YM, Hu Z, Lee KS, Yoon HJ, Cui Z, et al. Molecular cloning and characterization of a venom phospholipase A(2) from the bumblebee Bombus ignitus. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 2009 Oct;154(2):195-202.

[12] Qiu Y, Choo YM, Yoon HJ, Jin BR. Molecular cloning and fibrin(ogen)olytic activity of a bumblebee (*Bombus hypocrita sapporoensis*) venom serine protease. Journal of Asia-Pacific Entomology 2012 Mar;15(1):79-82.

[13] Choo YM, Lee KS, Yoon HJ, Kim BY, Sohn MR, Roh JY, et al. Dual Function of a Bee Venom Serine Protease: Prophenoloxidase-Activating Factor in Arthropods and Fibrin(ogen)olytic Enzyme in Mammals. PLoS ONE 2010 May 3;5(5):e10393.

[14] Argiolas A, Pisano JJ. Bombolitins, A New Class of Mast-Cell Degranulating Peptides from the Venom of the Bumblebee *Megabombus pennsylvanicus*. Journal of Biological Chemistry 1985;260(3):1437-44.

[15] Choo YM, Lee KS, Yoon HJ, Je YH, Lee SW, Sohn HD, et al. Molecular cloning and antimicrobial activity of bombolitin, a component of bumblebee *Bombus ignitus* venom. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 2010 Jul;156(3):168-73.

[16] Favreau P, Menin L, Michalet S, Perret F, Cheyneval O, Stocklin M, et al. Mass spectrometry strategies for venom mapping and peptide sequencing from crude venoms: Case applications with single arthropod specimen. Toxicon 2006 May;47(6):676-87.

[17] Argiolas A, Herring P, Pisano JJ. Amino-Acid-Sequence of Bumblebee Mcd Peptide - A New Mast-Cell Degranulating Peptide from the Venom of the Bumblebee *Megabombus pennsylvanicus*. Peptides 1985;6:431-6.

[18] Choo YM, Lee KS, Yoon HJ, Qiu YL, Wan H, Sohn MR, et al. Antifibrinolytic Role of a Bee Venom Serine Protease Inhibitor That Acts as a Plasmin Inhibitor. PLoS ONE 2012 Feb 16;7(2):e32269.

[19] Choo YM, Yoon HJ, Jin BR. Effects of the bumblebee (*Bombus ignitus*) venom serine protease inhibitor on serine protease and phospholipase A(2) of *B. ignitus* venom. Journal of Asia-Pacific Entomology 2012 Dec;15(4):543-5.

[20] de Graaf DC, Aerts M, Danneels E, Devreese B. Bee, wasp and ant venomics pave the way for a component-resolved diagnosis of sting allergy. Journal of Proteomics 2009 Mar 6;72(2):145-54.

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

[22] International Bumblebee Genomics Consortium. Two bumblebee genomes show the route to advanced social living. Unpublished.

[23] Peiren N, Vanrobaeys F, de Graaf DC, Devreese B, Van Beeumen J, Jacobs FJ. The protein composition of honeybee venom reconsidered by a proteomic approach. Biochimica et Biophysica Acta-Proteins and Proteomics 2005 Aug 31;1752(1):1-5.

[24] Munoz-Torres MC, Reese JT, Childers CP, Bennett AK, Sundaram JP, Childs KL, et al. Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera. Nucleic Acids Research 2011 Jan;39:D658-D662.

[25] The Uniprot Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Research 2012 Jan;40(D1):D71-D75.

[26] Honey bee Genome Sequencing and Analysis Consortium. Finding the missing honey bee genes: lessons learned from a genome upgrade. Unpublished.

[27] Peiren N, de Graaf DC, Vanrobaeys F, Danneels EL, Devreese B, Van Beeurnen J, et al. Proteomic analysis of the honey bee worker venom gland focusing on the mechanisms of protection against tissue damage. Toxicon 2008 Jul;52(1):72-83.

[28] Kupke J, Spaethe J, Mueller MJ, Rossler W, Albert S. Molecular and biochemical characterization of the major royal jelly protein in bumblebees suggest a non-nutritive function. Insect Biochemistry and Molecular Biology 2012 Sep;42(9):647-54.

[29] de Lima PR, Brochetto-Braga MR. Hymenoptera venom review focusing on *Apis mellifera*. Journal of Venomous Animals and Toxins including Tropical Diseases 2003;2(9):149-62.

[30] Kitamura H, Yokoyama M, Akita H, Matsushita K, Kurachi Y, Yamada M. Tertiapin potently and selectively blocks muscarinic K+ channels in rabbit cardiac myocytes. Journal of Pharmacology and Experimental Therapeutics 2000 Apr;293(1):196-205.

[31] Van Vaerenbergh M, Cardoen D, Formesyn EM, Brunain M, Van DG, Blank S, et al. Extending the honey bee venome with the antimicrobial peptide apidaecin and a protein resembling wasp antigen 5. Insect Mol Biol 2013 Apr;22(2):199-210.

[32] Blank S, Seismann H, McIntyre M, Ollert M, Wolf S, Bantleon FI, et al. Vitellogenins Are New High Molecular Weight Components and Allergens (Api m 12 and Ves v 6) of *Apis mellifera* and *Vespula vulgaris* Venom. PLoS ONE 2013 Apr 23;8(4):e62009.

[33] Crawford AM, Brauning R, Smolenski G, Ferguson C, Barton D, Wheeler TT, et al. The constituents of *Microctonus* sp. parasitoid venoms. Insect Molecular Biology 2008 Jun;17(3):313-24.

[34] Stolle E, Wilfert L, Schmid-Hempel R, Schmid-Hempel P, Kube M, Reinhardt R, et al. A second generation genetic map of the bumblebee *Bombus terrestris* (Linnaeus, 1758) reveals slow genome and chromosome evolution in the Apidae. Bmc Genomics 2011 Jan 19;12.

[35] Rokyta DR, Lemmon AR, Margres MJ, Aronow K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). Bmc Genomics 2012 Jul 16;13.

[36] Parkinson NM, Conyers CM, Keen JN, MacNicoll AD, Smith I, Weaver RJ. cDNAs encoding large venom proteins from the parasitoid wasp *Pimpla hypochondriaca* identified by random sequence analysis. Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 2003 Apr;134(4):513-20.

[37] Morgenstern D, Rohde BH, King GF, Tal T, Sher D, Zlotkin E. The tale of a resting gland: Transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. Toxicon 2011 Apr;57(5):695-703.

[38] Baracchi D, Francese S, Turillazzi S. Beyond the antipredatory defence: Honey bee venom function as a component of social immunity. Toxicon 2011 Nov;58(6-7):550-7.

[39] Stapel SO, de Raadt JWL, van Toorenenbergen AW, de Groot H. Allergy to bumblebee venom. II. IgE cross-reactivity between bumblebee and honeybee venom. Allergy 1998 Aug;53(8):769-77.

[40] Williams PH, Cameron SA, Hines HM, Cederberg B, Rasmont P. A simplified subgeneric classification of the bumblebees (genus Bombus). Apidologie 2008 Jan;39(1):46-74.

[41] Cruz S, Vega A, Fernandez S, Marques L, Baltasar M, Alonso A, et al. Report From the Hymenoptera Committee of the Spanish Society of Allergology and Clinical Immunology: Immunotherapy With Bumblebee Venom. Journal of Investigational Allergology and Clinical Immunology 2012;22(5):377-8.

Chapter 4

IgE recognition of novel chimeric isoforms of the honeybee (*Apis mellifera*) venom allergen Api m 10 evaluated by protein array technology

4.1 CONTRIBUTIONS

D. de Graaf, T. Jakob, L. De Smet, D. Ebo, E. Spillner and S. Blank assisted with the study design. For RT-PCRs and cloning, M. Van Vaerenbergh was assisted by master student K. De Crem. M. Van Vaerenbergh did the bacterial expressions, protein purification, protein array development and statistical analysis, and mass spectrometry searches. Concerning these activities, L. De Smet, G. Van Driessche and G. Debyser provided technical advice. During a 6 week internship of M. Van Vaerenbergh at the Institute of Biochemistry and Molecular Biology (Hamburg University, Germany), he was assisted by Y. Michel and S. Blank for variant 2 bacterial expression. S. Blank provided insect cell-produced variant 2. T. Jakob conducted the ImmunoCAPs for pre-screening of Api m 10 allergic patients and provided sera. D. Ebo delivered control sera and sera for protein array optimization. Innobiochips spotted the arrays and optimized the protein array protocol. B. Devreese searched for predicted isoform-specific tryptic peptides detectable by mass spectrometry. M. Van Vaerenbergh wrote the manuscript and was assisted by the co-authors through the writing phase.

4.2 ABSTRACT

The major allergen Api m 10 is an interesting candidate for increasing sensitivity of honeybee venom allergy component-resolved diagnosis. However, preceding studies provided indications for Api m 10 protein heterogeneity and this presently unexplored complexity may have implications for immunodiagnostics. In the present study, reverse transcription PCR revealed the expression of at least nine additional transcript isoforms by the honeybee venom glands. Two distinct mechanisms are responsible for the generation of these isoforms: while the previously known variant 2 is produced by an alternative splicing event, novel identified isoforms are intragenic chimeric transcripts. To the best of our knowledge, this is the first report of the identification of chimeric transcripts generated by the honeybee. Also, by a retrospective proteomic analysis we found some evidence for the presence of several of these isoforms in the venom proteome. Additionally, we explored the effect of Api m 10 protein heterogeneity on IgE reactivity by the colorimetric protein array technology. This revealed that the observed heterogeneity may have important consequences for honeybee venom allergy diagnosis, immunotherapy and allergic responses, as IgE recognition appears to be both isoform- and patient-specific. In addition to variant 2, which was previously demonstrated to be a good biomarker for Api m 10 IgE recognition, two other Api m 10 variants were found to have the potential to increase the sensitivity of component-resolved diagnosis, although this was not the case in our set of analyzed sera.

4.3 INTRODUCTION

Allergic reactions as a consequence of honeybee stings are often observed, especially in beekeepers and their relatives who come close to the hives and are frequently stung [1]. Honeybee venom allergy is mediated by IgE antibodies specific to protein allergens present within the venom. The protein composition of honeybee venom is highly complex, with at least 113 identified proteins and peptides [2]. The complexity is even increased by different glycosylation patterns and protein heterogeneity (phospholipase A_2 [3;4]; Api m 6 [5;6]). Within the order of Hymenoptera, honeybee venom provides the best immunologically characterized model: 12 allergens have been reported (http://www.allergen.org/). One of these allergens is icarapin, designated as Api m 10 in the IUIS nomenclature. Api m 10 is a highly interesting allergen, as it is clinically relevant but underrepresented in therapeutic

extracts [7]. In addition, as two alternatively spliced Api m 10 transcripts have been identified [5], protein heterogeneity may make this picture even more complex. Preceding analyses revealed that both variant 1 [5] and variant 2 [7] are IgE-binding and that their allergenicity is independent of cross-reactive carbohydrate determinants (CCDs). Indeed, both non-glycosylated variant 2 and variant 2 containing glycan structures devoid of the CCD hallmark α-1,3-core fucose residues exhibited IgE reactivity with approximately 50% of honeybee venom-sensitized patients [7]. Also IgE recognition of the non-glycosylated bacterial recombinant variant 1 has previously been reported [5].

Upon 2D-gel separation of pure venom, four protein spots were identified as icarapin [3]. However, the difference in theoretical molecular weight (MW) between the two characterized splice variants is small (0.3 kDa) and cannot account for the observed MW differences between these protein spots. Indeed, two icarapin spots from the higher MW region of the gel differ by about 3 kDa and these may correspond to different glycoforms of variant 1 and/or 2. Curiously, the MW of two other icarapin spots is about 30 kDa lower. This observation strongly suggests that additional Api m 10 isoforms exist, which could have immunological consequences. For other allergens, protein heterogeneity has been reported to be immunologically relevant. For example, the birch genome contains at least 7 pollen-expressed genes that encode distinct Bet v 1 isoforms with varying IgE reactivity [8;9]. Therefore, this study focused on the identification of (potential) additional Api m 10 isoforms and the impact of Api m 10 protein heterogeneity on IgE recognition.

4.4 MATERIALS AND METHODS

4.4.1 Venom gland RNA isolation and cDNA synthesis

Approximately 100 honeybee (*Apis mellifera carnica*) venom glands were dissected under anesthesia by chilling and were transferred to RNAlater (Ambion, Austin, TX, USA). RNA was extracted using the RNeasy Mini Kit (Qiagen, Valencia, CA, USA) following the protocol for purification of total RNA from animal cells using spin technology. Subsequently, cDNA was synthesized using the RevertAid™ H Minus First Strand cDNA synthesis kit (Fermentas, St Leon-Roti, Germany).

4.4.2 Screening of icarapin transcript heterogeneity

Reverse transcriptase-PCRs (RT-PCRs) were carried out as described before [10]. Primers for amplification of the mature (without secretory signal sequence) icarapin sequence were developed, allowing ligation-independent cloning by a 5' incorporated sequence (in italics): forward primer 5'-*GACGACGACAAGATG*TTCCCTGGTGCACACGATG-3' and reverse primer 5'-*GAGGAGAAGCCCGG*TCAAGCAGTTAATACATCTCCT-3'. RT-PCR-amplification products were electrophoretically separated on a 2% agarose gel over a 25 cm distance. The complete length of the gel lane was divided in ten gel pieces and separate DNA extractions were performed by the Genejet gel extraction kit (Fermentas, St Leon-Roti, Germany). Subsequently, RT-PCR fragments were cloned in the pIEx-7 Ek/LIC vector according to the instructions of the Ek/LIC cloning kit (Novagen, Madison, WI, USA). For each cloning reaction, plasmid DNA was extracted [11] from eight different colonies. Agarose gel electrophoresis of restriction digests allowed the selection of different transcripts. Plasmid DNA was purified by the Miniprep protocol (Fermentas, St Leon-Roti, Germany) and DNA sequencing was performed as described previously [10]. Sequences were analyzed by multiple sequence alignment (ClustalW2; [12]). Putative N- or O-linked glycosylation sites were determined with NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) and NetOGlyc 3.1 (http://www.cbs.dtu.dk/services/NetOGlyc/).

4.4.3 Proteomics

Several of our prior studies reported on proteomic analyzes of honeybee venom [2;3;13]. Here, we reanalyzed the generated MS/MS data by searching against the honeybee protein RefSeq database extended with all translated icarapin isoform sequences to identify isoform-specific tryptic peptides. Identical search parameters were used as those previously described [2;3;13].

4.4.4 Synthetic peptide production and recombinant production

Isoforms smaller than 50 amino acids were synthetically produced (Genscript, Piscataway, NJ, USA) at HPLC purities higher than 70%. Mass spectrometry peak analysis was done to analyze peptide quality.

Synthetic production of some of these isoforms was impossible due to technical limitations. These isoforms and isoforms larger than 50 amino acids were produced by

bacterial recombinant expression. Icarapin variant 2 was expressed and purified using the IMPACT protocol (New England Biolabs, Beverly, MA, USA) as previously described [7]. All other icarapin variants and uterocalin were cloned, sequenced and expressed using the pET system (Invitrogen, Carlsbad, CA, USA) [10]. They were overproduced in *E. coli* as His-tagged fusion proteins and purified by immobilized metal chelate affinity chromatography on Ni-NTA beads (adapted protocol from [13]). When native purification yielded insufficient amounts of soluble protein after bacterial pellet sonication, denaturing purifications were applied. Purified proteins were dialyzed to PBS and sample purity was evaluated by Coomassie staining of SDS-PAGE gels. Protein integrity was checked by Western blots visualizing His-tagged recombinants using an anti-His antibody [13]. Protein concentration was determined by a Bradford protein assay (Thermo-Scientific, Rockford, IL, USA). Proteins were diluted to a final concentration of 1 mg/ml, except variant 1 had a final concentration of 0.2 mg/ml as it was insoluble at higher concentrations.

4.4.5 Patients' sera

Sera from honeybee venom allergic patients were collected to test Api m 10 isoform IgE reactivity (Table S4.1). The diagnosis of honeybee venom allergy was based on the history of systemic allergic bee sting, positive skin test and sIgE to honeybee venom (≥ 0.35 kU/L, ImmunoCAP i1, Thermo Fisher Scientific, Uppsala, Sweden), as recently described [14]. Api m 10-specific IgE titers were determined by ImmunoCAP FEIA tests (Thermo Fisher Scientific, Uppsala, Sweden) with rApi m 10 variant 2, recombinantly produced in Sf9 insect cells [7]. Also serum total IgE titers and IgE titers for CCDs, serum tryptase and honeybee venom allergens Api m 1, Api m 4, Api m 5 were defined (Table S4.1). Additionally, negative control sera obtained from wasp-stung individuals without sting reaction and/or with negative honeybee venom IgE results were collected. All patients had given their informed written consent to draw a serum sample, and all experiments applying human sera were approved by the local ethics committee.

4.4.6 Protein array spotting and development

Synthetic peptides were dissolved at 0.1 mM in 0.01M phosphate buffer pH 7.4 with 50mM NaCl and recombinants were dissolved at 0.5 mg/ml in 0.01M phosphate buffer pH 7.4. Peptides and proteins were printed on nitrocellulose-coated glass slides (Sartorius,

Gottingen, Germany) in duplicate, which was conducted by Innobiochips (IBL, Lille, France). Besides the Api m 10 isoforms, an irrelevant control protein (uterocalin or BSA) was spotted. Additionally, two proteins to assess the quality of biochip development and spot normalization were spotted: protein G binds all goat-IgG and Goat Anti-Mouse (GAM) binds all antibodies from mouse.

Arrays were developed in a 16-well hybridization cassette (Arrayit). A protocol for human serum IgE recognition was developed: arrays were saturated for 1 hour at room temperature (RT) with 150 µl of saturation buffer (PBS-BSA 1%-Tween20 0.05%). Incubation with 50 µl of two-fold diluted (with saturation buffer) serum samples was performed overnight at 4 °C. Subsequently, arrays were incubated for 1 hour at RT with 100 µl of a secondary polyclonal HRP-linked goat-IgG anti-human IgE (2 µg/ml in saturation buffer) (Acris Antibodies, Hiddenhausen, Germany). Finally, 50 µl of TMB (3,3',5,5'-tetramethylbenzidine) solution (Calbiochem, La Jolla, CA) was added which generates a stable precipitate at the reaction site. Arrays were washed three times for 5 minutes with washing buffer (PBS-Tween20 0.05%) between all consecutive steps. After development, arrays were air-dried and scanned using the Spotware colorimetric array scanner (obtained from European Biotech Network, Dolembreux, Belgium) at a scan gain of 0.9 and a scan resolution of 5 µm (Spotware 1.1 software). ImageJ was used to adjust image size to a 1000 pixel width. The resulting image is quantified using Mapix (Innopsis, Carbonne, France).

4.4.7 Data analysis

For each spot, the local median background intensity is subtracted from the median spot intensity. Only spots with signal to noise ratios greater than a minimum threshold of two were included in the analysis. Normalization between arrays was done using protein G and the average of spot duplicates was calculated. The average spot intensity per protein on control arrays developed with sera of honeybee non-allergic patients was subtracted from the average spot intensity per protein on the arrays developed with sera from honeybee allergic patients. Next, for each array a cut-off value was calculated as the average of the intensity of the irrelevant control protein plus three times the standard deviation (SD) (Table S4.2). SDs were calculated as follows: $((SD\ of\ the\ control\ protein)^2 + (SD\ of\ the\ control\ protein\ spots\ on\ the\ negative\ control\ arrays)^2)^{1/2}$.

4.5 RESULTS AND DISCUSSION

4.5.1 Icarapin transcript heterogeneity

Mature icarapin transcripts were amplified from venom gland tissue by reverse transcription (RT)-PCR. Analysis of generated amplicons revealed at least ten fragments with distinct nucleotide lengths (Figure 4.1). Sequence analysis of seventeen cloned RT-PCR fragments, identified eleven different variants. Variants are named variant 1 to variant 11. All cloned nucleotide sequences (and their GenBank numbers) and protein sequences are shown in Figure S4.1 and S4.2, respectively.

The icarapin gene is positioned on chromosome LG1 (GeneID: 503505) and consists of 4 exons. Sequence alignment of all variants to the genome indicates the existence of a single genomic locus only (Figure S4.3). Two distinct mechanisms seem to be responsible for the generation of these variants. First, the splice sites of variant 1 conform to the general canonical GT-AG splicing rule [15]. Variant 1 and variant 2 differ by only 12 base pairs due to the presence of an alternative splice acceptor site in exon 3 and also this splicing is
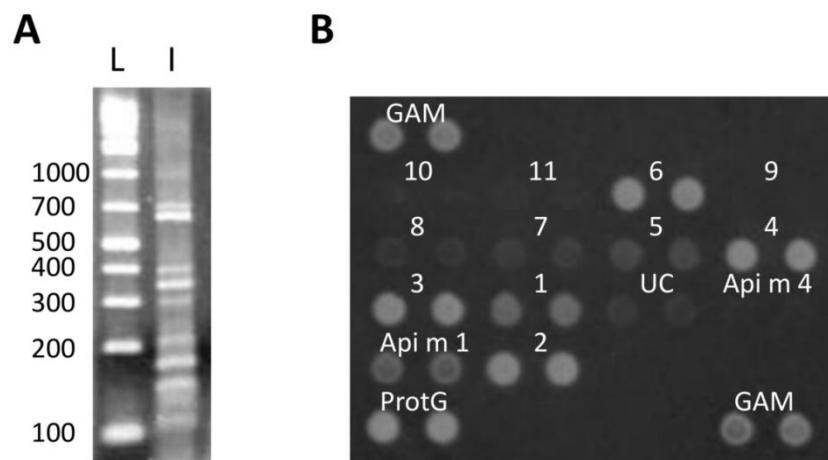
Figure 4.1: A) Reverse transcriptase-PCR on honeybee venom gland tissue reveals multiple Api m 10 isoform transcripts. L= GeneRuler 1 kb DNA ladder (Fermentas); I= icarapin amplicons. B) Api m 10 isoforms were printed as synthetic constructs or recombinants to determine their immunological relevance by protein array technology. A developed array with serum of patient 1 (see Table 4.1) is shown. Numbers represent spotted Api m 10 variants 1 until 11. Also a negative control protein (UC) and two positive control proteins [purified native honeybee venom allergens Api m 1 and Api m 4 (Sigma-Aldrich, St. Louis, MO)] were included. Additionally, protein G (ProtG) and Goat Anti-Mouse (GAM) allow to assess the quality of biochip development and spot normalization.

consistent with the GT-AG rule. In contrast, the nine newly detected variants are so-called chimeric transcripts, which are distinct from conventionally spliced mRNA isoforms as they are produced by joining exons from two or more different gene transcripts. Instead of splice sites, short homologous sequences (SHSs) are found at the junction sites of the source sequence. The icarapin gene contains multiple distinct SHSs in exon 2, which are joined to SHSs from exon 3 or exon 4 (Figure 4.2 and Figure S4.3). SHSs are 5-9 base pairs in length. In four variants, this process generates frame shifts and premature stop codons, leading to C-terminally truncated variants. While mature variant 1 consists of 204 amino acids, variants 8 to 11 are severely truncated with lengths of 41, 25, 19 and 12 amino acids. Protein sequence alignments of all variants are shown in Figure 4.3.
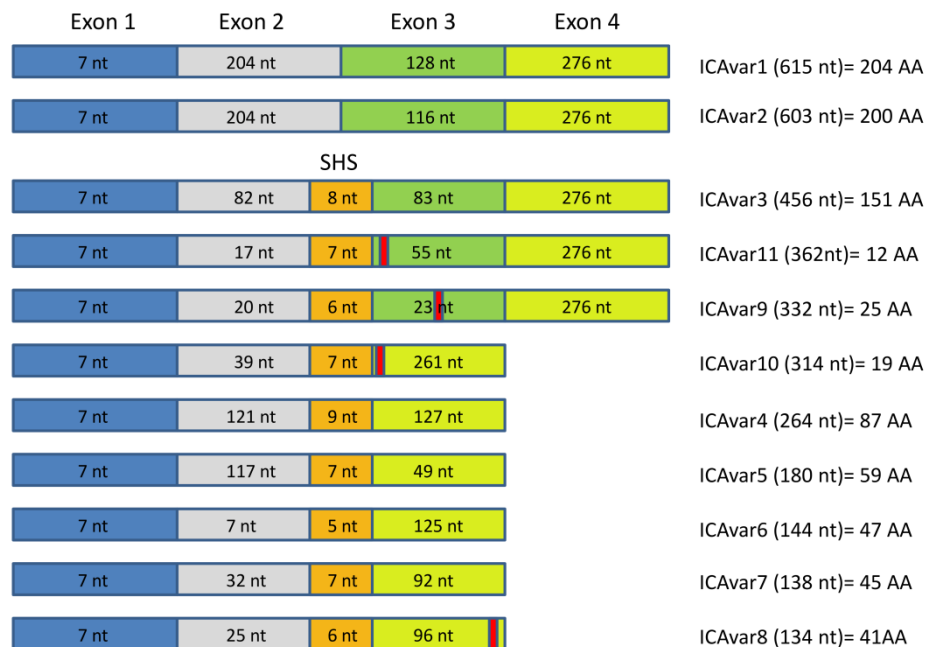
Figure 4.2: Schematic figure showing the exon structure of sequenced icarapin amplicons (named icarapin variant 1 to 11). Different exons are shown as colored boxes. Distinct short homologous sequences (SHSs) are found at the junction sites of the chimeric transcripts (orange boxes). Red boxes present alternative stop codons, which generate C-terminally truncated variants. The full coding sequence lengths are shown: number of nucleotides (nt), number of amino acids (AA).

To the best of our knowledge, we here report the first identification of chimeric transcripts generated by the honeybee. Although thousands of chimeric transcripts have been reported in fruit fly, mouse and human [16], evidence at the protein level is still scarce.

In humans, the few characterized chimeric proteins are associated with cancer, although some also appear to be expressed at low levels under normal physiological circumstances [17;18]. Their physiological role remains elusive but they certainly boost the complexity of the proteome [18]. As until now neither function nor enzymatic activity of icarapin could be determined, the impact of this heterogeneity on its function/activity is impossible to establish. However, as the sequence length of some isoforms is severely reduced, we hypothesize that some of them are loss-of-function mutants.

4.5.2 Proteomics

The specific and unambiguous identification of the icarapin protein isoforms in honeybee venom requires the detection of the isoform-specific junction peptides by bottom-up proteomics [18]. Searching honeybee venom MS/MS data from preceding research [2;3;13] against a database containing all translated icarapin isoform sequences identified multiple tryptic peptides (Figure 4.3). Unfortunately, the differentiation can only be based on a limited number of peptides. For instance, there is only a single difference between the theoretical set of tryptic peptides of variants 1 and 2. Moreover this peptide is larger than 50 amino acids, which typically gives poor signal intensity in MALDI peptide mass fingerprinting. Consequently, no isoform-specific peptides were found.

However, we found some evidence for the presence of a number of isoform proteins within the venom. Indeed, in a preceding study [3], venom proteins were separated using 2D-gel electrophoresis and icarapin tryptic peptides were identified in four spots. Two tryptic peptides (KNVDTVLVLPSIER and VREQMAGILSR) were detected in two high molecular weight spots (spot 8 and 9), which we suggest to be different glycosylation forms of variant 1 and/or variant 2. In contrast, only one icarapin tryptic peptide was found in two low molecular weight spots (spots 19 and 20). These spots are located at a relative low molecular weight (± 8 kDa) and therefore cannot represent full-size variant 1 and/or variant 2. This suggests that they correspond to some of the newly identified icarapin chimers, although protein degradation of the larger variants cannot be excluded. The VREQMAGILSR peptide, a variant 1 and variant 2 specific peptide (Figure 4.3), was not detected in these spots. The only identified tryptic peptide (KNVDTVLVLPSIER) is part of variant 1 and 2 and three of the newly identified chimers: variants 3 to 5. The theoretical pI/MW of variant 5 correlates best to the

observed pI/MW of both spots.

Proteomic evidence for the smallest isoforms (isoform 9-11) can hardly be obtained using our approach, as detectable tryptic peptides are not generated. Moreover, due to the poor resolution in classical SDS-PAGE separation they are not visualized on a 2D-gel. Besides, due to the process of nonsense-mediated mRNA decay, these smallest isoforms may not be expressed [20].

```
A   ICA_var1   FPGAHDEDSKEERKNVDTVLVLPSIERDQMMAATFDFPSLSFEDSDEGSNWNWNTLLRPN 60
    ICA_var2   FPGAHDEDSKEERKNVDTVLVLPSIERDQMMAATFDFPSLSFEDSDEGSNWNWNTLLRPN 60
    ICA_var3   FPGAHDEDSKEERKNVDTVLVLPSIERDQMMA--------------------------- 32
    ICA_var6   FPGAHD----------------------------------------------------- 6
    ICA_var7   FPGAHDEDSKEER---------------------------------------------- 13
    ICA_var4   FPGAHDEDSKEERKNVDTVLVLPSIERDQMMAATFDFPSLSFED--------------- 44
    ICA_var5   FPGAHDEDSKEERKNVDTVLVLPSIERDQMMAATFDFPSLSFED--------------- 44
               ******

    ICA_var1   FLDGWYQTLQSAISAHMKKVREQMAGILSRIPEQGVVNWNKIPEGANTTSTTKIIDGHVV 120
    ICA_var2   FLDGWYQTLQT----HMKKVREQMAGILSRIPEQGVVNWNKIPEGANTTSTTKIIDGHVV 116
    ICA_var3   -----------------------GILSRIPEQGVVNWNKIPEGANTTSTTKIIDGHVV 67
    ICA_var6   ----------------------------------------------------------
    ICA_var7   ----------------------------------------------------------
    ICA_var4   ----------------------------------------------------------
    ICA_var5   ----------------------------------------------------------

                                                                    ↓
    ICA_var1   TINETTYTDGSDDYSTLIRVRVIDVRPQNETILTTVSSEADSDVTTLPTLIGKNETSTQS 180
    ICA_var2   TINETTYTDGSDDYSTLIRVRVIDVRPQNETILTTVSSEADSDVTTLPTLIGKNETSTQS 176
    ICA_var3   TINETTYTDGSDDYSTLIRVRVIDVRPQNETILTTVSSEADSDVTTLPTLIGKNETSTQS 127
    ICA_var6   --------------------------------------------VTTLPTLIGKNETSTQS 23
    ICA_var7   ----------------------------------------------------KNETSTQS 21
    ICA_var4   ------------------------------------------SDVTTLPTLIGKNETSTQS 63
    ICA_var5   ----------------------------------------------------------

    ICA_var1   SRSVESVEDFDNEIPKNQGDVLTA 204
    ICA_var2   SRSVESVEDFDNEIPKNQGDVLTA 200
    ICA_var3   SRSVESVEDFDNEIPKNQGDVLTA 151
    ICA_var6   SRSVESVEDFDNEIPKNQGDVLTA 47
    ICA_var7   SRSVESVEDFDNEIPKNQGDVLTA 45
    ICA_var4   SRSVESVEDFDNEIPKNQGDVLTA 87
    ICA_var5   ---------FDNEIPKNQGDVLTA 59
               ***************

B   ICA_var8    FPGAHDEDSKEERMRPAPNLQGVWKASRISTTRYRRTKEMY 41
    ICA_var11   FPGAHDEDSK-----------VL----------------- 12
    ICA_var9    FPGAHDEDSKE----------------RTLPLPPRSSMDTW 25
    ICA_var10   FPGAHDEDSKEER--------------------KNVDTW 19
                **********
```

Figure 4.3: Protein sequence alignments of icarapin isoforms generated without (A) and with (B) frame shifts. (A) Yellow boxes show all identified tryptic peptides by searching honeybee venom MS/MS data from preceding research [2;3;13] against the RefSeq honeybee protein database extended with all translated icarapin isoform sequences. The experimentally determined antigenic site serine of the mouse IgE epitope from the venom icarapin of the Asian honeybee (*A. cerana*) [19] is indicated by an arrow. (B) Isoform-specific sequences are indicated in gray.

4.5.3 Testing Api m 10 isoform IgE reactivity

We additionally explored the effect of Api m 10 protein heterogeneity on IgE reactivity by protein array technology. A selection of sera of honeybee venom allergic patients was pre-screened with ImmunoCAP FEIA for the presence of IgE reactivity to variant 2 recombinantly produced using the Sf9 insect cell line. As this cell line produces glycans which are not recognized by IgEs, so lacking α-1,3-core fucoses (CCDs) [21], this allowed to assess if these sera possess IgEs binding to the protein backbone of variant 2. For spotting on the array, all 11 variants were produced as synthetic peptides or bacterial recombinant proteins (Figure S4.4) which lack glycosylations. Therefore, serum IgEs specific for CCDs are unable to bind the spotted variants, which otherwise would disturb our analysis focusing on IgE recognition of distinct Api m 10 isoform protein backbones.

As unambiguous protein evidence for a number of Api m 10 isoforms was not obtained, the complete panel of isoforms was evaluated. Variants were printed on nitrocellulose-coated glass slides (Figure 4.1). In contrast to other immunoassays, the applied protein array technology is a multiplex assay, which allows testing a broad protein panel with minute amounts of serum [22]. However, although colorimetric arrays have to be fully validated, they are believed to be equivalent to other immunoassays [23]. Indeed, it was shown that they can reliably detect allergen-specific IgE below 0.35 kU/l, the current WHO standard cut-off for sensitization [22]. Other advantages are its sensitivity, reproducibility, rapidity and simplicity. In addition, in contrast to fluorescent detection, the colorimetric detection allows to perform experiments on cost-effective instrumentation [23].

In the first array experiment, IgE reactivity of the Api m 10 isoforms was analyzed using 18 sera of variant 2-sensitized patients (Table S4.1). ImmunoCAP IgE reactivity of variant 2 was confirmed using the protein array technology in sixteen out of eighteen patients (Table 4.1). In contrast to the array-spotted non-glycosylated variant 2, Api m 10 IgE titers were determined by ImmunoCAP tests using glycosylated, CCD-free variant 2. A preceding study reported a similar IgE reactivity of non-glycosylated and glycosylated (without CCD) variant 2 [7]. However, the present study shows that two patients with rather low ImmunoCAP titers for glycosylated (without CCDs) variant 2 lacked IgE recognition of the non-glycosylated protein. This may indicate that the absence of multiple carbohydrates (variant 2 contains 2 N-linked glycosylation sites) may have altered the protein conformation and specific

Table 4.1. This table presents the result of the first protein array experiment, conducted with variant 2-sensitized sera (ImmunoCAP Api m 10 IgE titers > 0.35 kU/l).

For each serum sample, IgE recognition of the spotted recombinant and synthetic Api m 10 variants is shown: x=recognition; 0=no recognition.

	ImmunoCAP		Results protein array experiment										
Patients	Api m10		Recombinants						Synthetic peptides				
N°	kU/l	Class	ICA_VAR1	ICA_VAR2	ICA_VAR3	ICA_VAR4	ICA_VAR5	ICA_VAR7	ICA_VAR6	ICA_VAR8	ICA_VAR9	ICA_VAR10	ICA_VAR11
1	14.90	3	x	x	x	x	x	0	x	0	0	0	0
2	11.20	3	x	x	x	x	0	x	x	0	0	0	0
3	10.60	3	x	x	x	x	x	0	x	0	0	0	0
4	8.88	3	x	x	x	x	0	0	x	0	0	0	0
5	8.31	3	x	x	x	x	x	x	x	0	0	0	0
6	7.17	3	x	x	x	x	x	0	x	x	0	0	0
7	4.41	3	x	x	x	x	x	x	x	0	0	0	0
8	3.39	2	0	x	x	x	x	0	0	0	0	0	0
9	3.05	2	x	x	x	x	0	0	x	0	0	0	0
10	3.03	2	x	x	x	x	x	0	x	x	0	0	0
11	2.84	2	0	x	x	x	0	x	x	0	0	0	0
12	1.79	2	0	x	x	x	x	x	0	0	0	0	0
13	1.40	2	0	x	x	x	0	0	0	0	0	0	0
14	0.94	2	x	x	x	x	0	0	x	x	0	0	0
15	0.84	2	x	x	x	0	0	0	0	0	0	0	0
16	0.60	1	x	0	0	0	0	0	0	0	0	0	0
17	0.59	1	x	0	0	0	0	x	0	0	0	0	0
18	0.58	1	x	x	x	0	0	0	0	x	0	0	0
N° of positive sera			14	16	16	14	8	6	11	4	0	0	0

conformational epitopes. Alternatively, as both sera have rather low Api m 10-specific IgE titers in ImmunoCAP, the sensitivity of our array system may be lower than that of the ImmunoCAP system.

IgE recognition was found to be both isoform- and patient-specific (Table 4.1). In general, three groups of isoforms can be distinguished. Variants 1 to 4 are IgE recognized by most of the sera, variants 5 to 8 are recognized by less than half of the sera, and variants 9 to 11 are recognized by none of the sera. Unfortunately, unambiguous proteomic evidence for the presence of the novel identified variants in the venom is lacking. Therefore, we give two explanations for the observed differences in IgE reactivity between isoforms.

First, in case only the alternative splice variants 1 and 2 are present in the venom, the IgE recognition of several other variants may be explained by their conserved N- and C-terminal regions (Figure 4.3). These regions may contain linear or even conformational IgE epitopes identical to those of variant 2. Consequently, in addition to variant 2, other variants will be IgE recognized by the selection of sera showing IgE reactivity to variant 2. One of these conserved epitopes may correspond to the mouse IgE epitope of the venom icarapin of the Asian honeybee (*A. cerana*) [19], which is highly similar to variant 1. The experimentally determined antigenic site serine is present at the C-terminal region, and is also present in the *A. mellifera* icarapin variants 1 to 4 and 6 and 7 (Figure 4.3), which explains their IgE reactivity in the array experiment. The observed IgE recognition of variant 5, which lacks this antigenic site serine, may indicate that additional epitopes exist, which are conserved between variant 2 and 5. The smallest variants, variant 9, 10 and 11 (resp. 25, 19 and 12 AA), have only limited sequence identity with variant 2 due to frame shifts. This may explain why they are not recognized.

Second, besides variants 1 and 2, additional variants may be present in the venom. Variant 8 is recognized by 4 patients (Table 4.1). The N-terminal region of variant 8 and the unrecognized variant 10 are identical (Figure 4.3). Therefore, this experiment suggests that variant 8 contains a unique IgE epitope within the variant-specific region. This unique IgE epitope supports its presence in the venom. The lack of IgE recognition of the small variants 9 to 11 by all sera indicates that they have a low antigenicity or that these small peptides are not expressed. Variants 1 to 7 may have common IgE epitopes as their N- and C-terminal regions are conserved (Figure 4.3). In addition, as each variant lacks a different internal region, these variants have a different conformation and possibly variant-specific

Table 4.2. This table presents partial results of the second protein array experiment, conducted with variant 2-non-sensitized sera. Only the results are shown of arrays analyzed with sera containing ImmunoCAP Api m 10 IgE titers between 0.1 and 0.35 kU/l. Besides, 14 sera with ImmunoCAP Api m 10 IgE titers below 0.1 kU/l recognized none of the variants (results not shown). For each serum sample, IgE recognition of the spotted recombinant and synthetic Api m 10 variants is shown: x=recognition; 0=no recognition.

	ImmunoCAP		Results protein array experiment										
Patients	Api m10		Recombinants						Synthetic peptides				
N°	kU/l	Class	ICA_VAR1	ICA_VAR2	ICA_VAR3	ICA_VAR4	ICA_VAR5	ICA_VAR7	ICA_VAR6	ICA_VAR8	ICA_VAR9	ICA_VAR10	ICA_VAR11
1	0.34	0	0	0	x	x	0	0	0	0	0	0	0
2	0.29	0	0	0	0	0	0	0	0	0	0	0	0
3	0.27	0	0	0	0	0	0	0	0	0	0	0	0
4	0.27	0	0	0	0	x	0	0	0	0	0	0	0
5	0.26	0	0	x	0	0	0	0	0	0	0	0	0
6	0.26	0	0	0	0	0	0	0	0	0	0	0	0
7	0.24	0	0	0	0	0	0	0	0	0	0	0	0
8	0.23	0	0	0	0	0	0	0	0	0	0	0	0
9	0.23	0	0	0	0	0	0	0	0	0	0	0	0
10	0.22	0	0	0	0	x	0	0	0	0	0	0	0
11	0.20	0	0	0	0	0	0	0	0	0	0	0	0
12	0.18	0	0	0	0	0	0	0	0	0	0	0	0
13	0.18	0	0	0	0	x	0	0	0	0	0	0	0
14	0.17	0	0	0	0	0	0	0	0	0	0	0	0
15	0.12	0	0	0	0	0	0	0	0	0	0	0	0
N° of positive sera			0	1	1	4	0	0	0	0	0	0	0

conformational IgE epitopes. Analyzing the IgE reactivity of these variants with sera of honeybee venom allergic patients without variant 2-specific IgE should reveal if they possess variant-specific IgE epitopes, which also confirms that they are present in the venom. Therefore, a second array experiment was conducted.

In the first array experiment, analyses were executed with sera having variant 2-specific IgE levels higher than 0.35 kU/l, which suggests sensitization to variant 2. However, the ImmunoCAP system allows linear detection down to 0.1 kU/l, although the cut-off for sensitization is maintained at 0.35 kU/l. In the second array experiment, IgE reactivity of the Api m 10 variants was analyzed using 15 sera with variant 2-specific IgE titers between 0.10 and 0.35 kU/l, and 14 sera with variant 2-specific IgE titers lower than 0.10 kU/l (Table S4.1). This experiment shows that variant 3 and 4 are IgE recognized by respectively 1 and 4 of the variant 2-non sensitized sera. Non-specific IgE binding due to elevated total IgE concentrations has not influenced the obtained result (see Table S4.1 for total IgE titers). These sera belong to the group with variant 2-specific IgE titers between 0.10 and 0.35 kU/l (Table 4.2), while the sera with IgE titers lower than 0.10 kU/l recognized any variants. This observation indicates that a selection of patients possesses variant 3- and 4-specific IgEs directed towards unique epitopes not present in variant 2 or other variants. Therefore, variant 3 and 4 should be present in the venom.

4.5.4 Clinical and diagnostic consequences

In honeybee venom, allergen heterogeneity has only been described for Api m 6, for which four allelic protein isoforms have been described differing in their primary structure at the amino and carboxy terminus by a maximum of six amino acids. However, immunoblot analyzes revealed no isoform-specific IgE [6]. In contrast, our study shows that Api m 10 variants are recognized differently by patient IgE, an observation which has been described for several major allergens derived from different sources [24]. However, the biological mechanism at the level of effector cell activation triggered by interactions between individual IgEs and different isoallergens remains mostly unresolved [25].

Nowadays, rApi m 1 is the only honeybee venom allergen commercially available for component-resolved diagnosis (CRD) of honeybee venom allergy [26]. However, due to the

moderate rApi m 1 sensitization within the population of honeybee venom allergic patients, additional venom allergens should be used to increase sensitivity [27]. Api m 10 is an interesting candidate as it exhibits IgE reactivity with about half of the allergic patients [7]. This study shows that the observed Api m 10 protein heterogeneity may have important consequences for diagnostic tests as IgE recognition is both isoform- and patient-specific. Variant 2 was previously demonstrated to be a good biomarker for Api m 10 IgE recognition [7], which was confirmed by the present study. In addition, our study found that variant 3 and especially variant 4 may be of particular relevance for the diagnosis of honeybee venom allergy in those patients that are allergic to honeybee venom but who do not react to variant 2. However, since all of the analyzed variant 2 non-reactive sera displaying IgE reactivity to variant 3 and variant 4 also showed IgE reactivity to Api m 1 (Table S4.1), these Api m 10 variants did not increase sensitivity of CRD.

4.6 CONCLUSIONS

Already more than 100 compounds have previously been found within honeybee venom [2]. The successful identification of nine new icarapin chimeric transcripts produced by the honeybee venom glands and the indirect evidence for their presence in the venom proteome, shows that the venom complexity has not yet been fully unraveled. As neither the function nor the activity of this compound is known, the impact of this heterogeneity on its function remains elusive. However, this study demonstrates that Api m 10 protein heterogeneity may have important consequences for honeybee venom allergy diagnosis, immunotherapy and allergic responses, as it was shown that IgE recognition is both isoform- and patient-specific. In addition to variant 2, two variants were found to have the potential to increase the sensitivity of component-resolved diagnosis, although this was not the case in our set of analyzed sera.

4.7 ACKNOWLEDGEMENTS

Belgium), Dr. Gonzalez Van Driessche (L-PROBE, Ghent University, Belgium) and Kim De Crem for technical assistance.

4.8 ADDENDUM

Supplementary figures and tables can be found on the included CD-ROM or can be requested by e-mail from matthiasvanvaerenbergh@hotmail.com and Dirk.deGraaf@UGent.be.

Figure S4.1 presents the nucleotide sequences and the GenBank accession numbers of sequenced mature icarapin clones from honeybee venom glands. Alternative stop codons producing C-terminally truncated icarapin peptides are indicated in red.

Figure S4.2 presents all translated protein sequences of sequenced mature icarapin clones from honeybee venom glands.

Figure S4.3 shows the alignment of detected icarapin variants against the icarapin gene structure (Gene ID: 503505) on chromosome 1. Exons are shown in green, while introns are not colored. Intron 5'GT donor and 3'AG acceptor splice sites, conform to the general canonical GT-AG splicing rule, are presented in purple. For the chimeric transcripts, yellow boxes show short homologous sequences found at the junction sites of the source sequence.

Figure S4.4 shows the isoform sequences produced as synthetic constructs and as prokaryotic recombinants. His-tags are shown in red.

Table S4.1 includes the serum IgE titers determined by ImmunoCAP. Spreadsheets 'First experiment' and 'Second experiment' present the IgE titers of the sera used in the first and second array experiment respectively. Total IgE titers and specific IgE titers for honeybee venom, Api m 1, Api m 4, Api m 5, Api m 10, cross-reactive carbohydrates (CCDs) and serum tryptase were determined by ImmunoCAP.

Table S4.2: Spot intensity values of the developed protein arrays per experiment (see spreadsheets 'First experiment' and 'Second experiment'). For each spot, the local background intensity was subtracted from spot intensity. Spots with signal to noise ratios lower than a minimum threshold of 2 are indicated as '/'. Spot intensities of duplicates were averaged, normalized, and average spot intensities of control arrays developed with sera of non-honeybee venom allergic patients were subtracted per variant. For each array a cut-off value was calculated. Positive IgE reactivity is shown as bold intensity values.

4.9 REFERENCES

[1]   Muller UR. Bee venom allergy in beekeepers and their family members. Current Opinion in Allergy and Clinical Immunology 2005 Aug;5(4):343-7.

[2]   Van Vaerenbergh M, Debyser G, Devreese B, de Graaf DC. Exploring the hidden honeybee (*Apis mellifera*) venom proteome by integrating a combinatorial peptide ligand library approach with FTMS. Journal of Proteomics. In press 2013.

[3]   Peiren N, Vanrobaeys F, de Graaf DC, Devreese B, Van BJ, Jacobs FJ. The protein composition of honeybee venom reconsidered by a proteomic approach. Biochim Biophys Acta 2005 Aug 31;1752(1):1-5.

[4]   Blank S, Michel Y, Seismann H, Plum M, Greunke K, Grunwald T, et al. Evaluation of Different Glycoforms of Honeybee Venom Major Allergen Phospholipase A2 (Api m 1) Produced in Insect Cells. Protein and Peptide Letters 2011 Apr;18(4):415-22.

[5]   Peiren N, de Graaf DC, Brunain M, Bridts CH, Ebo DG, Stevens WJ, et al. Molecular cloning and expression of icarapin, a novel IgE-binding bee venom protein. FEBS Lett 2006 Sep 4;580(20):4895-9.

[6]   Kettner A, Hughes GJ, Frutiger S, Astori M, Roggero M, Spertini F, et al. Api m 6: A new bee venom allergen. Journal of Allergy and Clinical Immunology 2001 May;107(5):914-20.

[7]   Blank S, Seismann H, Michel Y, McIntyre M, Cifuentes L, Braren I, et al. Api m 10, a genuine A. mellifera venom allergen, is clinically relevant but underrepresented in therapeutic extracts. Allergy 2011 Oct;66(10):1322-9.

[8]   Schenk MF, Gilissen LJWJ, Esselink GD, Smulders MJM. Seven different genes encode a diverse mixture of isoforms of Bet v I, the major birch pollen allergen. Bmc Genomics 2006 Jul 4;7.

[9]   Wagner S, Radauer C, Bublin M, Mann-Sommergruber KH, Kopp T, Greisenegger EK, et al. Naturally occurring hypoallergenic Bet v 1 isoforms fail to induce IgE responses in individuals with birch pollen allergy. Journal of Allergy and Clinical Immunology 2008 Jan;121(1):246-52.

[10] de Graaf DC, Brunain M, Scharlaken B, Peiren N, Devreese B, Ebo DG, et al. Two novel proteins expressed by the venom glands of *Apis mellifera* and *Nasonia vitripennis* share an ancient C1q-like domain. Insect Mol Biol 2010 Feb;19 Suppl 1:1-10.

[11] Birnboim HC, Doly J. Rapid Alkaline Extraction Procedure for Screening Recombinant Plasmid Dna. Nucleic Acids Research 1979;7(6):1513-23.

[12] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007 Nov 1;23(21):2947-8.

[13] Van Vaerenbergh M, Cardoen D, Formesyn EM, Brunain M, Van DG, Blank S, et al. Extending the honey bee venome with the antimicrobial peptide apidaecin and a protein resembling wasp antigen 5. Insect Mol Biol 2013 Apr;22(2):199-210.

[14] Hofmann SC, Pfender N, Weckesser S, Huss-Marp J, Jakob T. Added value of IgE detection to rApi m 1 and rVes v 5 in patients with Hymenoptera venom allergy. J Allergy Clin Immunol 2011 Jan;127(1):265-7.

[15] Xiong F, Gao JJ, Li J, Liu Y, Feng GY, Fang WL, et al. Noncanonical and canonical splice sites: a novel mutation at the rare noncanonical splice-donor cut site (IVS4+1A > G) of SEDL causes variable splicing isoforms in X-linked spondyloepiphyseal dysplasia tarda. European Journal of Human Genetics 2009 Apr;17(4):510-6.

[16] Li X, Zhao L, Jiang HF, Wang W. Short Homologous Sequences Are Strongly Associated with the Generation of Chimeric RNAs in Eukaryotes. Journal of Molecular Evolution 2009 Jan;68(1):56-65.

[17] Frenkel-Morgenstern M, Lacroix V, Ezkurdia I, Levin Y, Gabashvili A, Prilusky J, et al. Chimeras taking shape: Potential functions of proteins encoded by chimeric RNA transcripts. Genome Research 2012 Jul;22(7):1231-42.

[18] Casado-Vela J, Lacal JC, Elortza F. Protein chimerism: Novel source of protein diversity in humans adds complexity to bottom-up proteomics. Proteomics 2013 Jan;13(1):5-11.

[19] Wong KL, Li H, Wong KKK, Jiang T, Shaw PC. Location and Reduction of Icarapin Antigenicity by Site Specific Coupling to Polyethylene Glycol. Protein and Peptide Letters 2012 Feb;19(2):238-43.

[20] Conti E, Izaurralde E. Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. Current Opinion in Cell Biology 2005 Jun;17(3):316-25.

[21] Seismann H, Blank S, Braren I, Greunke K, Cifuentes L, Grunwald T, et al. Dissecting cross-reactivity in hymenoptera venom allergy by circumvention of alpha-1,3-core fucosylation. Molecular Immunology 2010 Jan;47(4):799-808.

[22] Lebrun SJ, Petchpud WN, Hui A, McLaughlin CS. Development of a sensitive, colorometric microarray assay for allergen-responsive human IgE. Journal of Immunological Methods 2005 May;300(1-2):24-31.

[23] Mascini M, Guilbault GG, Lebrun SJ, Compagnone D. Colorimetric microarray detection system for ghrelin using aptamer-technology. Analytical Letters 2007;40(7):1386-99.

[24] Wurtzen PA. Grass allergen-specific T-cells of atopic patients. Apmis 2001 Sep;109(9):561-71.

[25] Christensen LH, Riise E, Bang L, Zhang CQ, Lund K. Isoallergen Variations Contribute to the Overall Complexity of Effector Cell Degranulation: Effect Mediated through Differentiated IgE Affinity. Journal of Immunology 2010 May 1;184(9):4966-72.

[26] Sturm GJ, Bilo MB, Bonadonna P, Hemmer W, Caruso B, Bokanovic D, et al. Ves v 5 can establish the diagnosis in patients without detectable specific IgE to wasp venom and a possible north-south difference in Api m 1 sensitization in Europe. Journal of Allergy and Clinical Immunology 2012 Sep;130(3):817.

[27] Hofmann SC, Pfender N, Weckesser S, Blank S, Huss-Marp J, Spillner E, et al. Detection of IgE to recombinant Api m 1 and rVes v 5 is valuable but not sufficient to distinguish bee from wasp venom allergy Reply. Journal of Allergy and Clinical Immunology 2011 Jul;128(1):248.

Chapter 5

C1q-like protein and PVF1 from honeybee venom show IgE reactivity but do not activate basophils

5.1 CONTRIBUTIONS

T. Jakob, E. Spillner, S. Blank and D. de Graaf assisted with the study design. Several people contributed to the technical work of this manuscript. S. Blank developed primers for cloning of C1q and PVF1. M. Van Vaerenbergh performed the cloning, analyzed the *Pvf1* transcript heterogeneity and conducted mass spectrometry searches. During a 6 week internship of M. Van Vaerenbergh at the Institute of Biochemistry and Molecular Biology (Hamburg University, Germany), he learned to perform insect cell expressions and protein purification with the help of F. I. Bantleon and S. Blank. During this internship, C1q and PVF1 were produced and purified. Upon his return, M. Van Vaerenbergh introduced the acquired techniques in the Laboratory of Zoophysiology and also produced C1q and PVF1. L. De Smet, assisted with the protein purification. Purified proteins produced in Hamburg and Ghent were used for performing ELISAs (technical work and data analysis executed by S. Blank) and basophil activation tests (technical work and data analysis executed by T. Jakob). T. Jacob conducted the ImmunoCAPs for pre-screening of honeybee venom allergic patients and provided the sera. M. Van Vaerenbergh wrote the article, except for sections 5.3.6, 5.3.7 and 5.4.3 which were written by T. Jakob. Figure 5.4 and Figure 5.5 were kindly provided by S. Blank and T. Jakob, respectively.

5.2 ABSTRACT

The honeybee venom composition and allergenic properties of individual compounds have been extensively investigated, which resulted in the identification of 12 allergens. However, additional allergens may exist as several venom compounds reported in literature remain immunologically uncharacterized. A preceding study showed that C1q lacked IgE reactivity in a preliminary experiment, while the allergenic properties of PVF1 have never been studied. The present study revealed that at least three PVF1 alternative splice variants are produced by the honeybee venom glands. Both C1q and the largest PVF1 variant were produced in a baculovirus-based insect cell expression system, which allows the production of glycoproteins without cross-reactive carbohydrate determinants interfering with the identification of proteinous IgE epitopes. In a population of 72 honeybee venom allergic patients, about 1/3 showed IgE reactivity to C1q and 1/4 to PVF1. In addition, we could demonstrate that a panel of honeybee venom-specific allergens in combination with C1q and PVF1 allows to increase the current sensitivity of the Api m 1-based component-resolved diagnosis of honeybee venom allergy to more than 95%. Remarkably, both compounds lack the potential to activate basophils, which requires further investigation.

5.3 INTRODUCTION

In Hymenoptera venom allergic patients, systemic reactions to stings have been recognized as a potentially fatal condition mediated by IgE antibodies. A detailed characterization of the venom and assessment of allergic potential of venom compounds is a prerequisite for understanding the molecular mechanisms of Hymenoptera venom allergy and improving efficacy of diagnosis and venom immunotherapy (VIT). Many venom components may contribute to the allergic sensitization, allergic symptoms and success of VIT [1]. In addition, component resolved diagnosis (CRD), using an adequate panel of species-specific allergenic compounds, can lead to an increased sensitivity and a better discrimination between different allergies than diagnostic tests using complete extracts [2;3]. Within the Hymenoptera, honeybee (*Apis mellifera*) venom provides the best immunologically characterized model. So far, 12 venom allergens have been identified (http://www.allergen.org/Allergen.aspx). This list consists of mainly highly and moderately abundant venom compounds. Indeed, most of the recently characterized allergens were

detected in the venom proteome by applying mass spectrometry of protein bands/spots visible on SDS-PAGE-or 2D-gels [1;4-7], an approach which lacks sensitivity to detect (very) lowly abundant proteins [8]. However, a recent in-depth venom proteome study showed that honeybee venom has a higher complexity than previously thought, as it successfully identified 83 novel venom compounds by enriching the (very) lowly abundant protein fraction [8]. Therefore, many more allergen candidates remain to be immunologically characterized.

One of the lowly abundant honeybee venom proteins is a C1q-like protein (C1q), which was found in two studies exploring the venom proteome [8;9]. A preliminary test using prokaryotically produced recombinant C1q failed to demonstrate IgE recognition by 5 sera from patients with a documented severe honeybee or wasp venom allergy [9]. However, as the folding of the bacterial recombinant may differ from the natural counterpart and as only very few sera were tested, conclusions from this study have to be taken with care and further research should determine the allergenic nature of this compound.

Additionally, several proteins known to be present in honeybee venom for several years have never been immunologically characterized. One of these compounds, PVF1, is a protein containing a platelet-derived growth factor domain, which was initially identified in honeybee venom in 2005 [6]. Its presence in honeybee venom was recently confirmed [8], while it was also identified within the venom gland tissue proteome [10]. The PVF1 spot density from 2D-gel separated venom [6] indicates that it is a venom compound of moderate abundance. However, recent studies showed that several moderately abundant honeybee venom proteins are clinically relevant allergens [1;4;5]. Therefore, the immunological characterization of PVF1 should reveal if it represents an at present unknown honeybee venom allergen.

In this study, the allergenicity of C1q and PVF1 has been addressed. Recombinant production of both proteins was performed by baculovirus-mediated infection of Sf9 insect cells. This expression system is the preferred choice to produce honeybee venom compounds for immunological characterization, as recombinants will carry the natural insect-specific post-translational modifications (PTM), such as glycosylations and disulfide bridges. PTMs can

influence the protein conformation and therefore also the formation or accessibility of peptide epitopes, which is crucial when analyzing IgE reactivity of individual compounds. In the preceding preliminary analysis which indicated that C1q is not IgE recognized [9], the C1q recombinant was produced in a prokaryotic expression system, which lacks the capacity to add the natural insect-specific PTMs. Therefore, in the present study we decided to produce C1q in the Sf9 insect cell line. Moreover, Sf9 cells produce authentic glycosylations while circumventing α-1,3-core fucose addition, which is the hallmark for cross-reactive carbohydrate determinants (CCDs) [11]. Therefore this system allows to produce recombinants with a conformation closely resembling the natural counterpart, but without α-1,3-core fucose residues which interfere with the identification of proteinous epitopes [11]. Sensitization to both compounds in a large population of honeybee venom allergic patients was analyzed via allergen-specific IgE measurement. In addition, the quantification of basophil activation evoked by these compounds is determined by basophil activation tests.

5.4 MATERIALS AND METHODS

5.4.1 Screening of PVF1 transcript heterogeneity

RNA isolation from honeybee venom gland tissue, cDNA synthesis and RT-PCR were executed as described before [9]. DNA elongation in PCR was adapted to 1 min at 72 °C. Primers were developed for amplification of the mature (without secretory signal sequence) PVF1 sequence (GenBank: XM_392204.4). Additionally, they allow ligation-independent cloning by a 5' incorporated sequence (in italics): forward primer 5'-*GACGACGACAAGATG*CAACTCGAGGATACCAGATAC-3' and reverse primer 5'-*GAGGAGAAGCCCGG*TTATTCTGGATCTGGTTTAGGT-3'. Subcloning for sequencing was done in the pIEx-7 Ek/LIC vector according to the instructions of the Ek/LIC cloning kit (Novagen, Madison, WI, USA).Plasmid DNA was extracted [12] from 24 different colonies and agarose gel electrophoresis of restriction digests allowed the selection of different transcripts. Plasmid DNA was purified by the Miniprep protocol (Fermentas, St Leon-Roti, Germany) and DNA sequencing was performed as described previously [9]. Sequences were analyzed by multiple sequence alignment (ClustalW2 [13]).

5.4.2 Proteomic evidence

PVF1 was successfully identified in the honeybee venom proteome by a profound preceding mass spectrometric analysis [8]. Searching the generated MS/MS data against the honeybee RefSeq database extended with all translated PVF1 variant sequences was performed to identify isoform-specific tryptic peptides. Identical search parameters were used as those previously described [8].

5.4.3 Cloning and expression in Sf9 insect cells

The subcloned mature C1q (GenBank: NM_001144839.1; cloning described previously [9]) and PVF1 cDNA was used for secondary amplification of the coding region using Platinum Taq polymerase (Invitrogen, Carlsbad, CA, USA). The primers 5'-GATCTCTAGAGGGATCGAGGGAAGGGCTATACCGGATCCACCAAATTC-3' and 5'-GATCGCGGCCGCTTATATTTTAGCAATTCTGTATCCAGAG-3' were used for C1q amplification. The PCR product was subcloned via XbaI and NotI into the digested baculovirus transfer vector pAcGP67B (BD Pharmingen, Heidelberg, Germany), which was modified by addition of an N-terminal 10-fold His-tag, V5 epitope as well as a XbaI restriction site [4]. A C-terminal V5 epitope and a 10-fold His-tag was fused to the PVF1 sequence by two consecutive PCR reactions using an identical 5'-GATCGGATCCCAACTCGAGGATACCAGATACC-3' forward primer, but distinct reverse primers: 5'-GTAGAATCGAGACCGAGGAGAGGGTTAGGGATAGGCTTACCTTCTGGATCTGGTTTAGGTTTTCTTC-3' and 5'-GATCCTGCAGTCAATGGTGATGGTGATGGTGATGGTGATGATGACCGGTACGCGTAGAATCGAGACCGAGGAG-3' for the first and second PCR respectively. The final PCR product was subcloned via BamHI and PstI into the non-modified pAcGP67B vector. Expression of C1q and PVF1 in baculovirus-infected insect cells was executed as previously described [4].

5.4.4 Protein purification

The supernatant of baculovirus-infected cells was collected, adjusted to pH 8, centrifuged at 4000 x g for 5 min and applied to a nickel-sepharose matrix (1 ml HisTrap FF column, GE Healthcare, Freiburg, Germany). The column was washed with phosphate buffered saline (PBS) pH 8.0 (100 mM NaCl, 40 mM Na_2HPO_4, 10 mM NaH_2PO_4 $4H_2O$) and the recombinant protein was eluted from the matrix using PBS pH 8.0 containing 300 mM imidazole. Proteins

were dialyzed to PBS pH 8.0 and purification was confirmed by Coomassie Brilliant Blue R-250 staining of an SDS-PAGE separated protein sample. Western blotting and anti-His staining of His-tagged recombinant proteins was performed as described previously [8].

5.4.5 Immunoreactivity of patient sera with recombinant proteins

For assessment of specific IgE immunoreactivity of human sera in ELISA, 384 well microtiter plates (Greiner, Frickenhausen, Germany) were coated with purified recombinant proteins (20 µg/mL) at 4°C overnight and blocked with 40 mg/mL milk powder in PBS. Thereafter, human sera was diluted 1:2 with PBS and incubated in a final volume of 20 µL for 4 h at room temperature. After washing four times with PBS, bound IgE were detected with a monoclonal AP-conjugated anti-human IgE antibody (BD Pharmingen, Heidelberg, Germany) diluted 1:1000. After washing four times with PBS, 50 µL of substrate solution (5 mg/mL 4-nitrophenylphosphate; AppliChem, Darmstadt, Germany) per well was added. The plates were read at 405 nm. The lower end functional cut-off indicated as lines was calculated as the mean of the negative controls plus two standard deviations.

5.4.6 Basophil activation test

In vitro basophil activation was determined by flow cytometry using the FlowCAST Assay, (BÜHLMANN Laboratories AG, Schönenbuch, Switzerland) as recently described [5]. Anti-FcεRI antibody and stimulation buffer served as positive and negative control, respectively. PVF1 and C1q were diluted in stimulation buffer and tested in a range of 0.01 to 1000 ng/ml. Native purified Api m 1 (Latoxan, Valence, France) was used at concentration ranging from 1 – 300 ng/ml. Flow cytometry was performed on a FACSCanto (Becton-Dicksinson, Heidelberg, Germany) using FACS Diva Software for measurement and FlowJo Software (Tree Star Inc, Ashland, OR, USA) for data analysis. In each assay a minimum of 500 basophils were assessed. Upregulation of the activation marker CD63 was calculated as the percentage of $CD63^+$ cells of total basophils ($CCR3^+$ SSC^{low}).

5.4.7 Sera and blood

Sera from 72 patients with anaphylactic reactions to honeybee stings were analyzed. Diagnosis of honeybee venom allergy was based on a combination of patient's history of an anaphylactic sting reaction, a positive skin test and positive IgE to honeybee venom

(ImmunoCAP i1), as recently described [14]. As defined by the inclusion criteria, all honeybee venom allergic patients displayed IgE to honeybee venom (≥0.35 kU$_A$/L) and 39 also tested positive to wasp venom (ImmunoCAP i3) (Table S5.1). Also serum total IgE titers and IgE titers for CCDs, serum tryptase, honeybee venom allergens (Api m 1, Api m 2, Api m 3, Api m 4, Api m 5, Api m 10) and wasp venom allergens (Ves v 5 and Ves v 1) were defined (Table S5.1).

5.5 RESULTS AND DISCUSSION

5.5.1 PVF1 heterogeneity

Mature *Pvf1* transcripts were amplified from honeybee venom gland tissue by reverse transcription-PCR. All cloned nucleotide sequences and their GenBank numbers are shown in Figure S5.1A. Sequence analysis of generated *Pvf1* amplicons revealed the existence of at least three variants (Figure S5.1B). The obtained nucleotide sequence of variant 1 matched with the GenBank record XM_392204.4, representing the predicted *Pvf1* mRNA sequence. Only one nucleotide substitution was found between the cloned fragment and the predicted NCBI sequence (Figure S5.1C), but the sequences are identical at the protein level (Figure S5.1D). Compared to variant 1, variant 2 lacks only a four-nucleotides sequence, while variant 3 contains an additional 82-nucleotide internal sequence. Besides, five nucleotide substitutions were found (Figure S5.1B), with three of them introducing an amino acid substitution (Figure 5.1). Alignment of these variants to the genome shows that all are generated by alternative splicing of the same gene (Figure S5.2). The *Pvf1* gene is positioned on chromosome LG2 and consists of six exons (GeneID: 408666). Three alternative 5' donor sites are present within exon 5: while variant 1 and variant 3 use an alternative canonical GT splice donor, variant 2 uses a non-canonical GC splice donor. BeeBase Blasts confirmed the existence of these *Pvf1* transcripts and their expression by honeybee tissues. A complete variant 1 (GenBank: gi|308418987) and variant 3 (GenBank: gi|308392023) EST was found in a brain/ovary and abdomen database, respectively. In contrast, only a partial variant 2 (GenBank: gi|10276753) EST is available from a honeybee antennae database.

Next, we tried to obtain proteomic evidence for the presence of specific variants within honeybee venom. Translation of the variant 1 transcript generates a protein of 292 amino

acids. In contrast, due to the alternative splice events, variant 2 and variant 3 use alternative stop codons, which generates C-terminally truncated variants with lengths of 272 and 249 amino acids respectively (Figure 5.2, Figure S5.1E). As the C-terminal sequence of these variants differs, we searched for variant-specific tryptic peptides from within this region using honeybee venom MS/MS data from preceding research (Figure 5.1). This analysis identified the ETQECSTGFYFDQNSCR peptide which is found in the C-terminal region of both variant 1 and variant 3, but not variant 2. Unfortunately, we were unable to detect additional distinguishing peptides.

```
variant3    QLEDTRYPDQRIVFPDRGRETANPALEGGPSGGGIGELAKSIQLAKKISSINSRDDFLKL  60
variant1    QLEDTRYPDQRIVFPDRGRETANPALEGGPSGGGIGELAKSIQLAKKISSINSRDDFLKL  60
variant2    QLEDTRYPDRRIVFPDRGRETANPALEGGPSGGGIGELAKSIQLAKKISSINSRDDFLKL  60
            *********:**************************************************

variant3    VKDVPKDISFFSSSSRMGETERSNAERPNQALCMPELQTVPLLENEPPVIYYPTCTRIKR  120
variant1    VKDVPKDISFFSSSSRMGETERSNAERPNQALCMPELQTVPLLENEPSVIYYPTCTRIKR  120
variant2    VKDVPKDISFFSSSSRMGETERSNAERPNQALCMPELQTVPLLENEPSVIYYPTCTRIKR  120
            **********************************************.*************

variant3    CGGCCTHSLLSCQPTATEIRNFEILVTILESSGKLKYQGKRIVPIEEHTQCTCDCKIKET  180
variant1    CGGCCTHSLLSCQPTATEIRNFEILVTILESSGKLKYQGKRIVPIEEHTQCTCDCKIKET  180
variant2    CGGCCTHSLLSCQPTATEIRNFEILVTILESSGKLKYQGKRIVPIEEHTQCTCDCKIKET  180
            ************************************************************

variant3    DCNKKQSYVPEECTCACNNVDEQKKCNESNIKMWHPDLCSCFCRETQECSTGFYFDQNSC  240
variant1    DCNKKQSYVPEECTCACNNVDEQKKCNESNIKMWHPDLCSCFCRETQECSTGFYFDQNSC  240
variant2    DCNKKQSYVPEECTCACNNVDEQRKCNESNIKMWHPDLCSCFCRETQECSTGFYFDQNSY  240
            ***********************:************************************

variant3    RCERN-----------NKDL------------------------------  249
variant1    RCLQVPLSRTWFTSTKGSDYRFGQTQRPDNVPPVIIALDSDDPRRKPKPDPE  292
variant2    ---------------ACKYRY---LEHGLHPQKVLIIDSDKHKDQIMYHR-  272
```

Figure 5.1: Protein sequence alignment of the PVF1 variants. Amino acid substitutions are indicated in red. All tryptic peptides are shown, which have been identified by searching honeybee venom MS/MS data from preceding research [8] against the RefSeq honeybee protein database extended with the three translated PVF1 isoform sequences (yellow). The typical eight cysteines of the central platelet-derived and vascular endothelial growth factor domain are indicated in boxes.

PVF1 belongs to the PDGF/VEGF (platelet-derived growth factor/vascular endothelial growth factor) family. While in vertebrates two networks, PDGF/PDGFR (PDGF receptor) and VEGF/VEGFR (VEGF receptor), have evolved, invertebrates possess a single PVF/PVR (PDGF/VEGF-like factors/PVF receptor-related) network [15]. In the fruit fly *Drosophila melanogaster*, three *Pvf* genes were found (*Pvf1*, *Pvf2* and *Pvf3* with GeneIDs: 32876; 33994;

33995), which are ligands for a single PDGF/VEGF receptor [16]. The *Pvf* genes have distinct expression patterns in the fruit fly embryo [17] and execute different functions during development (reviewed in [15]). Honeybee venom PVF1 shows highest similarity to the fruit fly PVF1. Also two splice variants of the fruit fly PVF1 have been described, but in contrast to the honeybee homologue these differ at their N-termini [16]. Alternative splicing of PVF1 transcripts in both species doesn't interfere with the central PDGF/VEGF domain, which includes the typical eight conserved cysteines important for dimerization and functional activity (Figure 5.1 and 5.2) [18]. Honeybee venom PVF1 was suggested to have acquired a new function similar to snake venom VEGF-like compounds which can facilitate venom spreading by increasing vascular permeability [8]. Two additional gene predictions containing a PDGF/VEGF domain were found in the honeybee genome, while these compounds were detected in none of the honeybee venom proteome analyses. These gene predictions (GeneIDs: 100577397 and 100579031) are positioned on different chromosomes and both show highest sequence resemblance to fruit fly *Pvf3*. While fruit fly *Pvf2* and *Pvf3* were suggested to be generated by a recent gene duplication due to their close proximity in the genome [16], no additional *Pvf* gene was found within the honeybee genome adjacent to both *Pvf3*-resembling genes.
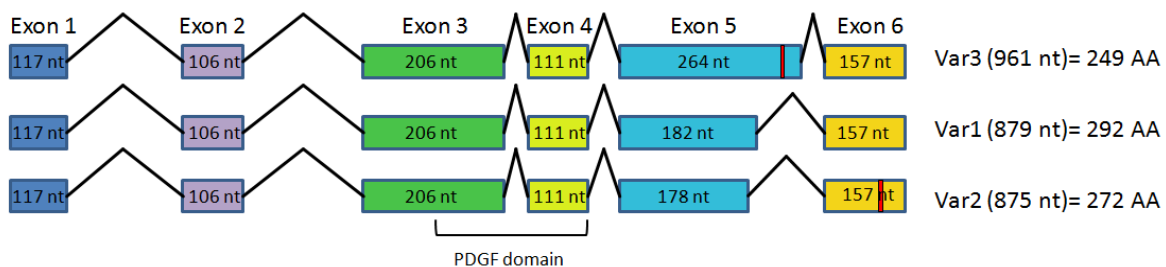

Figure 5.2: Schematic figure showing the intron-exon structure of sequenced *Pvf1* amplicons, named variant 1 to 3. Different exons are shown as colored boxes, while introns are presented as lines. Both are drawn to scale. Red boxes present alternative stop codons, which generate C-terminally truncated variants. The full coding sequence lengths are shown: number of nucleotides (nt), number of amino acids (AA).

5.5.2 Recombinant production of C1q and PVF1

To assess immunoreactivity of C1q and PVF1, both mature proteins were produced as recombinants by baculovirus-mediated infection of Sf9 (*Spodoptera frugiperda*) insect cells.

As variant 1 is the largest PVF1 protein variant and as a tryptic peptide from its C-terminal region was detected in the venom proteome, it was selected for immunological characterization. Both C1q and PVF1 proteins were obtained as secreted, soluble proteins, which were purified (Figure 5.3). Coomassie and/or anti-His staining of both SDS-PAGE separated proteins revealed a double protein band pattern in the region corresponding to their theoretical MW. This double protein band pattern indicates that a fraction of glycosylated (highest band) and non-glycosylated (lowest band) protein is produced. Indeed, according to the NetNGlyc Prediction server both compounds contain one predicted N-linked glycosylation site. Additionally, Sf9 insect cells are known to produce glycosylated proteins [11]. However, their produced carbohydrate structures are devoid of α-1,3-core fucosylations, also known as cross-reactive carbohydrate determinants (CCDs). As CCDs are recognized by IgEs, this cell line is an excellent choice for determination of immunoreactivity to proteinous epitopes exclusively [4].
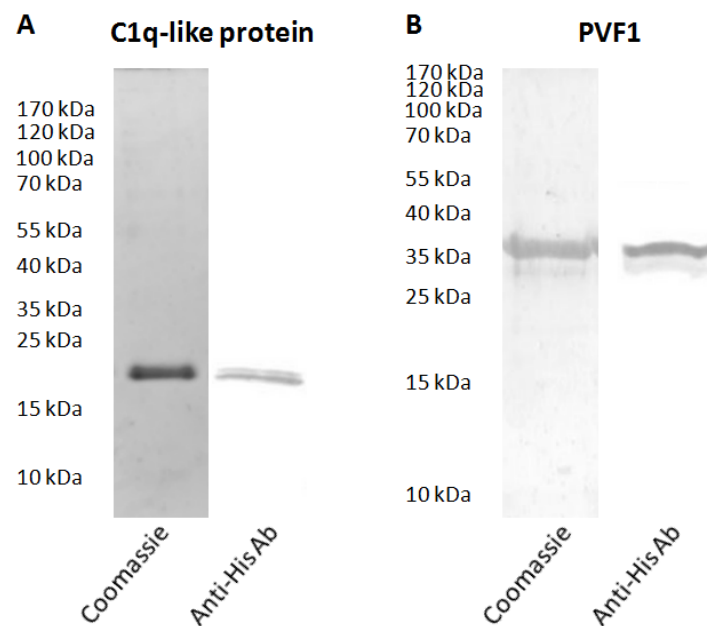

Figure 5.3: SDS-PAGE and Western blot analysis of C1q (panel A) and PVF1 (panel B) proteins recombinantly produced in Sf9 insect cells. Proteins are visualized by Coomassie blue staining and anti-His epitope antibody. Mature C1q has a theoretical molecular weight (MW) of 15 kDa, while PVF1 variant has a MW of approximately 33 kDa. Fusion tags increased the size of the protein 3 kDa in the gel and blot. A double protein band pattern is visualized which indicates that a fraction of glycosylated (highest band) and non-glycosylated (lowest band) protein is produced.

5.5.3 Serum specific IgE reactivity of C1q and PVF1 recombinants

IgE recognition of C1q and PVF1 was tested by ELISA using 72 sera of honeybee venom allergic patients. Specific IgE reactivity to C1q was observed in 24 (33.3%) sera, while 19 (26.3%) sera recognized PVF1 (Figure 5.4; Table S5.1). In addition, six and seven sera showed minor reactivity with C1q and PVF1, respectively. As their reactivity is only slightly above the cut-off, this might represent only background binding. In contrast to a preceding preliminary study which demonstrated the lack of IgE recognition of a non-glycosylated, prokaryotic produced C1q [9], this study revealed IgE reactivity of the glycosylated, insect cell-produced C1q protein. However, as the present study revealed IgE reactivity in only 1/3 of the 72 analyzed sera, the three sera of honeybee venom allergic patients used in the preliminary study were insufficient to detect IgE reactivity. Additionally, the preliminary study analyzed IgE reactivity of two wasp allergic patients, but these did not recognize the honeybee venom C1q recombinant. Further studies should confirm if cross-reactivity between honeybee and wasp venom C1q is lacking or if a C1q-like protein is absent from wasp venom. IgE binding of PVF1 may also have been reported in a previous immunoblot study [19]. Indeed, two patients showed weak IgE recognition to a 54 kDa protein with an N-terminal XXAERPNQAS sequence. However, this sequence is not present within the honeybee RefSeq protein database, while the AERPNQAL sequence can be found exclusively within all variant sequences of PVF1. Further processing of the mature PVF1 protein may position this sequence in an N-terminal position. Additionally, as the mature PVF1 variants have much lower molecular weights (32.9, 31 and 28.1 kDa) than the reported 54 kDa band, protein dimer formation or interaction with other venom proteins may allow detection of PVF1 in the 54 kDa molecular weight region.

In many European countries, the European honeybee (*A. mellifera*) and wasps (*V. vulgaris*) are the most prevalent stinging insects. As many patients fail to identify or name the species that stung and as a correct allergy diagnosis is required for the initiation of an appropriate immunotherapy, several research studies have recently focused on their differential allergy diagnosis by component-resolved diagnosis (CRD) [20-22]. CRD relies on quantification of sIgE antibodies to single components, which can help to distinguish between a true double sensitisation and cross-sensitization to several unrelated allergen sources. Using the combination of Ves v 1 and Ves v 5, between 92% and 96% of the wasp allergic patients can
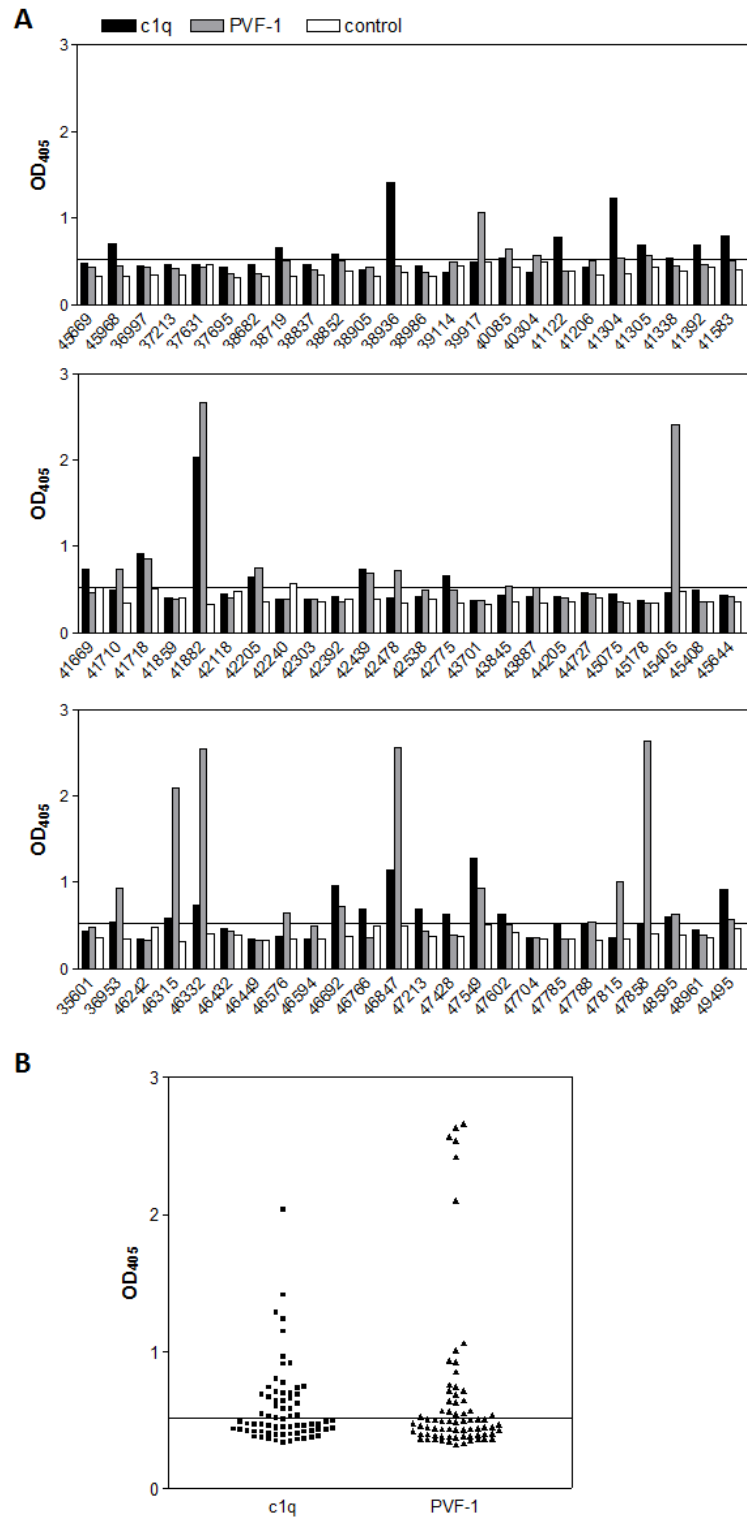
Figure 5.4: IgE immunoreactivity of individual honeybee venom-sensitized patient sera with recombinant C1q and PVF1. The IgE reactivity was assessed by ELISA with 72 sera of honeybee venom-sensitized patients. The lower end functional cut-off of the ELISA is represented by a solid line. (A) Schematic presentation of the optical density values (OD405) of C1q, PVF1 and a control for each serum sample. Serum numbers correlate to the numbers found in table S5.1. (B) Scatter chart of the OD405 values of C1q and PVF1 for all sera.

be diagnosed [23-25]. In contrast, rApi m 1 is the only honeybee venom allergen commercially available for CRD of honeybee venom allergy [26]. Due to the moderate rApi m 1 sensitization within the population of honeybee venom allergic patients [23;24;27-29], additional venom allergens should be used to increase sensitivity [3]. The panel of 72 sera of honeybee venom allergic patients used for our ELISA analyses was pre-screened by ImmunoCAP tests for IgE reactivity towards four honeybee venom-specific allergens, Api m 1, Api m 3, Api m 4 and Api m 10, and two honeybee-wasp cross-reactive allergens, Api m 2 and Api m 5. In total, nine sera lacked IgE reactivity to all of the honeybee venom-specific allergens, which indicates that CRD using these four allergens would not detect 12.5% of the honeybee venom allergic patients. Interestingly, our ELISA experiment demonstrated that five of these sera showed IgE reactivity to C1q and two to PVF1, including one serum sample showing IgE reactivity to both compounds (Table S5.1). Therefore, the panel of honeybee venom-specific allergens in combination with C1q and PVF1 allows to increase the current sensitivity of the Api m 1-based CRD for honeybee venom allergy to more than 95%. None of these patients showed IgE reactivity to wasp venom, demonstrating that the observed reactivity is not due to putatively existing cross-reactive IgEs directed towards possible C1q and PVF1 wasp venom homologs. Future proteomics studies focusing on wasp venoms should reveal if these contain C1q and PVF1 homologs.

5.5.4 Basophil activation

The capacity of recombinant C1q and PVF1 to activate basophils was addressed in bee venom allergic patients (n=7) that displayed positive IgE reactivity to C1q (n=4, OD range 0.702 – 1.236) and/or PVF1 (n=5, OD range 0.733 – 2.567) in the ELISA. In none of the patients C1q or PVF1 induced significant basophil activation when tested over a broad range of 0.01 to 1000 ng/ml. In contrast, stimulation with Api m 1 induced a dose dependent basophil activation in those patients that displayed sIgE reactivity to Api m 1 (Figure 5.5). The lack of basophil activation by C1q and PVF1 in all patients, suggests that both proteins harbour only one IgE epitope and are thus unable to cross link the FcεRI on basophils. Alternatively, the relative IgE reactivity to each of the proteins is too low to be detected by the basophil activation test. Finally, both proteins could exert inhibitory functions on basophil activation that neutralizes the activation signal provided by FcεRI crosslinking. Further studies will have to address these issues.
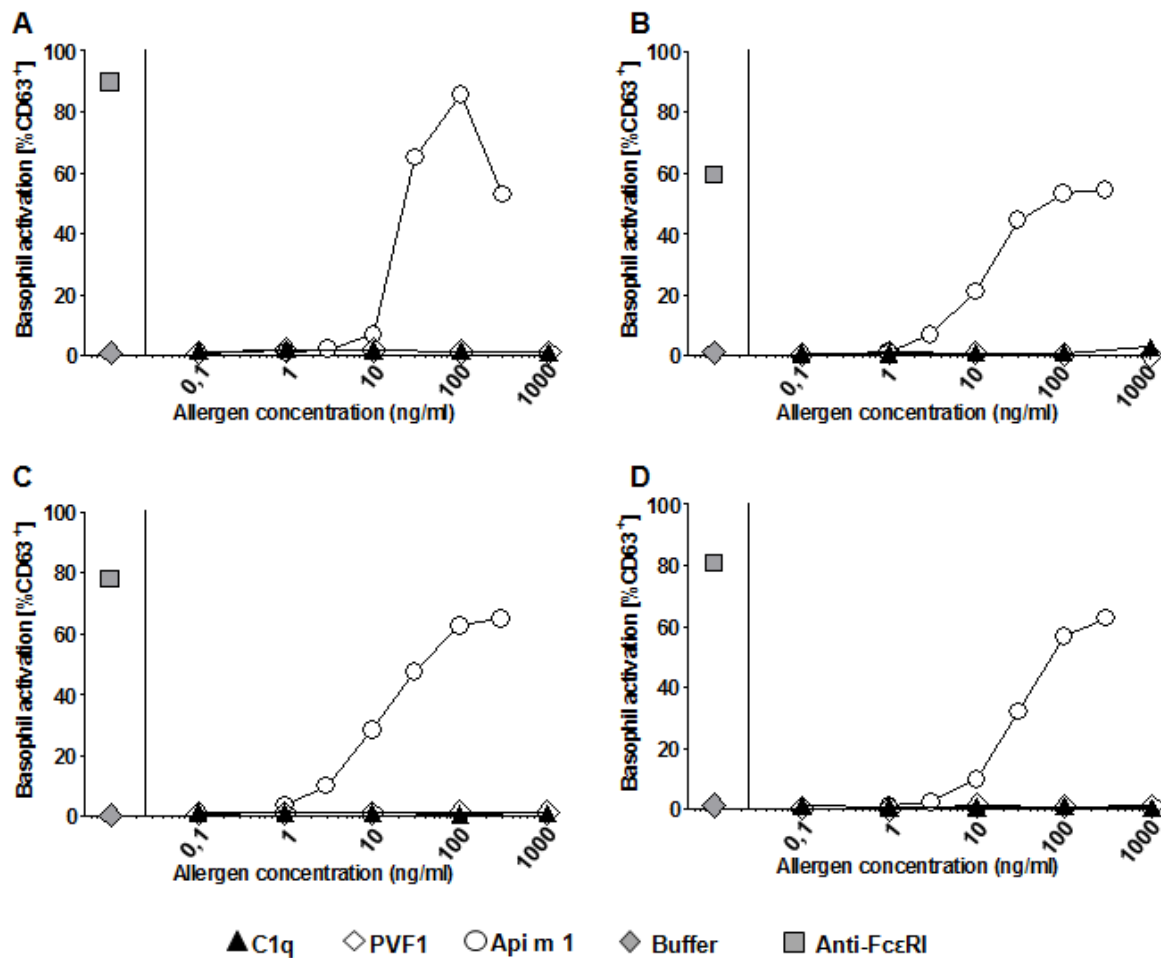
Figure 5.5: Basophil activation tests with recombinant C1q and PVF1. Human basophils from 4 exemplary honeybee venom-sensitized patients that displayed positive IgE reactivity to C1q and/or PVF1 were exposed to serial dilutions of recombinant C1q (filled triangles) and PVF1 (open rhombus) over a broad range of 0.01 to 1000 ng/ml. Api m 1 was used as an established reference allergen (open circles) at concentrations ranging from 1 – 300 ng/ml. Additionally, stimulation with anti-FcεRI antibody (filled squares) and plain stimulation buffer (filled rhombus) is shown. Activation is shown as percentage of CD63+ cells.

5.6 CONCLUSIONS

This study successfully identified three alternative splice variants of PVF1, which were found to be expressed by the venom glands using an RT-PCR-based approach. The baculovirus-based insect cell expression system showed to be appropriate for the recombinant production of the honeybee venom glycoproteins C1q and PVF1. A minor group of honeybee venom allergic patients showed serum IgE reactivity to these compounds. We could

demonstrate that, in combination with other honeybee venom-specific allergens, both compounds can significantly improve sensitivity of the current Api m 1-based CRD for honeybee venom allergy. However, both compounds are unable to activate basophils, rendering their relevance in the context of allergy questionable. This remarkable observation requires further investigation.

5.7 ACKNOWLEDGEMENTS

5.8 ADDENDUM

Supplementary figures and tables can be found on the included CD-ROM or can be requested by e-mail from matthiasvanvaerenbergh@hotmail.com and Dirk.deGraaf@UGent.be.

Figure S5.1: A) This panel presents the nucleotide sequences and the GenBank accession numbers of sequenced mature *Pvf1* amplicons, generated by reverse transcription-PCR on honeybee venom gland tissue. B) Sequence alignment of *Pvf1* nucleotide sequences (inferred by ClustalW2). Nucleotide substitutions are indicated in red. C) The obtained nucleotide sequence of mature variant 1 matched with the GenBank record XM_392204.4, representing the predicted honeybee *Pvf1* mRNA sequence. Only one nucleotide substitution was found between the cloned fragment and the predicted NCBI sequence (indicated in red). D) Amino acid sequence alignment of the obtained mature PVF1 variant 1 sequence with the GenBank honeybee PVF1 sequence (XP_392204.2). E) This panel presents the translated mature PVF1 variant sequences.

Figure S5.2 shows the alignment of detected *Pvf1* variants against the *Pvf1* gene structure (GeneID: 408666) on chromosome LG2. Intron 5'GT donor and 3'AG acceptor splice sites conform to the general canonical GT-AG splicing rule are presented in purple. The three

alternative 5' donor sites present within exon 5 are shown in green. Alternative stop codons are indicated in red.

Table S5.1 presents the ImmunoCAP data and ELISA results of the 72 sera of honeybee venom allergic patients. Specific IgE levels (in kU/l) for honeybee venom, Api m 1, Api m 2, Api m 3, Api m 4, Api m 5, Api m 10, wasp venom, Ves v 5, Ves v 1 and cross-reactive carbohydrate determinants (CCD) were determined using ImmunoCAP. In addition, total serum IgE, serum tryptase levels and anamnesis (B: honeybee venom; W: wasp venom; BB: bumblebee venom; NC: not clear) are included. The ELISA optical densities measured at 405 nm (OD405) are given for C1q, PVF1 and a control (buffer). Values indicated in red are the double of the control value, while values in yellow are clearly above the cut-off, but not the double of the control value. Values coloured in green are only slightly above the cut-off, which may be caused by background binding.

5.9 REFERENCES

[1] Blank S, Bantleon FI, McIntyre M, Ollert M, Spillner E. The major royal jelly proteins 8 and 9 (Api m 11) are glycosylated components of *Apis mellifera* venom with allergenic potential beyond carbohydrate-based reactivity. Clin Exp Allergy 2012 Jun;42(6):976-85.

[2] Valenta R, Vrtala S. Recombinant allergens for specific immunotherapy. Allergy 1999;54:43-4.

[3] Hofmann SC, Pfender N, Weckesser S, Blank S, Huss-Marp J, Spillner E, et al. Detection of IgE to recombinant Api m 1 and rVes v 5 is valuable but not sufficient to distinguish bee from wasp venom allergy Reply. Journal of Allergy and Clinical Immunology 2011 Jul;128(1):248.

[4] Blank S, Seismann H, McIntyre M, Ollert M, Wolf S, Bantleon FI, et al. Vitellogenins Are New High Molecular Weight Components and Allergens (Api m 12 and Ves v 6) of *Apis mellifera* and *Vespula vulgaris* Venom. PLoS ONE 2013;8(4):e62009.

[5] Blank S, Seismann H, Michel Y, McIntyre M, Cifuentes L, Braren I, et al. Api m 10, a genuine *A. mellifera* venom allergen, is clinically relevant but underrepresented in therapeutic extracts. Allergy 2011 Oct;66(10):1322-9.

[6] Peiren N, Vanrobaeys F, de Graaf DC, Devreese B, Van Beeumen J, Jacobs FJ. The protein composition of honeybee venom reconsidered by a proteomic approach. Biochimica et Biophysica Acta-Proteins and Proteomics 2005 Aug 31;1752(1):1-5.

[7] Blank S, Seismann H, Bockisch B, Braren I, Cifuentes L, McIntyre M, et al. Identification, Recombinant Expression, and Characterization of the 100 kDa High Molecular Weight Hymenoptera Venom Allergens Api m 5 and Ves v 3. Journal of Immunology 2010 May 1;184(9):5403-13.

[8] Van Vaerenbergh M, Debyser G, Devreese B, de Graaf DC. Exploring the hidden honeybee (*Apis mellifera*) venom proteome by integrating a combinatorial peptide ligand library approach with FTMS. Journal of Proteomics. In press 2013.

[9] de Graaf DC, Brunain M, Scharlaken B, Peiren N, Devreese B, Ebo DG, et al. Two novel proteins expressed by the venom glands of *Apis mellifera* and *Nasonia vitripennis* share an ancient C1q-like domain. Insect Mol Biol 2010 Feb;19 Suppl 1:1-10.

[10] Peiren N, de Graaf DC, Vanrobaeys F, Danneels EL, Devreese B, Van Beeurnen J, et al. Proteomic analysis of the honey bee worker venom gland focusing on the mechanisms of protection against tissue damage. Toxicon 2008 Jul;52(1):72-83.

[11] Seismann H, Blank S, Braren I, Greunke K, Cifuentes L, Grunwald T, et al. Dissecting cross-reactivity in hymenoptera venom allergy by circumvention of alpha-1,3-core fucosylation. Molecular Immunology 2010 Jan;47(4):799-808.

[12] Birnboim HC, Doly J. Rapid Alkaline Extraction Procedure for Screening Recombinant Plasmid Dna. Nucleic Acids Research 1979;7(6):1513-23.

[13] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007 Nov 1;23(21):2947-8.

[14] Hofmann SC, Pfender N, Weckesser S, Blank S, Huss-Marp J, Spillner E, et al. Detection of IgE to a panel of species specific allergens further improves discrimination of bee and wasp venom allergy. Journal of Allergy and Clinical Immunology 2011;128(1):247-8.

[15] Hoch RV, Soriano P. Roles of PDGF in animal development. Development 2003 Oct;130(20):4769-84.

[16] Cho NK, Keyes L, Johnson E, Heller J, Ryner L, Karim F, et al. Developmental control of blood cell migration by the *Drosophila* VEGF pathway. Cell 2002 Mar 22;108(6):865-76.

[17] Harris KE, Schnittke N, Beckendorf SK. Two ligands signal through the *Drosophila* PDGF/VEGF receptor to ensure proper salivary gland positioning. Mech Dev 2007 Jul;124(6):441-8.

[18] Tarsitano M, De FS, Colonna V, McGhee JD, Persico MG. The C. elegans pvf-1 gene encodes a PDGF/VEGF-like factor able to bind mammalian VEGF receptors and to induce angiogenesis. FASEB J 2006 Feb;20(2):227-33.

[19] Kettner A, Henry H, Hughes GJ, Corradin G, Spertini F. IgE and T-cell responses to high-molecular weight allergens from bee venom. Clinical and Experimental Allergy 1999 Mar;29(3):394-401.

[20] Seismann H, Blank S, Cifuentes L, Braren I, Bredehorst R, Grunwald T, et al. Recombinant phospholipase A1 (Ves v 1) from yellow jacket venom for improved diagnosis of hymenoptera venom hypersensitivity. Clin Mol Allergy 2010;8:7.

[21] Mittermann I, Zidarn M, Silar M, Markovic-Housley Z, Aberer W, Korosec P, et al. Recombinant allergen-based IgE testing to distinguish bee and wasp allergy. J Allergy Clin Immunol 2010 Jun;125(6):1300-7.

[22] Müller UR, Johansen N, Petersen AB, Fromberg-Nielsen J, Haeberli G. Hymenoptera venom allergy: analysis of double positivity to honey bee and Vespula venom by estimation of IgE antibodies to species-specific major allergens Api m1 and Ves v5. Allergy 2009 Apr;64(4):543-8.

[23] Korosec P, Valenta R, Mittermann I, Celesnik N, Silar M, Zidarn M, et al. High sensitivity of CAP-FEIA rVes v 5 and rVes v 1 for diagnosis of Vespula venom allergy. Journal of Allergy and Clinical Immunology 2012 May;129(5):1406-8.

[24] Hofmann SC, Pfender N, Weckesser S, Blank S, Huss-Marp J, Spillner E, et al. Detection of IgE to recombinant Api m 1 and rVes v 5 is valuable but not sufficient to distinguish bee from wasp venom allergy Reply. Journal of Allergy and Clinical Immunology 2011 Jul;128(1):248.

[25] Ebo DG, Faber M, Sabato V, Leysen J, Bridts CH, De Clerck LS. Component-resolved diagnosis of wasp (yellow jacket) venom allergy. Clinical and Experimental Allergy 2013 Feb;43(2):255-61.

[26] Sturm GJ, Bilo MB, Bonadonna P, Hemmer W, Caruso B, Bokanovic D, et al. Ves v 5 can establish the diagnosis in patients without detectable specific IgE to wasp venom and a possible north-south difference in Api m 1 sensitization in Europe. Journal of Allergy and Clinical Immunology 2012 Sep;130(3):817.

[27] Hofmann SC, Pfender N, Weckesser S, Huss-Marp J, Jakob T. Added value of IgE detection to rApi m 1 and rVes v 5 in patients with Hymenoptera venom allergy. Journal of Allergy and Clinical Immunology 2011 Jan;127(1):265-7.

[28] Sturm GJ, Hemmer W, Hawranek T, Lang R, Ollert M, Spillner E, et al. Detection of IgE to recombinant Api m 1 and rVes v 5 is valuable but not sufficient to distinguish bee from wasp venom allergy. Journal of Allergy and Clinical Immunology 2011 Jul;128(1):247-8.

[29] Jakob T, Köhler J, Blank S, Huss-Marp J, Spillner E, Lidholmh J. Comparable IgE-reactivity to nApi m 1 and rApi m 1 in CCD negative bee venom allergic patients. Journal of Allergy and Clinical Immunology 2012;130:276-8.

General discussion

1. THE HONEYBEE AND BUMBLEBEE VENOM PROTEOME

By integrating genome, transcriptome and proteome information, this PhD work obtained in-depth insights in the complexity of the honeybee and bumblebee venom composition. Our proteomic studies revealed the presence of more than 100 honeybee (Chapter 1 and 2) and 57 bumblebee venom (Chapter 3) compounds. Both venoms contain approximately 30 (putative) toxins which indicates that they have a rather low complexity compared to many cone snail, scorpion and spider venoms. These species have massive numbers of toxins, numbering in the tens of thousands, which are typically short peptides evolved by large-scale gene duplications [1]. This level of complexity has not been observed in any hymenopteran species. Honeybee and bumblebee venom contains many typical venom constituents. Indeed, in contrast to the extensive diversity of venomous organisms, the number of different protein scaffolds is restricted [2]. Throughout evolution, at least 14 protein types have been convergently recruited into the venom by two or more venomous lineages (Table 1). The identification of a C-type lectin in honeybee venom (Chapter 2) adds a new class of convergently recruited proteins to the group of insect sting toxins (Table 1). Moreover, many of the same protein families have also been convergently recruited for use in the hematophagous gland secretions of invertebrates (e.g. fleas, leeches, kissing bugs, mosquitoes, and ticks) and vertebrates (e.g. vampire bats) [2]. The high proportion of convergently recruited protein families suggests that there are structural and/or functional constraints that make a protein suitable for recruitment as a toxin. Toxins typically contain a secretory protein ancestor, functionally versatile protein ancestors with a fundamentally conserved basal activity, extensive disulfide cross-links, stable molecular scaffolds, and once recruited, adaptive evolution generates novel toxins by gene duplication [2]. Gene annotation revealed that also several identified honeybee and bumblebee venom genes have been generated by gene duplications (Chapter 2 and 3). Honeybee venom acid phosphatase 1 and 3 are tandemly positioned on chromosome 5, MRJP8 and MRJP9 on chromosome 11, and MCDP, apamin and tertiapin on chromosome 12, while the bumblebee

venom genes PLA$_2$-1 and PLA$_2$-2 are tandemly positioned on chromosome 13 and the six serine protease genes on chromosome 4. Additional toxin genes can increase expression levels in the glands and result in higher toxin doses. Newly created genes may also obtain different levels of potency, functions, or complementary specificities [2]. Besides the convergence of recruited proteins to serve as toxins, convergence of activities can be found between different venoms. For example, honeybee venom apamin is a neurotoxin blocking potassium channels, an activity which is also executed by cone snail conotoxins, snake venom dendrotoxins, spider venom atracotoxins and cnidaria kunitz-type protease inhibitors [2].

Table 1: Convergently recruited venom proteins. Abbreviations used: AVIT: AVIT/Colipase/Prokineticin proteins; CAP: CRISP (cysteine rich secretory proteins), antigen 5 (Ag5) and pathogenesis-related (PR-1) proteins; Chi: chitinase; Cys: cystatin; Def: defensin; Hya: hyaluronidase; Kun: kunitz; lec: lectin; Lip: lipocalin; Nat: natriuretic peptide; PS1: peptidase S1; PLA2: phospholipase A2;Sm-D: sphingomyelinase; SPRY: SPRY/Concavalin A–Like lectins. Table adapted from [2].

	AVIT	CAP	Chi	Cys	Def	Hya	Kun	Lec	Lip	Nat	PS1	PLA$_2$	Sm-D	SPRY
Amphibian	X													
Cephalopod		X	X			X					X	X		
Cnidarian							X					X		
Cone snail		X					X							
Fish						X		X						X
Insect (bristle)				X			X	X			X	X		
Insect (proboscis)		X		X		X	X	X	X		X	X		
Insect (stinger)		X	X			X	X	X			X	X		
Platypus					X					X				
Scorpion		X			X	X	X					X		
Shrew											X			
Spider	X	X				X	X(2)						X	
Reptile	X	X		X(2)	X	X	X	X	X	X	X	X(3)		X
Tick	X	X		X		X			X		X	X	X	

This PhD work analyzed the honeybee and bumblebee venom proteome using an identical technological approach, which allowed to compare both venom compositions (Chapter 2 and 3). Honeybees and bumblebees have diverged already 77-95 million years ago [3]. Nevertheless, 72% of the detected bumblebee venom compounds proved to have a honeybee venom homologue which reflects the similar defensive function of both venoms and the high degree of homology between both genomes. Previously, the venom of another hymenopteran species, *Nasonia vitripennis*, was explored by shotgun proteomics [4]. Sixty

compounds were found in the venom of this ectoparasitoid wasp. Honeybee, bumblebee and *Nasonia* venom appear to have only one venom homologue in common: DPP IV. This high diversity in venom composition between *Nasonia* and both bees is due to a different venom function and their long evolutionary divergence (honeybee and *Nasonia* have diverged about 190 million years ago; [5]). DPP IV has also been found in the venom of *Vespula vulgaris* and multiple snake species. A phylogenetic analysis based on these venom DPP IV sequences shows the honeybee and bumblebee as closest relatives (Figure 1).
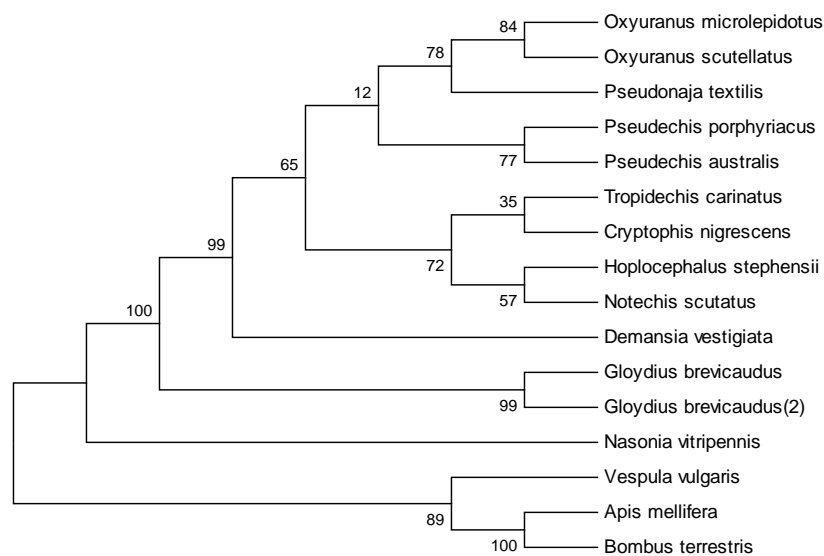

Figure 1: Phylogenetic tree based on venom dipeptidyl peptidase IV sequences, generated by maximum parsimony. Numbers indicate parsimony bootstrap scores for the branch.

Over the years, multiple complementary proteomic approaches have been applied to explore the honeybee worker venom proteome. Peiren and coworkers [6] and de Graaf and coworkers [7] separated honeybee venom proteins by 2D-PAGE and analyzed excised spots by MALDI-TOF/TOF and Q-TRAP LC-MS/MS. Besides six well-known honeybee venom compounds, four novel compounds were identified. However, these studies revealed that the conducted gel-based proteomics strategy faces several limitations. This PhD work applied two novel approaches to gain deeper insights in the honeybee venom composition. First, the venom was separated by HPLC and protein fractions were analyzed by MALDI-TOF/TOF (Chapter 1). This analysis confirmed the presence of nine honeybee venom compounds, including two peptides which remained undetected in the preceding gel-based studies due to their low molecular weight. In addition, a novel venom peptide with

antimicrobial properties, called apidaecin, was found. In contrast, the presence of several compounds found in gel-based studies, such as MRJP9, PVF1 and C1q, could not be confirmed by this approach (Figure 2). The second proteomic study conducted during this PhD was much more sensitive, as lowly abundant compounds were enriched by the ProteoMiner technology (Chapter 2). Moreover, no other mass analyzer matches the resolved power and mass accuracy of the applied FT-ICR technology, with the exception of the Orbitrap technology [8;9]. Both mass analyzers measure m/z values as frequencies, which can be obtained more accurately than any other experimental parameter. This approach allowed a 10-fold increase in the number of indentified compounds compared to the preceding approaches. In total, 102 venom compounds were found, including 83 newly discovered proteins. All compounds from the previously mentioned analyses were found, except MCDP (although found at less stringent search parameters) and apidaecin (Figure 2).

In addition, only very recently, the venom composition of the Africanized honeybees and two European subspecies, *A. m. ligustica* and *A. m. carnica*, has been studied applying a shotgun LC-MS/MS analysis incorporating the Orbitrap mass analyzer [10]. In total, 51 proteins were found with 42 being common to all three subspecies. This study identified 43 compounds in the venom of *A. m. carnica*, the subspecies which was also studied in our proteomic analyses. Remarkably, 8 venom compounds were found which were not detected in our proteomics studies (Figure 2). These include the venom peptide tertiapin, which has previously been found by Edman degradation sequencing of the chromatographically purified peptide [11], and odorant binding protein 14, which has previously been found in a MS analysis of the venom gland tissue [12]. Also, 6 novel compounds are described including MRJP2, MRJP3, a chymotrypsin inhibitor, multiple coagulation factor deficiency protein 2 homolog and two proteins with an unknown function. Several factors may explain this observed variation, including technological (venom collection method, proteomic approach, search parameters) and biological (geographical, seasonal, age-related variation in venom composition) factors. It is especially peculiar that besides MRJP8 and MRJP9 no other MRJP proteins have been detected in our proteomic analyses, while this study also reports on the detection of MRJP2 and MRJP3. An approximately equal number of tryptic peptides matching with MRJP8 was detected in the venoms of the three subspecies, which may indicate that this MRJP is equally abundant in the venoms of the three subspecies. The same observation has been made for MRJP9. In contrast, a high number of peptides matching with

MRJP2 and MRJP3 were found in the venom of the Africanized honeybees, while only 1 or 2 peptides were found in the venoms of European subspecies. Therefore, MRJP2 and MRJP3 may be much lower abundant in the venoms of the European subspecies. Moreover, MRJP1 and MRJP5 were exclusively found in the Africanized subspecies. As the MRJPs 1 to 5 are major compounds of the secreted royal jelly which is provided by nurses as food to the larvae and queen [13], their function in the venom requires further investigation. Contamination of the venom samples cannot be ruled out. In contrast, MRJP8 and MRJP9 were found to be the most ancient members of the MRJP protein family, which lack the later evolved repetitive regions suggested to have a nutritious function. Therefore, both MRJPs may possess the original but yet unknown pre-royal jelly function [12].
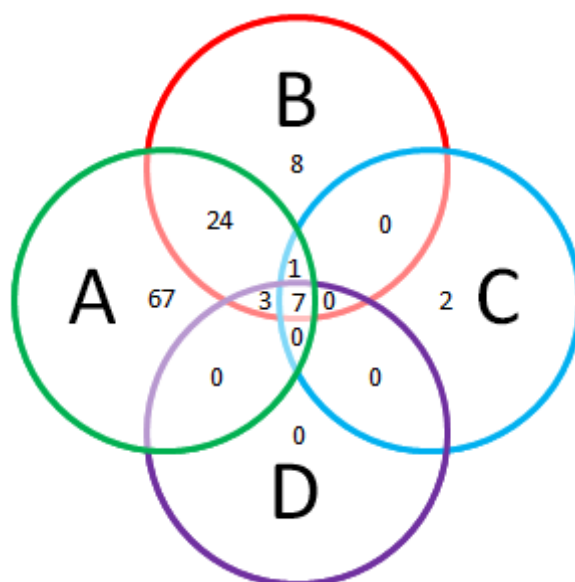

Figure 2: Overview of the number of honeybee (*A. mellifera carnica*) venom compounds identified in four different studies. A: CPLL + 1D-PAGE + LC-ESI-LTQ-FT-ICR-MS/MS (Chapter 2); B: shotgun Orbitrap-LC MS/MS [10]; C: HPLC + MALDI-TOF/TOF (Chapter 1); D: 2D-PAGE + MALDI-TOF/TOF [6;7].

The low sensitivity proteomic approaches (2D-PAGE or HPLC followed by MALDI-TOF/TOF) mainly identified highly and moderately abundant compounds. Most of these compounds are toxins, contributing to the defence and social immunity functions of the venom. Now, the novel high sensitivity approaches (sample pre-treatment by CPLL, Fourier transform-based MS) enabled to dig deeper in the complex honeybee venom proteome than ever before. The group of newly identified compounds includes several potential toxins. However, as their biological function was only predicted based on the experimentally determined

function of homologues, and the finding of GO-terms, functional domains and venom homologues, their function should be further experimentally investigated. In addition, these highly sensitive approaches identified many low and extremely lowly abundant compounds, which are called venom trace molecules as they serve no toxic functions. These include a group of common secretory proteins exerting essential functions in the extracellular space, e.g. immunity-related proteins and apolipophorins. However, most trace molecules are typical secretory pathway proteins being unintentionally released due to an inefficient retrieval and retrograde transport within the secretory pathway of the highly active secreting venom gland tissue.

Many research studies have focused on unravelling the honeybee venom proteome. Nevertheless, many unanswered questions remain. For example, prediction databases often miss peptide sequences as their determination from a genetic structure is very difficult [14]. As our mass spectrometry experiments applied a database search approach, in addition to melittin, apamin, MCDP, secapin and tertiapin, multiple unknown peptides may be present in the venom. Indeed, several venom peptides have been isolated in the 1970's and 80's by chromatographic means, e.g. minimine [15], cardiopep [16] and adolapin [17] but amino acid sequences are still lacking. Future studies should apply peptidomics, characterizing peptides by MS-driven de novo peptide sequencing [14], to explore the venom peptidome. Also the variation in the venom content has only been scarcely investigated. Within a single hive in Brazil, seasonal variation in the PLA_2 and melittin venom levels has been demonstrated, which did not correlate to climatic factors [18]. In addition, our finding that transcription of the Ag5-like gene by the venom glands is restricted to winter bees may be indicative for qualitative seasonal variation in the venom composition (Chapter 1). However, a proteomic study of the winter bee venom has not yet been conducted. Variation in the venom composition between the physiologically distinct summer and winter bees can be expected. In the moderate climate zone, summer bees actively forage and live about six weeks, while winter bees live up to six months and stay within the hive to generate heat to keep the hive warm. Also in-depth insights in age-, caste-, colony-related and geographical variability are lacking. During this PhD work, the venom of the queen was submitted to SDS-PAGE and pre-treated using the ProteoMiner technology (Figure 3). Queen SDS-PAGE patterns clearly differ from those of honeybee workers. An identical mass spectrometry analysis as that applied for

the honeybee worker venom should provide insights in this caste-related variability. Besides qualitative variation, also quantitative variability in venom composition may be present. A variety of techniques for differential quantification are available, such as iTRAQ™ (Isobaric Tags for Relative and Absolute Quantitation), DIGE (Differential Gel Electrophoresis) and spectral counting [19]. For absolute quantification of a particular protein compound, standard curves and AQUA™ peptides can be used. During this PhD work, no quantitative data have been generated as the CPLL sample-pre-treatment shifts relative protein abundances.
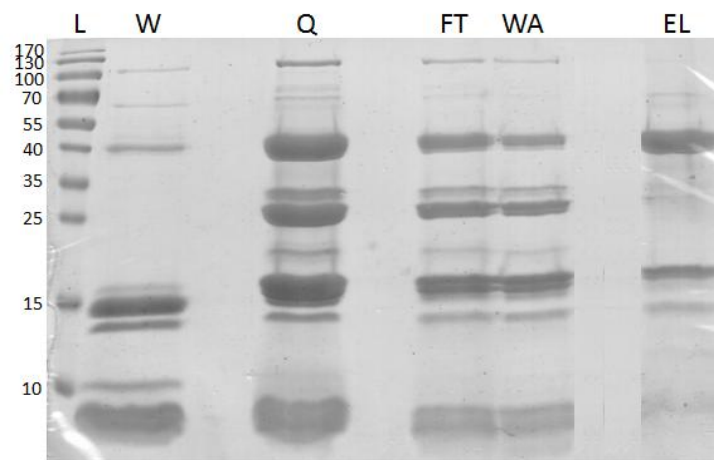

Figure 3: SDS-PAGE pattern of worker (W) and queen (Q) honeybee venom proteins. Queen venom was treated by the ProteoMiner technology. Flow-through (FT), wash (WA) and elution (EL) of this pre-treatment are shown.

The honeybee was the first hymenopteran species which genome was sequenced (2006; [20]). Later, also the genomes of three *Nasonia* species (2010; [5]) and seven ant species (2010-2011; [21-26]) have been sequenced. While the honeybee and *N. vitripennis* genomes were sequenced by Sanger sequencing methods, other genomes were sequenced using next-generation sequencing approaches. Next-generation sequencing has dramatically reduced costs in producing high-quality draft genomes. Therefore, several additional hymenopteran genomes were recently sequenced and many more will follow in the (near) future. Recently, re-sequencing the honeybee genome improved the genome assembly and allowed to increase the gene set by about 50% [27]. In addition, the genomes of the bumblebees *B. terrestris* and *B. impatiens* were sequenced [28]. In the near future, two other honeybee species, *A. dorsata* and *A. florea*, will be added to the list of hymenopteran

species with a sequenced genome. Moreover, since March 2011, the i5k initiative was announced, which will aim to sequence the genomes of 5000 insects and other arthropods. At present, more than 800 species have been nominated, including 276 Hymenoptera, and 58 genomes have already been sequenced (http://arthropodgenomes.org/wiki/i5K). Also multiple stinging bees, wasps and ants are nominated. As the available gene prediction sets facilitate the identification of venom proteins in mass spectrometry studies, sequencing the genomes of additional venomous hymenopterans will give a boost to the venom proteome research.

2. HYMENOPTERA VENOM ALLERGY

This PhD work identified several novel honeybee venom allergen candidates. Although PVF1 and C1q IgE reactivity was demonstrated with sera of honeybee venom allergic patients in ELISA, both compounds were unable to activate basophils in basophil activation tests (BAT) (Chapter 5). Consequently, both compounds cannot be incorporated in the official IUIS list of allergens as this requires both IgE recognition and in vitro cell activation or positive reactions in skin tests. Additional experiments are required to further elucidate the nature of this discrepancy between both tests. First, the specificity of IgE binding in the ELISA assay can be controlled by pre-incubating the serum samples with the protein, which should reduce IgE reactivity in the ELISA assay. Second, there may be several reasons for negative results in BATs. In case both compounds contain only one IgE epitope, ELISAs show a positive signal, while basophils may not be activated as this requires cross-linking of the FcεRI receptors on the basophil cell surface. However, it has been demonstrated that many allergens possess only one IgE epitope, but due to homodimerization on cell surface-bound antibodies they are able to activate basophils [29]. Dimerization would be very common and essential for many allergens. PVF1 proteins are reported in literature to be homodimers [30]. Also the honeybee venom PVF1 sequence contains the typical eight cysteines involved in dimerization. Non-denaturing SDS-PAGE separation should reveal if also the PVF1 recombinant exists in dimeric form, however this is often concentration- and/or pH-dependent. Alternatively, relative IgE reactivity to each of the proteins is too low to be detected by the BAT or both proteins may exert inhibitory functions on basophil activation

that neutralize the activation signal provided by FcεRI cross-linking. The latter can be tested by adding the proteins to activated basophils in BATs.

Also testing the IgE reactivity of several novel Api m 10 protein variants by protein array technology identified two novel allergen candidates (Chapter 4). The experiment revealed that variant 3 and 4 are present in the venom and that they may possess unique IgE epitopes. Therefore, the allergenicity of these variants should be further examined using BATs.

Component-resolved diagnosis (CRD) becomes an important approach for distinguishing between different Hymenoptera venom allergies. This PhD work provides insights in the honeybee and bumblebee venom proteome, and in antigenicity of several honeybee venom compounds, which is the fundament to further improve CRD for Hymenoptera venom allergy. First, an Ag5-like gene was found to be solely expressed by the honeybee venom glands during the winter months when bees stay inside the hive (Chapter 1). In addition, we demonstrated that it was absent in the venom during the summer months when bees actively forage and come in contact with humans (Chapter 2). Also phospholipase A_1 was found to be absent from honeybee venom (Chapter 2). Both phospholipase A_1 and antigen 5 proteins are major allergens of the venom of many wasp and ant species. Therefore, the absence of similar allergens in honeybee venom allows to distinguish between honeybee and wasp/ant venom allergy using CRD. Our findings are in favor of the currently applied CRD which aims to allow a differential diagnosis between venom allergies caused by the honeybee, *A. mellifera*, and wasp, *V. vulgaris*, as these are the most prevalent stinging insects in many European countries. The wasp-specific venom allergens Ves v 1 and Ves v 5, which correspond to the phospholipase A_1 and antigen 5 proteins respectively, and the honeybee-specific venom allergen Api m 1 are currently commercially available to establish a correct diagnosis. However, while Ves v 1 and Ves v 5 allow to diagnose between 92 and 96% of the wasp venom allergic patients [31-33], CRD for honeybee venom allergy solely based on Api m 1 lacks sensitivity [31;32;34-36]. Several studies have shown that adding additional honeybee venom allergens can increase sensitivity of CRD for honeybee venom allergy. Sturm and co-workers [35] described that ImmunoCAP assays with rApi m 1, rApi m 2 and nApi m 4 diagnosed honeybee venom allergy in 82.5% of the patients (n=40). Hofmann and co-workers [32] found that the combination of the same allergens led to a positive result in

89% of the patients in a larger study population (n=82). In addition, a brand new study [37] describes that the use of six honeybee venom allergens (rApi m 1, rApi m 2, rApi m 3, nApi m 4, rApi m 5, rApi m 10) increases diagnostic sensitivity to 94.4% in a population of 144 honeybee venom allergic patients. However, as Api m 2 and Api m 5 have cross-reactive homologues in the wasp venom, including both allergens decreases specificity of CRD. This study also reports that in case only the honeybee venom-specific allergens Api m 1, Api m 3, Api m 4 or Api m 10 were used, positive results were obtained in 89.6% of the honeybee venom allergic patients. Therefore, research should further focus on the identification of novel honeybee venom-specific allergens which provide both sufficient sensitivity for CRD of honeybee venom allergy and sufficient specificity to distinguish honeybee and wasp venom allergy. This PhD work showed that C1q and PVF1 were IgE recognized by respectively 24 and 19 out of 72 sera of honeybee venom allergic patients (Chapter 5). About 40% of the analyzed sera of honeybee venom allergic patients lacked IgE reactivity to rApi m 1. Sensitivity was increased to 87.5% using four honeybee venom-specific allergens, rApi m 1, rApi m 3, nApi m 4 and rApi m 10. We demonstrated that also adding C1q and PVF1 can further increase CRD sensitivity with 8.3%, reaching a sensitivity above 95%. Further studies should reveal if wasp venom contains cross-reactive C1q and PVF1 homologs.

In addition, this PhD work investigated the effects of Api m 10 protein heterogeneity on IgE recognition. As about 50% of the honeybee venom allergic patients shows IgE reactivity towards the immunologically characterized Api m 10 variant 2, this allergen is important for increasing CRD sensitivity. Our array-based experiment showed that Api m 10 protein heterogeneity has important consequences for diagnostic tests, as IgE recognition is both isoform- and patient-specific (Chapter 4). Variant 2 was previously demonstrated to be a good biomarker for Api m 10 IgE recognition [38], which was confirmed by the present study. In addition, we found that two additional variants, variant 3 and especially variant 4, may be of particular relevance for the diagnosis of honeybee venom allergy in those patients that are allergic to honeybee venom but who do not react to variant 2. However, since all of the analyzed variant 2 non-reactive sera displaying IgE reactivity to variant 3 and variant 4 also showed IgE reactivity to Api m 1, these Api m 10 variants did not increase the sensitivity of CRD.

Finally, this PhD work offers a long list of potential new honeybee venom allergens which can even further increase CRD sensitivity. However, the presently used insect cell

expression system is labor-intensive and time-consuming, making it impossible to produce all novel compounds as recombinants for immunological characterization. Unfortunately, at present no efficient high-throughput insect cell expression system exists, which enables the production of a high number of high quality venom proteins. Therefore, an efficient pre-screening strategy should be developed to select a group of expected venom allergens, which can then be produced to experimentally determine IgE reactivity and basophil activation. Several bioinformatics approaches have been developed to predict allergens. AlgPred (http://www.imtech.res.in/raghava/algpred/index.html) has integrated these approaches to predict allergenic proteins with high accuracy. However, according to this software, none of the experimentally confirmed honeybee venom allergens are predicted to be allergens. Therefore, we conclude that allergen prediction software is presently unreliable for selecting potential novel venom allergens. A second strategy is the use of immunoblots to screen for novel IgE recognized compounds. In a preliminary experiment, a CPLL-treated honeybee venom sample was separated by 1D-SDS-PAGE, blotted and incubated with three sera of honeybee venom allergic patients without CCD reactivity (Figure 4). As a CPLL-treated venom sample was also separated by SDS-PAGE in our mass spectrometry study (Chapter 2), we were able to identify which proteins are present within the IgE detected bands. Unfortunately, a broad list of proteins was identified in each of the IgE-detected bands, including multiple already known allergens. This makes it impossible to conclude if unknown allergens contribute to the observed IgE recognition. Moreover, many enriched proteins from the CPLL elution fraction have a similar molecular weight of 40 kDa (Figure 4). Over 30 venom compounds were found in this IgE recognized band, including most allergens. Therefore, the CPLL elution fraction should be separated by 2D electrophoresis, separating proteins with an identical MW by isoelectric point. IgE recognition can then be tested by immunoblotting using a well-selected set of sera of honeybee venom allergic patients. Interesting sera would be those with high IgE titers to honeybee venom, but low IgE titers for already known honeybee venom allergens. As many venom compounds carry CCDs, also CCD-specific IgE titers should be absent. A 2D-gel run in duplicate allows to cut out spots for protein identification by mass spectrometry.

So far, only the venom extract of the bumblebee *B. terrestris* is commercially available for allergy diagnosis. Due to the reported high cross-reactivity between honeybee and
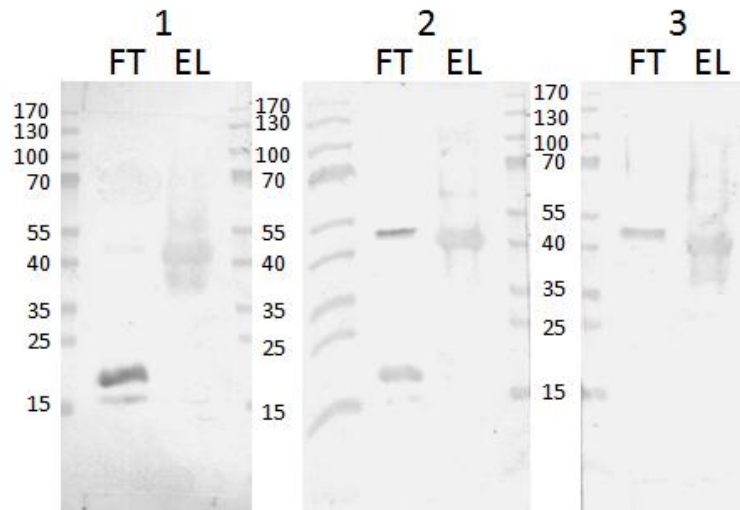
173

Figure 4: Immunoblots of combinatorial peptide ligand library (CPLL)-treated honeybee venom with sera of three honeybee venom allergic patients. FT= flow-through fraction of the CPLL, EL= elution fraction of the CPLL. Molecular weights of the markers are presented. ImmunoCAP IgE titers of sera can be found in Table S1.

bumblebee venoms, a diagnostic test with high specificity should be developed which allows to distinguish between both venom allergies. However, as so far only very few insights in the panel of cross-reactive and species-specific honeybee and bumblebee venom allergens have been obtained, it is currently impossible to develop a diagnostic test which allows to make this distinction. Our proteomic analyses revealed bumblebee venom homologues for all honeybee venom allergens (Chapter 3), except for Api m 6 (not found in the *B. terrestris* venom proteome although a putative homologue is found in its genome). It would be interesting to see if these compounds also represent important bumblebee venom allergens. Vice versa, further research should reveal if the CLIP serine protease from honeybee venom has allergenic properties, as it is homologous to the Bom t 4 allergen. However, as only very few sera of bumblebee venom allergic patients are available for research purposes, thoroughly analyzing the allergenic properties of individual bumblebee venom compounds will be difficult to complete. Therefore, an immunoblot study with 2D-separated bumblebee venom and few available sera of bumblebee venom allergic patients may make it possible to get insights in the bumblebee venom allergen repertoire. Later, further studies may unravel the nature of the immunological cross-reactivity between honeybee and bumblebee venom.

Genomes of several stinging hymenopterans will become available in the future, which will enable to explore the venom proteomes of these species. This information will eventually provide deeper insights in the venom allergen repertoire and allow to increase the efficacy of allergy diagnosis. As only few patients recognize the hymenopteran species that stung, diagnostic tests should be developed which allow to distinguish between allergies to venoms of different species present in a specific geographical region. Compared to singleplex test, multiplex tests allow to simultaneously analyze IgE reactivity towards a broad panel of species-specific allergens using only a limited amount of serum. Therefore, the protein array technology may play an important role in the future allergy diagnosis.

3. ADDENDUM

The supplementary table can be found on the included CD-ROM or can be requested by e-mail from matthiasvanvaerenbergh@hotmail.com and Dirk.deGraaf@UGent.be.

Table S1 presents ImmunoCAP IgE titers of three sera of honeybee venom allergic patients used in the preliminary immunoblot experiment (see Figure 4).

4. ACKNOWLEDGEMENTS

Special thanks to Olivier Christiaens for generating the phylogenetic tree (Figure 1).

5. REFERENCES

[1]   Wong ESW, Belov K. Venom evolution through gene duplications. Gene 2012 Mar 15;496(1):1-7.
[2]   Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, et al. The Toxicogenomic Multiverse: Convergent Recruitment of Proteins Into Animal Venoms. Annual Review of Genomics and Human Genetics 2009;10:483-511.
[3]   Cardinal S, Danforth BN. The Antiquity and Evolutionary History of Social Behavior in Bees. PLoS ONE 2011 Jun 13;6(6):e21086.
[4]   de Graaf DC, Aerts M, Brunain M, Desjardins CA, Jacobs FJ, Werren JH, et al. Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies. Insect Mol Biol 2010 Feb;19 Suppl 1:11-26.
[5]   Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, et al. Functional and Evolutionary Insights from the Genomes of Three Parasitoid Nasonia Species. Science 2010 Jan 15;327(5963):343-8.

[6] Peiren N, Vanrobaeys F, de Graaf DC, Devreese B, Van BJ, Jacobs FJ. The protein composition of honeybee venom reconsidered by a proteomic approach. Biochim Biophys Acta 2005 Aug 31;1752(1):1-5.

[7] de Graaf DC, Brunain M, Scharlaken B, Peiren N, Devreese B, Ebo DG, et al. Two novel proteins expressed by the venom glands of *Apis mellifera* and *Nasonia vitripennis* share an ancient C1q-like domain. Insect Mol Biol 2010 Feb;19 Suppl 1:1-10.

[8] Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: A primer. Mass Spectrometry Reviews 1998 Jan;17(1):1-35.

[9] Perry RH, Cooks RG, Noll RJ. Orbitrap Mass Spectrometry: Instrumentation, Ion Motion and Applications. Mass Spectrometry Reviews 2008 Nov;27(6):661-99.

[10] Resende VMF, Vasilj A, Santos KS, Palma MS, Shevchenko A. Proteome and phosphoproteome of Africanized and European honeybee venoms. Proteomics. In press.

[11] Ovchinnikov YA, Miroshnikov AI, Kudelin AB, Kostina MB, Boikov VA, Gotgilf IM. Structure and presynaptic activity of tertiapin-neurotoxin from bee venom *Apis mellifera*. Bioorganicheskaya Khimiya 1980;6(3):359-65.

[12] Peiren N, de Graaf DC, Vanrobaeys F, Danneels EL, Devreese B, Van BJ, et al. Proteomic analysis of the honey bee worker venom gland focusing on the mechanisms of protection against tissue damage. Toxicon 2008 Jul;52(1):72-83.

[13] Albert S, Klaudiny J. MRJP9, an ancient protein of honeybee MRJP family with non-nutritional function. Journal of Apicultural Research 2007;46(2):99-104.

[14] Hummon AB, Richmond TA, Verleyen P, Baggerman G, Huybrechts J, Ewing MA, et al. From the genome to the proteome: Uncovering peptides in the Apis brain. Science 2006 Oct 27;314(5799):647-9.

[15] Lowy PH, Sarmiento L, Mitchell HK. Polypeptides minimine and melittin from bee venom: effects on *Drosophila*. Arch Biochem Biophys 1971 Jul;145(1):338-43.

[16] Vick JA, Shipman WH, Brooks R, Jr. Beta adrenergic and anti-arrhythmic effects of cardiopep, a newly isolated substance from whole bee venom. Toxicon 1974 Mar;12(2):139-44.

[17] Shkenderov S, Koburova K. Adolapin--a newly isolated analgetic and anti-inflammatory polypeptide from bee venom. Toxicon 1982;20(1):317-21.

[18] Ferreira RS, Sciani JM, Marques-Porto R, Lourenco A, Orsi RD, Barraviera B, et al. Africanized honey bee (*Apis mellifera*) venom profiling: Seasonal variation of melittin and phospholipase A(2) levels. Toxicon 2010 Sep 1;56(3):355-62.

[19] Fox JW, Serrano SMT. Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures. Proteomics 2008 Feb;8(4):909-20.

[20] Weinstock GM, Robinson GE, Gibbs RA, Worley KC, Evans JD, Maleszka R, et al. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 2006 Oct 26;443(7114):931-49.

[21] Bonasio R, Zhang GJ, Ye CY, Mutti NS, Fang XD, Qin N, et al. Genomic Comparison of the Ants *Camponotus floridanus* and *Harpegnathos saltator*. Science 2010 Aug 27;329(5995):1068-71.

[22] Nygaard S, Zhang GJ, Schiott M, Li C, Wurm Y, Hu HF, et al. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. Genome Research 2011 Aug;21(8):1339-48.

[23] Smith CD, Zimin A, Holt C, Abouheif E, Benton R, Cash E, et al. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). Proceedings of the National Academy of Sciences of the United States of America 2011 Apr 5;108(14):5673-8.

[24] Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, et al. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. Proceedings of the National Academy of Sciences of the United States of America 2011 Apr 5;108(14):5667-72.

[25] Suen G, Teiling C, Li L, Holt C, Abouheif E, Bornberg-Bauer E, et al. The Genome Sequence of the Leaf-Cutter Ant *Atta cephalotes* Reveals Insights into Its Obligate Symbiotic Lifestyle. PLoS Genetics 2011 Feb;7(2):e1002007.

[26] Wurm Y, Wang J, Riba-Grognuz O, Corona M, Nygaard S, Hunt BG, et al. The genome of the fire ant *Solenopsis invicta*. Proceedings of the National Academy of Sciences of the United States of America 2011 Apr 5;108(14):5679-84.

[27] Honey bee Genome Sequencing and Analysis Consortium. Finding the missing honey bee genes: lessons learned from a genome upgrade. Unpublished.

[28] International Bumblebee Genomics Consortium. Two bumblebee genomes show the route to advanced social living. Unpublished.

[29] Rouvinen J, Janis J, Laukkanen ML, Jylha S, Niemi M, Paivinen T, et al. Transient Dimers of Allergens. PLoS ONE 2010 Feb 5;5(2):e9037.

[30] Tarsitano M, De Falco S, Colonna V, Mcghee JD, Persico MG. The C. elegans pvf-1 gene encodes a PDGF/VEGF-like factor able to bind mammalian VEGF receptors and to induce angiogenesis. Faseb Journal 2006 Feb;20(2):227-33.

[31] Korosec P, Valenta R, Mittermann I, Celesnik N, Silar M, Zidarn M, et al. High sensitivity of CAP-FEIA rVes v 5 and rVes v 1 for diagnosis of Vespula venom allergy. Journal of Allergy and Clinical Immunology 2012 May;129(5):1406-8.

[32] Hofmann SC, Pfender N, Weckesser S, Blank S, Huss-Marp J, Spillner E, et al. Detection of IgE to recombinant Api m 1 and rVes v 5 is valuable but not sufficient to distinguish bee from wasp venom allergy Reply. Journal of Allergy and Clinical Immunology 2011 Jul;128(1):248.

[33] Ebo DG, Faber M, Sabato V, Leysen J, Bridts CH, De Clerck LS. Component-resolved diagnosis of wasp (yellow jacket) venom allergy. Clinical and Experimental Allergy 2013 Feb;43(2):255-61.

[34] Hofmann SC, Pfender N, Weckesser S, Huss-Marp J, Jakob T. Added value of IgE detection to rApi m 1 and rVes v 5 in patients with Hymenoptera venom allergy. Journal of Allergy and Clinical Immunology 2011 Jan;127(1):265-7.

[35] Sturm GJ, Hemmer W, Hawranek T, Lang R, Ollert M, Spillner E, et al. Detection of IgE to recombinant Api m 1 and rVes v 5 is valuable but not sufficient to distinguish bee from wasp venom allergy. Journal of Allergy and Clinical Immunology 2011 Jul;128(1):247-8.

[36] Jakob T, Köhler J, Blank S, Huss-Marp J, Spillner E, Lidholmh J. Comparable IgE-reactivity to nApi m 1 and rApi m 1 in CCD negative bee venom allergic patients. Journal of Allergy and Clinical Immunology 2012;130:276-8.

[37] Köhler J, Blank S, Müller S, Bantleon F, Frick M, Huss-Marp J, et al. Component resolution reveals additional major allergens in bee venom allergic patients. Journal of Allergy and Clinical Immunology. In press.

[38] Blank S, Seismann H, Michel Y, McIntyre M, Cifuentes L, Braren I, et al. Api m 10, a genuine *A. mellifera* venom allergen, is clinically relevant but underrepresented in therapeutic extracts. Allergy 2011 Oct;66(10):1322-9.

Summary

Honeybees and bumblebees defend the hive against predators and external threats using venom, which contains several toxic compounds that cause death in other insects or inflict pain in higher organisms. Besides, in man, early exposure to bee venom evokes IgG1, IgG2 and to a lesser extent IgG4 antibody responses, whereas long-term exposure often found in beekeepers drives the immunity to an IgG4 type of humoral response. Allergy to a bee sting is mediated by IgE antibodies and, so far, 12 honeybee and 2 bumblebee venom allergens have been listed by the International Union of Immunological Societies (http://www.allergen.org/Allergen.aspx).

This PhD thesis consists of two main parts. The first part focused on further unraveling the venom composition of the honeybee (*A. mellifera*) and bumblebee (*B. terrestris*). Several gel-based proteomics studies conducted in the past suggested the existence of unknown venom compounds in the honeybee venom proteome. Also a genome mining study conducted in 2006 identified multiple novel genes encoding for putative venom constituents. Moreover, recently the honeybee genome was re-sequenced using next-generation sequencing technologies and improved gene prediction sets became available which may include novel venom genes. Therefore, this PhD work tried to obtain deeper insights in the honeybee worker venom composition by integrating genome, transcriptome and proteome information.

To overcome the issues of gel-based proteomics, this PhD work analyzed the honeybee worker venom composition by liquid chromatography-mass spectrometry. This analysis confirmed the presence of nine honeybee venom compounds, including two peptides which remained undetected in preceding gel-based studies due to their low molecular weight. In addition, a novel venom peptide with antimicrobial properties, called apidaecin, was found. In the second proteomic study, the honeybee venom proteome was investigated using a combinatorial peptide ligand library sample pretreatment to enrich for minor components, followed by shotgun LC-FT-ICR MS analysis. This strategy revealed an

unexpectedly rich venom composition: in total 102 proteins and peptides were found, with 83 of them never described in bee venom samples before.

Also genome and venom gland transcriptome data were used during this PhD to obtain insights in the honeybee venom composition. Genome mining revealed a list of compounds with resemblance to known insect allergens or venom toxins, one of which showed homology to proteins of the antigen 5 (Ag5)/Sol i 3 cluster. We also demonstrated that the honeybee Ag5-like gene is expressed by the venom gland tissue of winter bees but not of summer bees. Further proteomic experiments should confirm the presence of the Ag5-like protein in the venom of winter bees. In addition, this PhD work obtained evidence for transcript heterogeneity of known venom compounds, as three alternative splice variants of PVF1 and 9 novel chimeric variants of icarapin were found to be expressed by the venom glands. To the best of our knowledge, this is the first report of the identification of chimeric transcripts generated by the honeybee.

So far, only little research has focused on bumblebee venom. Recently, the genome sequence of the European large earth bumblebee (*Bombus terrestris*) became available and this allowed the first in-depth proteomic analysis of its venom composition. We identified 57 compounds, with 52 of them never described in bumblebee venom. Remarkably, 72% of the detected compounds were found to have a honeybee venom homologue, which reflects the similar defensive function of both venoms and the high degree of homology between both genomes. However, both venoms contain a selection of species-specific toxins, revealing distinct damaging effects that may have evolved in response to species-specific attackers.

The second part of this PhD work involved the immunological implications of the venom proteome. A first analysis revealed the lack of IgG4 recognition of both apidaecin and Ag5-like protein by beekeepers' sera. In case of the antimicrobial peptide apidaecin, a low immunogenicity can be explained by its short length. For the Ag5-like protein, its restricted expression in winter time certainly lowers the exposure to this venom compound significantly, as beekeepers are then hardly stung. Second, this PhD work identified several novel honeybee venom allergen candidates. Our ELISA assay showed that C1q and PVF1 are IgE recognized by respectively 1/3 and 1/4 of the honeybee venom allergic patients. However, both compounds were unable to activate basophils in basophil activation tests, which requires further investigation. Also, a protein array experiment showed that IgE

recognition of a panel of icarapin isoforms is both isoform- and patient-specific. Moreover, two novel icarapin isoforms represent interesting allergen candidates as they may possess unique IgE epitopes. Therefore, their allergenicity needs to be further investigated.

In addition, this PhD thesis provides insights which aid to further improve component-resolved diagnosis (CRD) for Hymenoptera venom allergy. Our proteomic analysis confirmed that phospholipase A_1 (PLA_1) and Ag5 proteins are absent from the venom of honeybees during summer months. As both are major allergens of the venom of many wasp and ant species, our findings are in favor of the currently applied CRD which aims to allow a differential diagnosis between honeybee (*A. mellifera*) and wasp (*V. vulgaris*) venom allergy, the most prevalent stinging insects in many European countries. The wasp-specific venom allergens Ves v 1 and Ves v 5, which correspond to the phospholipase A_1 and antigen 5 proteins respectively, and the honeybee-specific venom allergen Api m 1 are currently used to establish a correct diagnosis. However, while Ves v 1 and Ves v 5 allow to diagnose more than 95% of the wasp venom allergic patients, CRD for honeybee venom allergy solely based on Api m 1 lacks sensitivity. Our findings indicate that adding C1q and PVF1 to a panel of honeybee venom-specific allergens (Api m 1, Api m 3, Api m 4 and Api m 10) can increase CRD sensitivity with 8.3%, reaching a sensitivity above 95%. In addition, our novel insights in the honeybee venom proteome offer a long list of potential new honeybee venom allergens which can increase CRD sensitivity even further. Finally, our in-depth proteomic analysis of the bumblebee venom composition is the fundament for unraveling its allergen repertoire and the immunological cross-reactivity between honeybee and bumblebee venom. This may allow to develop a CRD strategy to distinguish between honeybee and bumblebee venom allergy.

Samenvatting

Het gif van de honingbij en de hommel is samengesteld uit verschillende toxische componenten. Deze zijn dodelijk voor andere insecten en induceren pijn in hogere organismen, wat hen toelaat om zichzelf en de kast te beschermen tegen ongenode gasten. Hiernaast veroorzaakt een steek bij sommige personen een IgG1, IgG2 en in mindere mate IgG4 immuunrespons. Bij personen die frequent worden gestoken, zoals imkers, evolueert dit vaak naar een IgG4 humorale respons. Hiernaast kan zich ook een allergie ontwikkelen die wordt gemedieerd door IgE herkenning van specifieke gifallergenen. Momenteel worden 12 honingbij en 2 hommel gifallergenen erkend door de 'International Union of Immunological Societies' (http://www.allergen.org/Allergen.aspx).

Deze thesis is opgedeeld in twee delen. In het eerste deel werd getracht de gifsamenstelling van de honingbij (*A. mellifera*) en aardhommel (*B. terrestris*) verder te ontrafelen. In het verleden werden reeds verschillende proteoomstudies uitgevoerd gebaseerd op 2D-PAGE en deze wezen op de aanwezigheid van ongekende gifcomponenten in het honingbij gifproteoom. Ook bij een honingbij 'genoom-mining' studie die reeds werd uitgevoerd in 2006 werden verschillende nieuwe genen gevonden die mogelijks coderen voor gifcomponenten. Bovendien werd recent het genoom van de honingbij opnieuw gesequeneerd via next-generation sequencing en werden nieuwe verbeterde genpredictie datasets beschikbaar gesteld die mogelijks nieuwe gifgenen bevatten. Aldus werd in deze thesis getracht diepere inzichten te verkrijgen in de gifsamenstelling van de honingbij werkster door integratie van informatie afkomstig van het genoom, transcriptoom en proteoom.

Om de tekortkomingen van proteoomstudies gebaseerd op 2D-PAGE te omzeilen, werd de gifsamenstelling van de honingbij werkster geanalyseerd via vloeistofchromatografie-massaspectrometrie. Deze analyse bevestigde de aanwezigheid van 9 honingbij gifcomponenten, inclusief 2 peptiden die door hun laag moleculair gewicht niet werden gedetecteerd in voorgaande 2D-PAGE-gebaseerde proteoomstudies. Bovendien

werd apidaecine, een peptide met antimicrobiële eigenschappen, voor het eerst gedetecteerd in het gif. In de tweede proteoomstudie werd het gifstaal behandeld met een bibliotheek van combinatoriële peptide liganden die toelaat laag abundante componenten aan te rijken, waarna een 'shotgun' LC-FT-ICR massaspectrometrie analyse werd uitgevoerd. Deze strategie onthulde een rijke gifsamenstelling: in totaal werden 102 proteïnen en peptiden geïdentificeerd, waarvan er 83 componenten nog nooit werden beschreven in bijengif.

Ook werden genoom- en gifklier transcriptoomdata gebruikt om inzichten te verwerven in de gifsamenstelling van de honingbij. Via 'genome mining' werd een aantal componenten gevonden die gelijkenissen vertonen met gekende insect allergenen of gif toxines. Eén hiervan vertoont homologie met proteïnen van de antigen 5 (Ag5)/Sol i 3 cluster. We toonden aan dat het honingbij Ag5-like gen wordt geëxpresseerd door de gifklieren van winterbijen, maar niet door die van zomerbijen. Bijkomend proteoomonderzoek moet de aanwezigheid van het Ag5-like proteïne in het gif van winterbijen bevestigen. Deze thesis identificeerde ook nieuwe transcript varianten van verschillende gekende bijengif componenten: zowel 3 alternatief gesplicete PVF1 varianten als 9 chimerische icarapine varianten worden geproduceerd door de gifklier. Deze thesis beschrijft voor het eerst de identificatie van chimerische varianten geproduceerd door de honingbij.

Totnogtoe werd het gif van de hommel slechts in beperkte mate onderzocht. Recent werd echter het genoom van de Europese aardhommel (*Bombus terrestris*) gesequeneerd en dit liet toe de gifsamenstelling via een diepgaande proteoomanalyse te onderzoeken. We identificeerden 57 componenten, waarvan er 52 nooit werden beschreven in hommelgif. 72% van de gedetecteerde proteïnen en peptiden bleken een homoloog te hebben in het gif van de honingbij, wat kan worden verklaard door de sterk vergelijkbare functie van beide giffen en de hoge homologie van beide genomen. Beide giffen bevatten echter een selectie van species-specifieke toxines wat wijst op een verschillende werking van beide giffen, mogelijks geëvolueerd in respons op species-specifieke vijanden.

In het tweede deel van deze thesis werd het immunologisch belang van het gifproteoom onderzocht. Een eerste analyse toonde aan dat zowel apidaecine als het Ag5-like proteïne niet worden herkend door serum-IgG4 antilichamen van imkers immuun tegen bijengif. In het geval van apidaecine kan dit mogelijks worden verklaard door een lage immunogeniciteit

van dit klein peptide, terwijl het Ag5-like proteïne enkel blijkt te worden geëxpresseerd tijdens de wintermaanden wanneer imkers nauwelijks worden gestoken waardoor geen immuunrespons wordt opgebouwd tegenover deze component. Ten tweede werden verschillende nieuwe kandidaat gifallergenen geïdentificeerd. Via een ELISA assay toonden we IgE herkenning aan van C1q en PVF1 door respectievelijk 1/3 en 1/4 van de bijengif allergische patiënten. Beide componenten bleken echter niet in staat om basofielen te activeren in basofiel activatie tests, wat verder moet worden onderzocht. Ook bleek uit een protein array experiment dat IgE herkenning van 11 icarapine isovormen zowel isovorm- als patiënt-specifiek is. Bovendien zijn twee van deze icarapine isovormen interessante kandidaat allergenen, aangezien ze mogelijks unieke IgE epitopen bevatten. Aldus moet hun allergeniciteit verder worden onderzocht.

Deze thesis verstrekt ook inzichten die kunnen bijdragen aan een hogere efficiëntie van de 'component-resolved diagnose' (CRD) voor Hymenoptera gifallergie. Onze proteoomanalyse bevestigde dat phospholipase A_1 (PLA_1) en Ag5 proteïnen niet voorkomen in het gif van honingbijen tijden de zomermaanden. Aangezien beide componenten belangrijke allergenen zijn in het gif van wespen en mieren, laat dit toe om via CRD een onderscheid te maken tussen allergieën veroorzaakt door bijengif dan wel wespen- of mierengif. Onze bevindingen ondersteunen de CRD die hedendaags wordt toegepast om een onderscheid te maken tussen gifallergieën veroorzaakt door de honingbij (*A. mellifera*) en wesp (*V. vulgaris*), die verantwoordelijk zijn voor het merendeel van de Hymenoptera steken doorheen Europa. De wesp-specifieke gifallergenen Ves v 1 en Ves v 5, respectievelijk PLA_1 en Ag5, en het honingbij-specifiek gifallergeen Api m 1 worden hedendaags gebruikt om een correcte diagnose te stellen. Via de combinatie van Ves v 1 en Ves v 5 kan een wespengif allergie worden vastgesteld in meer dan 95% van de wespengif allergische patiënten. De sensitiviteit van CRD voor honingbij gifallergie enkel gebruik makend van Api m 1 is momenteel echter ontoereikend. Via deze thesis kon worden aangetoond dat deze sensitiviteit kan toenemen met 8.3% door het toevoegen van C1q en PVF1 aan een selectie van bijengif-specifieke allergenen (Api m 1, Api m 3, Api m 4, Api m 10). Bovendien wordt via de combinatie van deze 6 gifcomponenten een sensitiviteit van meer dan 95% bereikt. Hiernaast werd via ons bijengif proteoomonderzoek een lange lijst met nieuwe potentiële gifallergenen gegenereerd die de CRD sensitiviteit verder kunnen verhogen. Ten slotte vormt onze diepgaande analyse van het hommelgif proteoom de basis voor een verdere

ontrafeling van het hommelgif allergeen repertoire en de immunologische kruisreactiviteit tussen bijengif en hommelgif. Verder kan dit toelaten een CRD strategie te ontwikkelen die toelaat het onderscheid te maken tussen allergieën veroorzaakt door bijengif en hommelgif.