

**PROTEOMIC ANALYSIS OF *APIS MELLIFERA* LARVAL  
IMMUNE RESPONSE AGAINST *PAENIBACILLUS*  
*LARVAE***

by

Queenie Wai Tsz Chan

B.Sc., University of British Columbia, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA  
(Vancouver)

January 2010

© Queenie Wai Tsz Chan, 2010

## ABSTRACT

The honey bee *Apis mellifera* is an extremely beneficial insect due to its role in pollination and honey production. Honey bees are vulnerable to many diseases, one of which is caused by the bacterium *Paenibacillus larvae*. Ingestion of its spores by honey bees early in their larval stage results in death before adulthood. This disease is known as American foulbrood. Interestingly, this deadly consequence does not occur when older larvae or adults consume the spores, a phenomenon which we hypothesize to be the result of an underdeveloped immune system in young larvae. To address this issue, we mainly employed mass spectrometry-based proteomics techniques to learn about the protein effectors that are present in the host. We focused on the hemolymph (insect blood) where a significant repertoire of immunity-related proteins exists. In our comparison of adults and larvae, we saw age-correlated differences in the levels of the antimicrobial peptide hymenoptaecin, the phenoloxidase enzyme which lead to the production of cytotoxic free radicals, and several bacterial recognition proteins. This prompted a detailed study of the larval development in which many new expression trends were revealed; for example, the age-related decrease of antioxidant proteins, proteins associated with translational machinery, and enzymes related to protein turnover. Unexpectedly, most immunity-related proteins showed no significant age correlation, except the antimicrobial peptide apisimin and phenoloxidase. The latter protein was particularly intriguing, given that its increased expression and activity with respect to age agreed with the buildup of resistance against *P. larvae*. Furthermore, this protein was upregulated in *P. larvae* infected larvae compared to healthy controls. Taken together, this enzyme may be an important factor of the host immune response. The decrease of nutrient storage proteins in the diseased state also implied that the defense response was mounted with an associated energetic cost. Data from the work covered in this thesis has helped explain, from a molecular point-of-view, the pathogenesis of American foulbrood. Although our focus was on honey bee immunity, the proteomics-based approach taken here has provided protein evidence and expression data that will serve as a resource for the wider scientific community.

# TABLE OF CONTENTS

<b>Abstract</b> .....	<b>ii</b>
<b>Table of contents</b> .....	<b>iii</b>
<b>List of tables</b> .....	<b>vii</b>
<b>List of figures</b> .....	<b>viii</b>
<b>List of abbreviations</b> .....	<b>x</b>
<b>Acknowledgements</b> .....	<b>xii</b>
<b>Dedication</b> .....	<b>xiii</b>
<b>Co-authorship statement</b> .....	<b>xiv</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Rationale for honey bee research .....	4
1.1.1 Major honey bee diseases .....	4
1.1.1.1 American Foulbrood and other bacterial diseases .....	5
1.1.2 Research aims and hypothesis.....	7
1.2 Developmental biology of honey bee larvae .....	7
1.2.1 Basic anatomy .....	8
1.2.2 Nutrition .....	8
1.2.3 Proteins in developing larvae.....	10
1.2.3.1 Juvenile hormone.....	10
1.2.3.2 Insect storage proteins .....	10
1.2.3.3 Major royal jelly proteins .....	11
1.2.4 Current status of honey bee larvae molecular biology research.....	12
1.3 Biochemistry of honey bee immunity .....	13
1.3.1 Cell-mediated immune responses .....	14
1.3.1.1 Phagocytosis.....	15
1.3.1.2 Nodulation .....	15
1.3.1.3 Encapsulation.....	15
1.3.2 Humoral responses .....	16
1.3.2.1 Pathogen recognition.....	16
1.3.2.2 Toll pathway .....	17
1.3.2.3 Immune deficiency pathway .....	18

1.3.2.4	Antimicrobial peptides .....	18
1.3.2.5	Lysozyme .....	20
1.3.2.6	Complement .....	21
1.3.2.7	Melanization.....	22
1.4	Experimental approach: mass spectrometry-based proteomics.....	24
1.4.1	Tandem mass spectrometry .....	25
1.4.1.1	Orbitrap.....	27
1.4.1.2	Fourier transform ion cyclotron resonance .....	28
1.4.1.3	Linear trapping quadrupole.....	29
1.4.2	Database searching .....	29
1.4.3	Relative protein quantitation .....	30
1.4.3.1	Peptide correlation profiling/ion intensities.....	30
1.4.3.2	Isotopic labeling.....	31
1.4.4	Functional assignment .....	32
1.5	References .....	33
<b>2</b>	<b>Quantitative comparison of caste differences in honey bee hemolymph.</b>	<b>41</b>
2.1	Abstract.....	41
2.2	Introduction.....	42
2.3	Experimental Procedures .....	43
2.3.1	Materials.....	43
2.3.2	Hemolymph collection.....	44
2.3.3	Electrophoresis.....	45
2.3.4	Proteolytic digestion .....	45
2.3.5	Liquid chromatography tandem mass spectrometry .....	45
2.3.6	Analysis of mass spectrometry data .....	46
2.3.7	Gene ontology.....	48
2.4	Results .....	48
2.4.1	Hemolymph extraction .....	48
2.4.2	Qualitative analysis of caste and stage differences in hemolymph proteins	50
2.4.3	Honey bee biomarkers.....	54
2.4.4	Gene ontology.....	55
2.4.5	Quantitative differences in hemolymph .....	58

2.4.6	Unassigned spectra .....	61
2.5	Discussion .....	62
2.6	Conclusions .....	66
2.7	References .....	67
<b>3</b>	<b>Protein profiles during honey bee larval development .....</b>	<b>72</b>
3.1	Abstract.....	72
3.2	Introduction.....	72
3.3	Experimental Procedures .....	73
3.3.1	Reagents .....	73
3.3.2	Obtaining larvae of known ages .....	74
3.3.3	Protein collection.....	74
3.3.4	Denaturing protein gel electrophoresis.....	75
3.3.5	Sample preparation for mass spectrometry analysis.....	75
3.3.6	Raw data processing.....	76
3.3.7	Automated protein annotation to Gene Ontology terms.....	77
3.3.8	Semi-automated protein annotation and manual categorization.....	77
3.3.9	Statistical analysis .....	78
3.3.10	Comparison to <i>Drosophila</i> .....	78
3.4	Results .....	79
3.5	Discussion .....	94
3.6	Conclusions .....	97
3.7	References .....	98
<b>4</b>	<b>The innate immune and systemic response in honey bees to a bacterial pathogen, <i>Paenibacillus larvae</i>.....</b>	<b>101</b>
4.1	Abstract.....	101
4.2	Introduction.....	102
4.3	Experimental Procedures .....	103
4.3.1	Honey bees and infection experiments.....	103
4.3.2	Sample collection and processing for MS.....	103
4.3.3	MS data analysis.....	104
4.3.4	Hemolymph collection for PO activity assay .....	104
4.3.5	Hemolymph fractionation .....	105

4.3.6	In solution phenoloxidase assay .....	105
4.4	Results .....	106
4.4.1	Different strains of P. larvae produce equivalent outcomes.....	106
4.4.2	Diseased honey bee larvae express higher levels of mitochondrial metabolic enzymes.....	108
4.4.3	Infected larvae deplete their energy stores during infection .....	110
4.4.4	The protein-folding/quality control machinery is over-expressed in response to infection .....	113
4.4.5	Lysozyme and hymenoptaecin levels increase with bacterial challenge....	114
4.4.6	Phenoloxidase activity is not found in the first two days of larval development, but increases sharply afterwards.....	116
4.5	Discussion .....	119
4.6	Conclusions .....	121
4.7	References .....	122
<b>5</b>	<b>Conclusion .....</b>	<b>124</b>
5.1	Addressing the project aims.....	124
5.2	Addressing the project hypothesis.....	126
5.3	Other contributions to honey bee knowledge .....	127
5.4	Future directions .....	129
5.4.1	Phenoloxidase knockdown study.....	129
5.4.2	Treatments and prevention against AFB.....	129
5.4.3	PeptideAtlas and inclusion of novel proteins .....	129
5.4.4	Comparison to transcriptomics data .....	130
5.5	Closing .....	130
5.6	References .....	131
	<b>Appendix A.....</b>	<b>132</b>

## LIST OF TABLES

Table 3.1 Using peptide ratios to derive relative protein expression: an example .....	82
Table 3.2 Expression trends of proteins categorized under Gene Ontology terms.....	84
Table 3.3 Expression trends of manually annotated and categorized proteins .....	86
Table 3.4 Enrichment analysis of tissue protein functional classes following hierarchical clustering .....	87
Table 3.5 Enrichment analysis of hemolymph protein functional classes following hierarchical clustering.....	88
Table 3.6 Protein expression levels during larval development: honey bee versus fruit fly larvae .....	93

## LIST OF FIGURES

Figure 1.1 The three castes of adult honey bees.....	1
Figure 1.2 Haplodiploid sex determination .....	2
Figure 1.3 Honey bee larval development.....	8
Figure 1.4 Insect immune responses.....	14
Figure 1.5 Conversion of DOPA to melanin .....	22
Figure 1.6 General experimental scheme used for MS-based proteomics .....	25
Figure 1.7 A simplified schematic of the LTQ-Orbitrap.....	27
Figure 1.8 A simplified schematic of the LTQ-FT .....	27
Figure 2.1 Hemolymph extraction from an adult honey bee.....	49
Figure 2.2 Hemolymph extraction from a honey bee larva.....	49
Figure 2.3 Polyacrylamide gel electrophoresis of honey bee hemolymph .....	51
Figure 2.4 Schematic of data analysis procedures following hemolymph digestion .....	52
Figure 2.5 XITandem versus Mascot.....	53
Figure 2.6 Number of proteins in adult hemolymph.....	54
Figure 2.7 Fragment spectrum of a peptide derived from a DMV polyprotein .....	55
Figure 2.8 Protein categorization by Molecular Function GO terms .....	56
Figure 2.9 Protein categorization by Biological Activity GO terms .....	57
Figure 2.10 Quantitation of protein expression comparing worker larvae and adults....	59
Figure 2.11 Quantitation of protein expression comparing adult castes .....	60
Figure 3.1 Polyacrylamide gel electrophoresis of honey bee larvae proteins.....	80
Figure 3.2 Histogram of heavy/light peptide ratios within a sample.....	81
Figure 3.3 Developmental changes of protein content in larval hemolymph .....	83
Figure 3.4 Average-linkage clustering of proteins quantified in the honey bee larvae ..	89
Figure 3.5 Expression profile of imaginal disc growth factor during larval development	91
Figure 3.6 Expression profile of odorant binding protein 14 during larval development	91
Figure 3.7 Expression profile of apisimin during larval development .....	92
Figure 4.1 Comparing two infection methods based on peptide ratios .....	107
Figure 4.2 Comparing two infection methods based on peptide ratios .....	108
Figure 4.3 American Foulbrood-induced changes in the levels of mitochondrial metabolic proteins.....	109

Figure 4.4 American Foulbrood-induced changes in the levels of food proteins .....	111
Figure 4.5 American Foulbrood-induced changes in the levels of energy storage proteins .....	112
Figure 4.6 American Foulbrood-induced changes in the levels of protein-folding chaperones.....	113
Figure 4.7 American Foulbrood-induced changes in the levels of proteins with immune function.....	115
Figure 4.8 Mass spectrometry-based peptide analysis for prophenoloxidase .....	116
Figure 4.9 Phenoloxidase activity in fractionated hemolymph.....	117
Figure 4.10 Method for relative peptide quantitation by mass spectrometry: an example .....	118
Figure 4.11 Phenoloxidase activity profile during larval development.....	119

## LIST OF ABBREVIATIONS

1DGE	one-dimensional gel electrophoresis
2DGE	two-dimensional gel electrophoresis
ALDH	aldehyde dehydrogenase
ATP	adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
cDNA	coding DNA
csd	complementary sex determiner
DAP	diaminopimelic
DNA	deoxyribonucleic acid
DOPA	3,4-dihydroxyphenylalanine
DWV	deformed wing virus
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tags
fem	feminizer
FTICR	Fourier transform ion cyclotron resonance
GO	Gene Ontology
HEX110	hexamerin 110
HEX70a	hexamerin 70a
HEX70b	hexamerin 70b
HEX70c	hexamerin 70c
HSP	heat-shock protein
IMD	Immune deficiency
JAK/STAT	Janus kinases/Signal Transducers and Activators of Transcription
LC	liquid-chromatography
LD50	lethal dose 50
LTQ	Linear trapping quadrupole
MALDI	matrix-assisted laser desorption ionization
MDH	malate dehydrogenase
MP	melanization protease
MRJP	major royal jelly protein

mRNA	messenger ribonucleic acid
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NPY	neuropeptide Y
OBP	odorant-binding protein
PBS	phosphate buffered saline
PGRP	peptidoglycan recognition protein
PO	phenoloxidase
ppm	parts-per-million
proPO	prophenoxidase
r.c.f.	relative centrifugal force
RJ	royal jelly
RNA	ribonucleic acid
RNAi	RNA interference
SCX	strong cation exchange
SDS	sodium dodecyl sulfate
serpin	serine protease inhibitors
STAGE	STop And Go Extraction
TCA	tricarboxylic acid
TEP	thioester-containing proteins
TNF-alpha	tumor necrosis factor
TOF	time of flight
VDV	varroa destructor virus

## **ACKNOWLEDGEMENTS**

I offer my gratitude to UBC, the faculty and staff, and my fellow students and colleagues who have allowed me to learn from them. I wish to thank my supervisor, Dr. L.J. Foster, who was not only concerned with developing me as a scientist, but also gave me the freedom to learn a great deal more that improved me as a person in general, such as the skills of leadership and communication. It is a precious thing to have a mentor that knows when to guide and when to trust.

I would also like to thank my supervisory committee: Dr. M. Roberge, Dr. E. Jan, and Dr. R. Beavis. I especially wish to thank Dr. Roberge for allowing me to learn so much in his laboratory even when I was only an undergraduate at the time, and all the thoughtful advice he has given me since then.

I thank my parents for their love, encouragement, and support. I also thank them for the bravery it took them to bring me to Canada, a country which is so very far from their own place of birth, and so very different from the culture in which they were raised. If I had not come to Canada, the country which I now consider my home, a place which supports the pursuit of knowledge and scientific research, I am sure I could not have walked the path to become a scientist.

I thank my partner and best friend Nelson for giving me happiness and a purpose to my life. It is his unwavering patience, love, and understanding that has given me the strength to tackle whatever difficulties I have faced in my life.

To my parents and Nelson.

## **CO-AUTHORSHIP STATEMENT**

Portions of this thesis involved the work of my co-authors.

The idea and design of the experiments was guided by my supervisor (Foster, L.J.). All the experimental work in this thesis were my own, except in Chapter 4 where my co-authors' (Melathopoulos, A.P., Pernal, S.F. and Foster, L.J.) role was in the design and set-up of the honey bee colonies from which the initial samples were obtained. One of my co-authors (Howes, C.G.) developed the database schema and its user interface that was used for online presentation of the data in Chapter 2. The data analysis was performed by myself and guided by my supervisor, who also wrote some of the bioinformatic software critical to the analysis. This thesis was written by myself with the editing help of my supervisor as well as my supervisory committee.

# 1 INTRODUCTION<sup>1</sup>

To the untrained eye, a “bee” generally refers to a six-legged flying insect that has alternating black and yellow strips on the abdomen, that is equipped with a stinger, and that collects honey. This description could describe almost any organism under the order *Hymenoptera*, including wasps and hornets; the true honey bee *Apis mellifera*, discussed in this thesis (and hereinafter may be referred to simply as “bee”) is under the superfamily *Apoidea*, shared only by bumblebees.

**Figure 1.1 The three castes of adult honey bees**



Honey bee adult of the three castes are shown: queen (left), worker (middle), drone (right). (Scale bar = 5mm.)

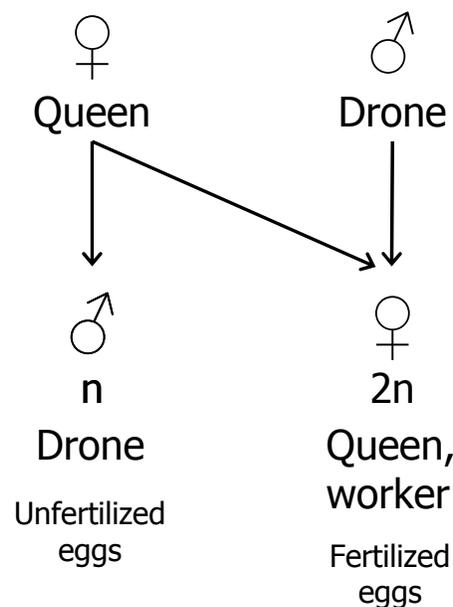
Honey bees are eusocial animals, placing them in the highest level of social organization where they are joined only by wasps, ants, termites (Wilson et al., 2005) and two species of mole rats (Holmes et al., 2007). The definition of eusociality features the presence of overlapping generations in a single colony, caring cooperatively for the young, division of labour among the sterile members, with reproductive tasks limited to a subset of individuals (Wilson et al., 2005). The division of labour in bees gives rise to three castes in each hive: queen, drone, and worker (Figure 1.1). In normal situations, a colony contains only a single fertile female, the queen. At a very early stage of her adult

---

<sup>1</sup> A version of this chapter has been accepted for publication. Chan, Q.W. and Foster, L.J. (2009) *Apis mellifera* proteomics: where will the future bee? *Current Proteomics*. 6(2):70-83.

life, she mates with typically one to two dozen males, the drones, after which they die immediately. The queen will lay up to 2000 eggs per day for the next one to three years, until she dies or her sperm stores are depleted (Winston, 1987). She can control the gender of her offspring, with females arising from fertilized diploid eggs and males from unfertilized haploid eggs: a sex-determination system called haplodiploidy (Beukeboom, 1995) (Figure 1.2). The males, having 16 chromosomes, are entirely hemizygous for every gene; females have 32 chromosomes.

**Figure 1.2 Haplodiploid sex determination**



Haplodiploidy dictates that the sex is determined by the number of sets of chromosomes an individual has. Drones receive only one set ( $n$ ) from unfertilized eggs and are therefore male. Those that receive two sets of chromosomes ( $2n$ ) are females, arising from diploid eggs. In a laboratory setting where highly inbred bees can be produced, diploid males are possible, indicating that sex determination is more complex than is outlined by haplodiploidy (Dzierzon, 1845).

It was Dzierzon who, in 1845, discovered that unmated queens, thus unable to fertilize her eggs, are only capable of producing males (Dzierzon, 1845). Sex determination in bees as understood today, however, involves at least two genes located in the Sex Determination Locus: complementary sex determiner (*csd*) and feminizer (*fem*). Through a series of RNAi experiments, it has been shown that knockdown of *csd* and *fem* in females results in sex transformation, causing male organs to be produced (Gempe et al., 2009). Highly inbred bee lines produced in the laboratory

can result in diploid male eggs, which if placed into a natural hive setting, are cannibalized by workers shortly after hatching. It is now known that the *csd* gene induces the female pathway early in embryogenesis, initiating the transcription of *fem*, which mediates its own positive feedback system that maintains the female phenotype throughout development (Gempe et al., 2009).

All eggs proceed through three developmental stages, spanning about three weeks in total, before reaching adulthood: three days as an egg, about one week as a larva during which its soft exoskeleton is shed (moults) to accommodate its increasing size in stages called instars at the rate of about once per day, and the remaining time as a pupa during which metamorphosis occurs within a sealed hexagonal wax cell. Adult workers perform all the necessary tasks involved in maintaining and protecting the hive, as well as providing food for all its members. These are all sterile female bees with very small ovaries compared to the queen; in rare situations when a colony becomes queenless, these sterile individuals are capable of laying unfertilized, haploid eggs that all become drones. Adult workers themselves can never become queens, so to replace a lost or under-performing queen a young female larva is moved to a special cup within the colony and fed a protein-rich diet of royal jelly (RJ). Workers secrete RJ when it is needed and its feeding to the young larva triggers a developmental switch that causes the larva to develop into a queen (Winston, 1987).

As with all insects, bee blood, called hemolymph, is distributed throughout the body in an open-circulatory system. Like mammalian blood, this fluid contains various nutrients, proteins, hormones and waste products from the organs and their individual cells; yet perhaps unlike mammalian blood, the protein content of hemolymph varies drastically according to age, sex, and season, implying that the study of fluid is a potentially rich source of information with regards to the individual's physiological state. While the anatomy for the majority of the bee organs had been described in great detail already in the 1920s (Nelson, 1924, Snodgrass, 1925), their physiology was explored much later and many gaps in knowledge remain to be filled. Nearly nothing is known about bee organs at the molecular biology level.

## 1.1 Rationale for honey bee research

Humans' fascination with bees began even before recorded history, as seen from cave paintings depicting bee hives (Clark et al., 1999), and bees being worshipped as goddesses by the ancient Greeks (Cook, 1895). Medicinal uses of bee products, known as apitherapy, has been practiced globally (Clark et al., 1999) and is still under active research. Based on fossil records, it appears that honey bees have been a separate species for more than 80 million years (Michener et al., 1988); as a testament of their importance to the ecosystem, not only have they evaded extinction, but we, as humans, are increasingly dependent on them for pollination (Aizen et al., 2008). Bees are constantly under threat due to combined damage from bacteria, parasites, viruses, and pesticides, insecticides, and artificial bee food (Aliouane et al., 2008, Cox-Foster et al., 2007, Higes et al., 2008, Oldroyd, 2007, Stokstad, 2007, Stokstad, 2007); this is no more evident than in the case of 'Colony Collapse Disorder', which resulted in United States beekeeping operations losing an average of 40% of their colonies, and in extreme cases as much as 90% (Cox-Foster et al., 2007). Currently, the cause of this disorder is not clear and remains under active investigation. Bee research is not only necessary to help ensure that bees remain available to help us but their unique behavior and biology, and as such we still have much to learn.

### 1.1.1 Major honey bee diseases

The honey bee is an excellent model of disease control in high population densities, where individuals are extremely social and have frequent transfer of body fluids: tropholaxis (mouth-to-mouth food sharing) occurs frequently between adults, and adults feed glandular secretions of food to immature bees in the larval stage, during which they are immotile. Because of their unique hive environment, honey bees are susceptible to a number of species-specific pathogens, many of which are asymptomatic unless the host is already under stress from other diseases (Brødsgaard et al., 2000). This is particularly true of viral infections, with Israel acute paralysis virus being an excellent example: this virus has been linked to Colony Collapse Disorder (Chen et al., 2007, Cox-Foster et al., 2007), yet there is mounting evidence that fungal infection by *Nosema ceranae* may also be related (Higes et al., 2008), in addition to possible threats from bacteria, parasites, viruses, and pesticides, insecticides, and artificial bee food

(Aliouane et al., 2008, Cox-Foster et al., 2007, Higes et al., 2008, Oldroyd, 2007, Stokstad, 2007, Stokstad, 2007). Other very common viruses include sacbrood virus, deformed wing virus (DWV), and Kashmir bee virus, all of which have positive-sense, single-stranded RNA genomes (Chen et al., 2006). In recent years, significant losses can also be attributed to the parasitic mite *Varroa destructor*, which feeds on the hemolymph of immature bees (Sammataro et al., 2000), and to a lesser extent *Acarapis woodi* (McMullan et al., 2006), which resides the trachea of adult bees. Fungal infections that are most problematic are *N. ceranae* and the more recent growing threat of *N. apis*. Especially in Canada, major losses (34% over the winter of 2008-2009) have been attributed to *N. ceranae* and *V. destructor*, adding to background conditions of starvation and weather (Pernal, 2009).

#### 1.1.1.1 American Foulbrood and other bacterial diseases

The viral, fungal, and parasitic diseases mentioned above share one common effect on the host: diseased adult bees may live shorter and less productive lives. The most common bacterial infections, however, are lethal at the larval stage. With adult production stalled, hive population dwindles quickly. American foulbrood, caused by the endospore-forming, Gram-positive bacterium *Paenibacillus larvae* (Shimanuki, 1997), was first noted to be widespread in the United States in the beginning of the last century (White, 1906). The infectious spores, which contaminate the food stores in the hive, cannot be cleaned by honey bees or human intervention. Consequently, infected hives must be incinerated as the only practical means of bacteria removal, and as a preventative measure against disease spreading to neighbouring hives, which can result in major economic losses. The less pervasive and less severe European foulbrood, caused by *Mellisococcus pluton* (Govan et al., 1998, Sanford, 2003), has a similar effect on the honey bee host, but will not be discussed further as *P. larvae* is the focus of this work.

Due to its economic and agricultural impact, American foulbrood is an active area of research. Despite significant ongoing efforts for more than 100 years, the mechanism of infection has only begun to reveal itself; the full picture, especially at the molecular biological level, remains far from complete. In the larval gut, the spores germinate into a rod-shaped vegetative state. It was initially thought that they breach the gut epithelium by phagocytosis (Davidson, 1973, Gregorc et al., 1998), and proliferate in bulk

hemolymph, causing death by systemic infection (Bailey et al., 1991). However, recent studies (Yue et al., 2008) showed that proliferation occurs in the midgut and bacteria enter the hemocoel by a paracellular route, between cells of the midgut epithelium. There have been screens for proteinaceous virulence factors secreted by *P. larvae* which have suggested proteases (Antunez et al., 2009, Dancer et al., 1997), however it has not been clear if these factors are the definitive players.

As of today, a mystery about American foulbrood that has yet to be solved and is also perhaps the most interesting: despite the lethality of infection by *P. larvae*, bees are in fact almost completely resistant to its effects for most of their lives, except for an approximately 48-hour window after egg-hatching. The LD50 for susceptible larvae is as low as 8.49 spores (Brødsgaard et al., 1998), while older larvae and adults are resistant (Bamrick et al., 1961, Brødsgaard et al., 1998, Hoage et al., 1966, Shimanuki, 1997). Theories to explain this phenomenon include physical and biochemical barriers. For example, there is an age-dependent increase in thickness of the peritrophic matrix that envelops incoming food and which may help contain the bacteria in the midgut (Bamrick, 1964) to be later excreted. This has been counter-argued by the observation that this membrane is present as early as 8 hours after hatching, and thus cannot completely explain the susceptibility of larva at this age (Davidson, 1973). Indeed, substances extracted from larval homogenates have the ability to inhibit the growth of *P. larvae*, suggesting that resistance is through chemical or biochemical, rather than physical means. Interestingly, homogenates of larvae that are 2-5 days old were found to be most inhibitory against bacterial growth (Wedenig et al., 2003), roughly mirroring the emergence of *P. larvae* resistance in the natural environment. A qualitative investigation of infected bee larval hemocytes (blood cells) found shifts in the numbers and percentages of cells with certain morphologies, suggesting cellular response to the infection (Zakaria, 2007). Two gene expression studies from the same authors, each exploring different ages and infection methods, show that in young larvae, a very small number of genes respond to *P. larvae* challenge. However, the results were not consistent between the studies (i.e. opposite regulation trends for a gene encoding for abaecin, an antimicrobial peptide), and most of the genes with putative roles in immunity in fact showed a reduction in gene expression (Evans, 2004, Evans et al., 2006). Furthermore, it is not known whether the transcript levels correlate with

biosynthesis of the corresponding enzymes. How *P. larvae* infect a bee within a narrow time window to induce a fatal consequence remains an unanswered question.

### **1.1.2 Research aims and hypothesis**

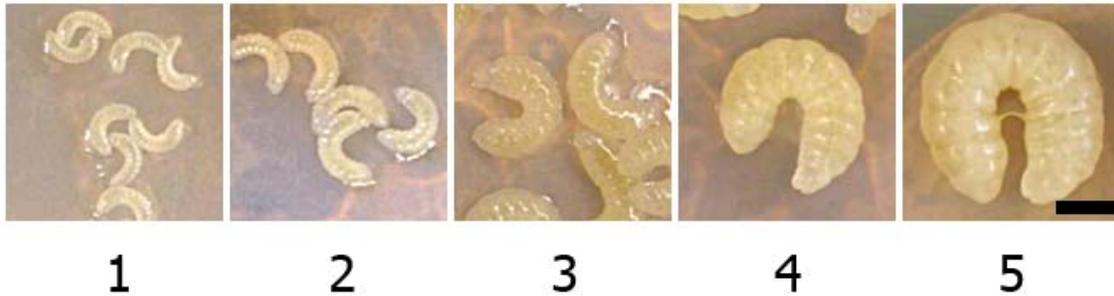
The goal of this project is to understand the molecular basis of young honey bee larvae vulnerability to *P. larvae* infection, as well as the mechanism of resistance found in older larvae. Our hypothesis that the lethal susceptibility of very young larvae to this bacterium is the result of an underdeveloped immune system, manifested in the absence or reduced expression of key immunity proteins combined with their insufficient targeting to sites of high susceptibility to pathogen invasion.

This dissertation involves the investigation of proteins related to both immunity and larval development, with a focus on those that are pertinent to our hypothesis and research aims.

## **1.2 Developmental biology of honey bee larvae**

There are four stages of development in the life of a honey bee: egg, larva, pupa, and adult. The pre-adult stages are sometimes collectively called "brood". Apart from feeding and expanding in size, the larva does little else; on the molecular level, however, several major changes that take place imply the presence of intricate biochemical controls at work. For example, dietary levels of protein in the form of major royal jelly proteins (MRJPs) within two days of eclosion determine whether a larva is destined to be a queen or a worker, due to epigenetic controls (Elango et al., 2009). Huge metabolic demands and a corresponding increase in efficiency of protein synthesis allow a body mass to increase 900-fold (in the case of workers) within this five- to six-day-long developmental stage (Figure 1.3) to about 140 mg, and growing from a 2.7 mm egg to nearly its adult size of 1.7 cm (Winston, 1987).

**Figure 1.3 Honey bee larval development**



Larvae of the honey bee shown for five consecutive days from eclosion (egg-hatching). (Scale bar = 2mm).

### **1.2.1 Basic anatomy**

The overall anatomy of larvae does not change drastically over this five- to six-day long phase, despite dramatic size changes. It has a simple body structure, containing mostly the organs essential for feeding and digestion, namely the mouth and midgut, with the latter taking up the majority of space within the body cavity. Excretory tubules and a hindgut (rectum) are present but are not connected to the midgut until the end of larval development, which is about the time when feeding stops (Snodgrass, 1925). They have an exoskeleton, which is white and soft relative to that of adults, and is shed about once a day during this stage of rapid growth to reflect its increasing size. With the matured silk glands, bees begin to spin a cocoon in the last half of larval development. Despite having no legs, they are able to wiggle and uncurl from the C-shaped position taken up earlier in larval life, so that their heads point towards the cap of their hexagonal wax cells. By the fifth or sixth day from eclosion, their cell is sealed with wax by an adult worker bee.

### **1.2.2 Nutrition**

The basic building blocks of honey bees are derived from nectar and pollen, providing carbohydrates and proteins, respectively. Workers take up nectar in their foregut (crop) and redeposit the solution into an individual cell in the honeycomb; in doing so, they also add enzymes from their hypopharyngeal glands. These enzymes are either involved in the breakdown of sugars or confer anti-bacterial properties. Pollen grains, which are the male germ plasm of plants, are treated with phytocidal acid before

storage to prevent germination and bacterial spoilage. Bees, which are not equipped with teeth, are presumed to have enzymes in their gut for breakdown of the pollen wall.

Larvae, having no legs or other means of transport, are completely dependent on adults for food. One report stated that, on average over the larval period, each brood is inspected 1926 times and fed during 143 visits (Lindauer, 1952); another group observed 7200 visits and 1140 feedings. Whichever may be the case, the frequency of contact between bees is a major factor for disease transmission (Kuwabara, 1947, Lineburg, 1924, Winston, 1987). Larvae are semi-immersed in a milky white, slightly viscous fluid, which is deposited by workers. This is a mixture containing clear secretion from the hypopharyngeal glands and a milky white substance from the mandibular gland of adult nurse bees. As the larva approaches the third instar, the white component decreases, and the food is directly mixed with increasing amounts of honey and pollen. Contents of glandular secretions have not been well documented, but it presumably contains vital growth factors, since to this day there is no completely artificial substitute for brood food: a formulation containing vitamins, minerals, and free amino acids supplemented with RJ extracts is able to support the development of normal-appearing adults (Shuel et al., 1986). Royal jelly is the milky white substance rich in proteins and lipids that is fed to honey bees during larval development. Larvae, prior to the third instar, can become a queen or worker depending on the amount of RJ fed: higher amounts will generate queens and lower amounts will produce workers. It is also used to feed adult foragers (Crailsheim, 2004). RJ has been shown to have many different enzymatic properties, including ascorbinoxidase, amylase, invertase, catalase, acid phosphatase, and more (Albert et al., 1999). Its reputation as a health food product comes from numerous reports, with data to indicate that it decreases blood pressure (Sultana et al., 2008), lowers triglyceride and very-low-density lipoprotein levels (Guo et al., 2007), and inhibits or induces T-cell proliferation depending on the concentration or component used (Gasic et al., 2007), to name a few. Due to the commercial value of this bee product, it is not surprising that RJ was the topic of the first bee proteomics publication (Scarselli et al., 2005) and several subsequent studies (reviewed in (Chan et al., 2009)). Despite this research, the value of human consumption of RJ is still not clear. Though it contains many biological activities that may benefit human health, for bees it

is generally thought of as a source of building blocks for growth and helps keep away infectious diseases.

### **1.2.3 Proteins in developing larvae**

Major proteins that are highly consequential to honey bee larval development are discussed.

#### **1.2.3.1 Juvenile hormone**

Juvenile hormone is a small molecule produced in an organ found on the sides of the esophagus in larva and adults, called the corpora allata (Winston, 1987). Of the several subtypes found in insects, the honey bee expresses only juvenile hormone III (Hagenguth et al., 1978). As the name suggests, it maintains larvae in the juvenile state by preventing metamorphosis: when the levels are sufficiently high in an individual it remains a larva. By degradation of juvenile hormone with juvenile hormone esterase or juvenile hormone epoxide hydrolase, such that the hormone levels drop below the threshold, the larva develops to the pupa stage. This temporal gradient of the hormone controls the level of proteins and biological activities associated with larval maturation. One example is the abundance of the major hemolymph protein hexamerin 70b, seen to be tightly negatively correlated with juvenile hormone levels (Cunha et al., 2005) in the larval stage. In Chapter 3 of this thesis, many time-sensitive expression trends were found for specific proteins (in some cases entire protein families) that are likely induced by changing juvenile hormone levels.

Closely associated with juvenile hormone in the control of development is 20-hydroxyecdysone (or ecdysterone). It regulates larval moulting, or ecdysis, the shedding of insect exoskeleton to accommodate its expanding size. Between moults, it coordinates a set transcription factors that allow the reabsorption of old cuticle and the production of new cuticle (Riddiford et al., 1999).

#### **1.2.3.2 Insect storage proteins**

During the honey bee pupal or metamorphosis stage, adults seal the cells containing these developing individuals using wax. Despite the metabolic demands of growing legs, wings, eyes, and all other adult organs (many of which are non-existent in the larval form), the pupa are not fed during this time. This process is fueled by the

large amounts of lipid accumulated during the larval stage and the amino acids necessary for this massive biosynthetic activity are kept in the form of storage proteins called hexamerins (Levenbook et al., 1984), which are composed of six identical or similar subunits. The fat body, which is an organ scattered throughout the body, is the major source of hexamerins. This organ undergoes two phases during larval development: first, hexamerin synthesis and subsequent release into the hemolymph, and second, reuptake of this protein by receptor-mediated endocytosis before transition into the pupal stage (de Oliveira et al., 2003).

Before protein sequencing became commonplace, hexamerins were named according to their approximate molecular weight as observed in polyacrylamide gel electrophoresis experiments. The nomenclature of these subunits was more unified after it was discovered that several distinct bands were actually derived from the same transcript. Today there are four known genes encoding hexamerin subunits: hexamerin 110 (HEX110) (Bitondi et al., 2006, Danty et al., 1998), hexamerin 70a (HEX70a) (Martins et al., 2008), hexamerin 70b (HEX70b), and hexamerin 70c (HEX70c) (less studied). Subunits HEX70b and HEX110 are highly abundant during larval development and are significantly decreased in pupa and adults. These two are best studied, likely because they are abundant enough to be detectable by earlier methodologies, but the less understood HEX70a and HEX70c are generally expected to have similar expression patterns. Of these, HEX70a is the only subunit documented to be prevalent also in adult hemolymph, suggesting that they have a more specialized role.

Hexamerin expression is highly regulated: their levels in the hemolymph increase sharply in the last two days of larval development, and decrease quickly almost immediately upon entry into the pupa stage. This cessation of hexamerin expression is linked with juvenile hormone decrease, an event that signals the larva-pupa transition.

### 1.2.3.3 Major royal jelly proteins

Royal jelly, as discussed previously, contains a number of different biological activities. However, 90% of its protein content is made up of the MRJP protein family (Schmitzova et al., 1998). They are produced mainly in the hypopharyngeal glands of young adult workers as a form of nutrient to feed immature bees. All of them have N-terminal regions that direct their secretion (Schmitzova et al., 1998). Currently there are eight known MRJPs, with MRJP1-4 being the most abundant, and therefore better

studied; MRJP5-8 are relatively recent discoveries. Three of these in particular – MRJP1, MRJP2, MRJP5 – have large amounts of the ten essential amino acids (Schmitzova et al., 1998). Though their role in nutrition is clear, the more recent discovery that their genes are also expressed in the brain (Garcia et al., 2009, Kucharski et al., 1998) suggests that they might play a far more complex, context-specific role in bees (Drapeau et al., 2006) than previously thought. The unusually high polymorphism and variable numbers of repeating units in MRJP3 that is not seen in other MRJPs further mystifies their functional capabilities (Albert et al., 1999). Two-dimensional gel electrophoresis (2DGE) patterns of RJ and the subsequent proteomic identification suggests that MRJPs are highly phosphorylated and glycosylated (Furusawa et al., 2008).

#### **1.2.4 Current status of honey bee larvae molecular biology research**

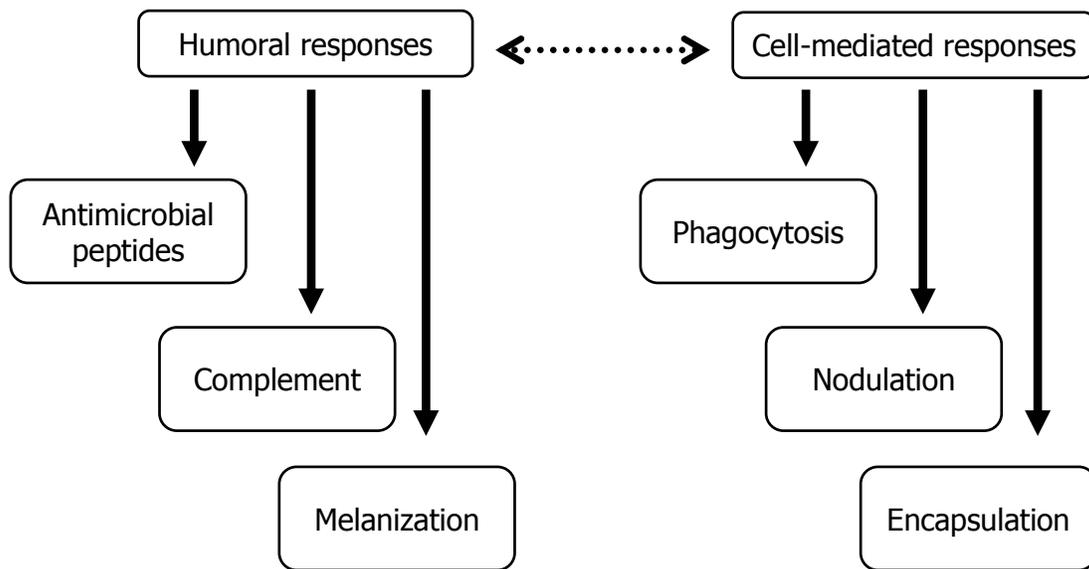
Perhaps due to their outward appearance of being an inert and unresponsive organism, doing little more than just eating, studies on larval development as compared to other stages of development have been relatively unexplored. However, the unique phenomenon of queen-worker caste determination that occurs in young bee larvae has drawn attention for many decades. It has long been known that nutrition is related to this, and it is now clear that DNA methylation is responsible for silencing certain genes that generate a queen instead of a worker (Kucharski et al., 2008). Though it is not clear which ones, it is apparent that the expression levels of many genes are affected (Evans et al., 2001). In terms of longitudinal research on gene or protein expression purely on the developmental aspect of larvae, a few highly abundant proteins have been examined as outlined in Section 1.2.3, and immunity-related proteins are discussed in Section 1.3. In 2007, twenty-two proteins of the whole larval proteome were quantified with respect to age, reporting up-regulation of proteins for carbohydrate metabolism, down-regulation of heat shock proteins, and an increase of an imaginal disc growth factor, among other findings (Li et al., 2007). Our understanding of genes and proteins expressed during this developmental period has barely begun, and with this lack of basic knowledge, it is difficult to make advances in the investigation of larval diseases such as American foulbrood. This is a major area which this thesis hopes to address.

### 1.3 Biochemistry of honey bee immunity

Bees, like other insects, are equipped with an innate immune system. They have no adaptive immunity, so they do not have lymphocytes, or any of their subtypes such as eosinophils and neutrophils and they cannot synthesize antibodies. As a result, it is not suitable to treat mammalian and insect immunity as entirely analogous, although some aspects are similar, e.g., Toll and Toll-like receptors. Instead of mapping pathways, the honey bee colony presents an interesting opportunity to study disease progression in a community environment, in contrast for example to *D. melanogaster* which is a solitary species. However, the social nature of bees and their high population density within a hive presents complications for research: genes for hygienic behaviour (Lapidge et al., 2002) and queen bee mating promiscuity (Seeley et al., 2007) have been linked to disease resistance, factors which must be considered in experiment design, albeit with great difficulty.

Given the high pathogen load conferred by their eusocial lifestyle, one might expect that bees have more immunity genes than solitary insects; however, comparative genomics tells us that bees have significantly fewer of these genes than *Anopheles* and *Drosophila* (Consortium, 2006, Evans et al., 2006). Bees' unique living environment, relatively clean food sources such as honey (Mundo et al., 2004), and hygienic behaviors (Lapidge et al., 2002) mean that they may face fewer and more niche-specific pathogens, and thus require fewer defenses; contrast this to the fruit fly, which feeds notoriously on rotting food and other wastes and thus may be driven to evolve more immunity genes.

Insect innate immunity can be divided into cellular and humoral responses, though the categorization is only an artificial one, since the regulation of both are tightly linked. Encapsulation, nodulation, and phagocytosis are considered part of the cellular response. These aspects are generally less well understood than humoral responses, which include antimicrobial peptides, complement, and melanization. This is especially true with regards to antimicrobial peptides, where their expression is regulated by the Toll and Immune Deficiency pathways. Most of these pathways have been elucidated in *Drosophila* (extensively reviewed in (Lemaitre et al., 2007)), and are believed to exist in the honey bee as well (Evans et al., 2006).



**Figure 1.4 Insect immune responses**

The immune system of insects can be generally divided into humoral and cell-mediated responses. Humoral responses consist of factors that function in the hemolymph, while cell-mediated responses are those requiring the direct action of host cells. Despite this categorization, the separation is only artificial since the responses of both are affected by one another.

### 1.3.1 Cell-mediated immune responses

Insect defense mechanisms of phagocytosis, nodulation, and encapsulation are typically categorized as cell-mediated immunity, since they all require physical action of the cell as a whole to deal directly with a microbial invader. Different pathogens will elicit different cellular responses, with varying efficiency, usually occurring within minutes of the initial exposure (Lavine et al., 2002). Hemolymph also contains complement-like insect host proteins, or lectins, which bind specific terminal sugar moieties present on certain bacterial walls and trigger hemocytes to act accordingly. In general, cellular immunity is far less studied than humoral immunity, due in part to the diversity of insect hemocytes types, making it difficult to isolate a homogeneous population for *in vitro* experiments. Naming of cells is also mostly based on morphological properties, causing the nomenclature to vary widely; contrast this to mammalian research where cell types can be defined by their binding specificity to certain antibodies. Judging from the single known article published in the past five years regarding the particular bee cell types in response to *P. larvae* (Zakaria, 2007), it is

apparent that the study of cell-mediated immunity in bees is an area that requires further investigation.

#### 1.3.1.1 Phagocytosis

Phagocytosis is hallmarked by a series of three events: recognition, engulfment, and destruction of pathogens, apoptotic cells, or infected cells. The mediators of bacterial phagocytosis are not well known, but it is known that the transmembrane proteins Eater (Kocks et al., 2005) and Nimrod (Kurucz et al., 2007), both containing epidermal growth factor-like repeats are necessary (Marmaras et al., 2009). It is possible that complement proteins binding to microbial surfaces will encourage their phagocytosis and subsequent destruction.

#### 1.3.1.2 Nodulation

Nodulation involves hemocytes forming aggregates around large numbers of bacteria, thus entrapping them and restricting their spread to other parts of the body (Lavine et al., 2002). This response is viewed as the predominant cellular immune response in insects (Gandhe et al., 2007). Interestingly, very little is known about this process; indeed, only in 2007 was the first protein directly associated with this process discovered. Using RNAi to suppress the expression of this protein, which was afterwards named Nodular, the number of nodules decreased significantly. During infection, its expression was upregulated within minutes of exposure to bacteria. Nodular appears to bind bacteria and yeast via their surface molecules such as lipopolysaccharide and beta-1,3 glucan. It was even shown to bind other hemocytes, which is reasonable given that nodulation is an accumulation of more than one layer of host hemocytes (Gandhe et al., 2007). This immune response is also known to require eicosanoid synthesis, as shown in the tobacco hornworm *Manduca sexta* (Miller et al., 1994). Nodules are often melanized, which occurs by a biochemical pathway that gives rise to reactive oxygen species that cause further damage to the microbes.

#### 1.3.1.3 Encapsulation

For pathogens such as parasites and nematodes, encapsulation occurs instead of nodulation, yet the two are similar. Hemocytes bind to the invader and form multilayer capsules to exclude from them from the circulation and kill them by asphyxiation.

Melanization also often occur, and the resulting free radical production will also inhibit the viability of invaders within the host.

### 1.3.2 Humoral responses

Humoral responses in insects consist of complement, antimicrobial peptides, and melanization. They are categorized together as such because their action against pathogens occurs in the hemolymph, despite their intracellular origin. Some humoral factors, such as complement and certain antimicrobial peptides, are found constitutively in healthy hemolymph as a first line of defense. Through the receptors located on the hemocytes surface, one or more pathways can be triggered, such as the Toll or Immune deficiency (IMD) pathway, which induces the upregulation of defensive mechanisms, most notably that of antimicrobial peptide synthesis. Melanization, which relies heavily on phenoloxidase, occurs most potently following apoptosis as this enzyme is released. There are few bee-specific reports of these humoral responses compared to the wealth of information regarding *Drosophila* immunity. It is generally assumed that bees and flies should be very similar in this regard, and grounds for this assumption have been presented thoroughly in comparative analyses such as (Evans et al., 2006). One major reason for the lagging research in honey bees is the lack of an immortalized honey bee cell line that is available in fruit flies: the Schneider 2 cell line isolated from late-stage embryonic cells (Schneider, 1972) exhibits macrophage-like properties and has been a useful cell type for insect immunity research. A cell line is a powerful tool with which gene mutations, RNA interference, transfections, and precisely controlled treatment conditions can be applied, attributing greatly to the deduction of biochemical pathways in cells.

#### 1.3.2.1 Pathogen recognition

Because of their open circulatory system, once a pathogen breaches physical barriers such as the exoskeleton or gut epithelium, it is likely to distribute very quickly within the entire body. The hemolymph contains a number of recognition molecules that, upon binding the appropriate pathogen-derived molecule, will trigger downstream host defense responses. To prevent autoimmunity, the molecules must not be similar to any that originate from the host: peptidoglycans, lipopolysaccharide, and flagellin are excellent examples. However, the majority of work has focused on the detection of

peptidoglycans by insect hosts, which are bound by the family of peptidoglycan recognition proteins (PGRPs). Mostly through research in *Drosophila*, a number of insect PGRPs are known. It is generally assumed that honey bees have a repertoire of similar-functioning enzymes, though gene count comparisons between the two species (Evans et al., 2006) clearly show that *Drosophila* have more of almost every gene type with putative roles in immunity. This is particularly true of the PGRP family. *Drosophila* has seven of the short-type (PGRP-S) and six of the long-type (PGRP-L), whereas *Apis* only has three and one, respectively. Since bees appear to have all or most of the pathways that act downstream of these PGRPs, it may be reasonable to assume that some of the PGRPs that flies have may be duplicate or serve very similar functions to another. From *Drosophila* studies, most of the PGRP-S and some of the PGRP-L are secreted proteins while other PGRP-Ls are single-pass transmembrane proteins. The specific functions of some are quite well-established, while those of many others are not known. Certain ones can bind lysine (Lys)-type peptidoglycans found on most Gram-positive bacteria, such as PGRP-SA and PGRP-SD, which activate the Toll signaling pathway, ultimately resulting the mobilization of defenses. Diaminopimelic (DAP)-type peptidoglycan, normally found on Gram-negative bacteria, is bound by PGRP-LC and PGRP-LE. The PGRPs of honey bees are named slightly differently in most cases (PGRP-LC, PGRP-S1, PGRP-S2, PGRP-S3) compared to their *Drosophila* orthologues but no research has been done directly on any of the bee PGRPs to confirm whether they perform similar functions to their *Drosophila* orthologues (Aggarwal et al., 2008, Pal et al., 2009).

Another group of receptors that bind pathogen-derived molecules are the Gram-negative bacteria binding protein (GNBP) family. Falsely named, but left unchanged for historical reasons, GNBP1 is known in *Drosophila* to detect, in combination with PGRP-SA, Lys-type peptidoglycan on most Gram-positive bacteria cell walls, and correspondingly, activate the Toll pathway (Pili-Floury et al., 2004). Also linked to this pathway is the detection of  $\beta$ -1,3-glucan from fungal cell walls by GNBP3. By homology, honey bees have only two GNBP, and both resemble GNBP1, and thus they have been named GNBP1-1 (GI:110755978) and GNBP1-2 (GI:254911140).

### 1.3.2.2 Toll pathway

The Toll pathway is activated by beta-1,3 glucan from yeast and fungi and Lys-type peptidoglycan, a major molecule on most Gram-positive bacteria. In this signaling

cascade, PGRP-SA and GGBP1 (Pili-Floury et al., 2004) or PGRP-SD in the hemolymph is believed to activate Grass, a serine protease. Grass cleaves the zymogen Spirit, which is also a serine protease. Activated Spirit cleaves the propeptide of the Spatzle-processing enzyme to activate it, which cleaves the proprotein Spatzle, leaving a C-terminus portion that binds the extracellular domain of Toll, a transmembrane receptor. Toll, when dimerized, recruits through its cytoplasmic portion, a pre-existing complex consisting of Myd88 and Tube. Together, the kinase Pelle is also recruited, causing it to phosphorylate Cactus. This post-translational modification of Cactus, whose homolog is the mammalian inhibitor of  $\kappa$ B protein, leads to its degradation through a ubiquitin-mediated pathway. Without Cactus to retain the proteins DIF and Dorsal in the cytoplasm, they enter the nucleus where they induce transcription of immune genes, including several antimicrobial peptides. The Toll pathway is also known to control the melanization pathway, which also plays an important role in immunity. This is discussed in Section 1.3.2.7.

### 1.3.2.3 Immune deficiency pathway

Defense against Gram-negative bacteria is mounted by a pathway called IMD, named after one of its proteins. The *Drosophila* IMD pathway is homologous to the mammalian tumor necrosis factor (TNF- $\alpha$ ) signaling pathway. The activation begins in the hemolymph, where a DAP-type peptidoglycan interacts with the host receptor which consists of the transmembrane protein PGRP-LC and secreted PGRP-LE. With the cytoplasmic domain of PGRP-LC, the adaptor protein IMD recruits the Fas-associated via death domain protein, then the death-related ced-3/Nedd2-like protein, followed by Relish. For recruitment to this complex, Relish must also have been already phosphorylated by the I $\kappa$ B kinase complex, which is activated by transforming growth factor-beta activated kinase 1. After this double chain of events, Relish, also known as nuclear factor- $\kappa$ B, is cleaved in the cytoplasm. Subsequently, the amino-terminus can enter the nucleus to induce the transcription of various genes, including antimicrobial peptides that act against Gram-negative bacteria.

### 1.3.2.4 Antimicrobial peptides

An important component of the innate immune system are antimicrobial peptides, which are a large class of small proteins (usually less than 100 amino acids in length)

that act on pathogens, usually bacteria, but also fungi and even some enveloped viruses (Diamond et al., 2009). Honey bees share some of the conserved peptides found also in *Drosophila* and even in most organisms, exerting bactericidal or bacteriostatic effects.

For insects, antimicrobial peptides circulate freely in the hemolymph, secreted by the major production source, the fat body (a liver-like organ), and smaller amounts by hemocytes and some epithelial cells (Bulet et al., 2003). Some of them are ubiquitously present in the hemolymph, regardless of infection, while others are only detected in appreciable amounts upon microbial challenge; some have wide and some have very narrow specificities. Most of the research surrounds the anti-bacterial activities of these peptides, though a few of them may also possess anti-fungal properties. Honey bees, likely due to their highly specialized and relatively clean living environment, have only six antimicrobial peptides, compared to the 20 which the notoriously organic waste-eating *Drosophila* express (Evans et al., 2006).

The antimicrobial peptide defensin, which is found in plants, insects, and mammals, is also seen in honey bees in two forms. Both are suspected to disrupt the bacterial membrane by weakening the permeability barrier, causing partial membrane depolarization and consequent reduction in ATP production, killing bacteria within a minute for most species (Bulet et al., 2005). It is thought to be mostly active against Gram-positive bacteria, and its expression is predicted to be downstream of the Toll pathway (Evans et al., 2006). One particularly well known defensin is royalisin (Bilikova et al., 2001), named for its original discovery in bee royal jelly (Fujiwara et al., 1990), currently renamed defensin 1 (NP\_001011616) (Klaudiny et al., 2005). Defensin 2 (NP\_001011638), distinctive because it is cysteine-rich, is generally different from defensin 1 except for a stretch of 47 residues over which they share 56% sequence identity (Altschul et al., 1990).

Another common class of antimicrobials is the proline-rich peptide family, distributed across many species of organisms, including bees. Apidaecin, only 18 amino acids long, is found in several isoforms in bees: 1a, 1b, II, and possibly a fourth one called III (Li et al., 2006). It is thought to be a slow-acting peptide against mostly Gram-negative strains (Casteels et al., 1989) with broad activity, but correspondingly low specificity. Its mode of action is not entirely known yet, but the hypothesis is that apidaecins cross the outer membrane, then binds irreversibly to a component of the

periplasmic space or inner membrane (Li et al., 2006). The active version of apidaecin is detectable in adults, while in the larvae it exists in an inactive, precursor form (Casteels-Josson et al., 1993). Another proline-rich antimicrobial peptide expressed by honey bees is related to the amino-terminal portion of apidaecins, called abaecin, which is a longer peptide (34 amino acids long) effective against a smaller number of Gram-positive and Gram-negative bacteria compared to the apidaecins. It appears to increase the permeability of the outer membrane, thus allowing lysozyme to weaken the cell wall through its glycosidase activity (Casteels et al., 1990).

It is not surprising, given the specialized pathogen niche found in the honey bee hive, there would be bee-specific antimicrobial peptides. Hymenoptaecin is one such protein: 93 amino acids in length, with a highly unique 2-pyrrolidone-5-carboxylic acid at the amino-terminus. It is bactericidal against some bacteria of both Gram stains. By testing on *E. coli*, it was concluded that the mode of action was sequential permeabilization of the outer and inner membrane (Casteels et al., 1993).

Currently knowledge of bee antimicrobial peptides is relatively poor, despite having some confirmation or reasonable hypotheses about their inhibitory mechanisms. One peptide that was discovered, called apisimin, still remains largely uninvestigated beyond several basic chemical/physical properties such as its charge and fold structure (Bilikova et al., 2002). It is even possible that other bee antimicrobial peptides have not yet been predicted, mainly because genome annotation tends to miss small genes, erring on the side of caution to reduce false positives.

#### 1.3.2.5 Lysozyme

Lysozyme is an enzyme that hydrolyzes the 1,4 glycosidic linkage of alternating units of N-acetylglucosamine and N-acetylmuramic acid residues that make up peptidoglycan of the bacterial cell wall. This protein, only about 100 amino acids long (depending on the species) and small enough to occasionally be categorized as an antimicrobial peptide, is constitutively present at low levels in insect hemolymph. Three lysozyme genes are predicted in bees (Evans et al., 2006), of which lysozyme-1 and lysozyme-2 (GI:110762174, GI:110762162 respectively) are highly similar to each other, and are classified as c-type lysozymes. They are mainly active against Gram-positive bacteria since the cell wall of those bacteria is composed largely of peptidoglycan. In a study of transcript levels of *P. larvae*-challenged bees, lysozyme-1 was upregulated in

larvae and downregulated in adults upon infection. It was also shown that lysozyme concentration in healthy larvae is 5-15 µg/ml, while bacterial infection increases the level to 1300 µg/ml (Glinski et al., 1993).

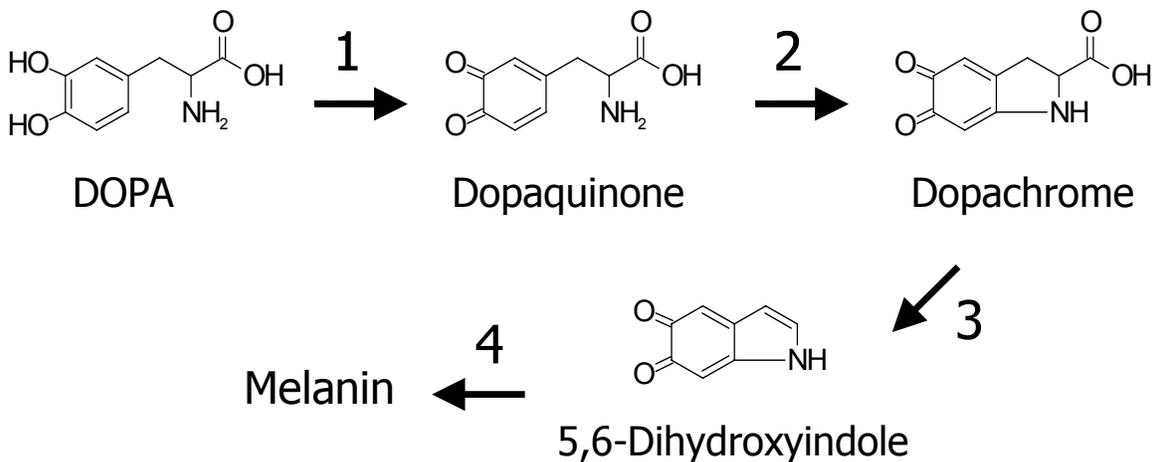
Apart from the best studied c-type lysozyme, there is a destabilase-lysozyme in bees (GI:66565246). Originally purified from the medicinal leech *Hirudo medicinalis* (Zavalova et al., 1996), it has both the activity of c-type lysozyme (Zavalova et al., 2000) and a unique isopeptidase activity, specifically cleaving the bond between gamma-carboxamide of glutamine and epsilon-amino group of lysine (Zavalova et al., 1996). For the leech, this is an important enzyme for the breakdown of blood clots, in which this isopeptide bond is made by a transglutaminase in vertebrate blood. In bees, the purpose of this enzyme is not known but Evans et al. reported that a protein they call lysozyme-3 is a destabilase-lysozyme which is downregulated at the transcript level after infection of both adults and larvae by *E. coli* and *P. larvae* (Evans et al., 2006).

#### 1.3.2.6 Complement

It is known that insects express a set of complement-like proteins. Best studied in flies and mosquitoes, these belong to a group called thioester-containing proteins (TEPs), which attach to target surfaces of pathogens by forming covalent bonds. In *Drosophila*, six TEPs have been identified, some of which are upregulated by bacterial challenge to larvae, and others to adults. *Anopheles* expresses 19 TEP homologs and the most well known is TEP1, which is transcribed in the fat body and hemocytes, has activity that is highly reminiscent of vertebrate complement factors (Blandin et al., 2004). Knockdown of this protein decreases phagocytic activity of latex beads by 50% (Levashina et al., 2001). It has been shown that the Janus kinase (JAK)/Signal Transducers and Activators of Transcription (STAT) pathway may contribute to its expression, a pathway that includes the extracellular Upd protein, which binds the transmembrane receptor domeless; intracellularly, domeless activates the JAK tyrosine kinase Hopscotch, then activates the enzyme called Signal-transducer and activator of transcription protein at 92E that activates TEP transcription. Honey bee, which has three TEP homologs, also contains homologs of the JAK/STAT pathway except for the initiating ligand, Upd. It is likely that another protein exists in its place that has only not yet been found.

### 1.3.2.7 Melanization

Melanization is the process by which tyrosine is converted to melanin, a black pigment. Most eukaryotes have the necessary molecular machinery to carry out the series of necessary reactions, including humans and insects, where the most obvious manifestation is seen in eyes, hair and skin (in the case of insects, exoskeleton). Due to the side production of free radicals, this biochemical pathway is also used by insects as a significant part of immunity in a way that humans cannot, because in humans this process is the specialty of the small population melanocytes (Sulaimon et al., 2003). Insects, on the other hand, have many types of hemocytes that are capable of this reaction. In encapsulation or nodulation responses, where pathogens are inhibited by asphyxiation and entrapment respectively, melanization can often occur, further damaging the invader (Lavine et al., 2003) by free radicals. At the site of localized infection, melanin can be seen as darkened spots. This can also occur as a response to septic injury (Marmaras et al., 2009). The synthesis of melanin occurs by the reaction shown in Figure 1.5.



**Figure 1.5 Conversion of DOPA to melanin**

3,4-dihydroxyphenylalanine (DOPA) is converted to melanin in a series of both enzymatic and spontaneous reactions. Reaction 1 is catalyzed by phenoloxidase to yield dopaquinone, which non-enzymatically reacts (Reaction 2) to form dopachrome. From that, the dopachrome-conversion enzyme (Reaction 3) generates 5,6-dihydroxyindole, and from that, oxidizes and polymerizes, sometimes aided by phenoloxidase to form the black molecule melanin (Reaction 4).

Because of the high potential of free radicals to damage DNA, proteins, and organelles, it is important for the host to regulate the melanization pathway, both spatially and temporally. Upon the detection of a microbial invader, a signaling cascade consisting mainly of serine proteases is triggered before melanization can occur and is catalyzed by the key enzyme, PO. The cascade is important for both signal amplification, as well as multi-level regulation by way of serine protease inhibitors (serpins). The final step in the activation of PO is the cleavage of an arginine-phenylalanine bond that removes its propeptide, an approximately 50-amino acid long sequence, depending on the species. This cut is made by a serine protease, known generally as the phenoloxidase activating enzyme, which is itself activated by cleavage of its propeptide (Cerenius et al., 2004). This activation process is not well known, and may vary between species with regards to the identities and number of regulatory proteases that signal between PO, the terminal component of this cascade, and the receptor that binds a pathogen-associated molecule, located at the start of this cascade. In *Drosophila*, for example, melanization protease (MP) 2 acts upstream of MP1, since their overexpression resulted in semi-lethal constitutive melanization, while their knockdown compromised PO activation (Tang et al., 2006). Another class of proteins called serpins can counter the activity of MP1 and MP2: Spn77Ba in the trachea and Spn27A in the hemolymph of *Drosophila*, which upon knockdown resulted in lethal melanization that can be reverted in a concomitant knockdown of MP1 or MP2 (De Gregorio et al., 2002, Tang et al., 2008).

Besides the serine protease cascade that directly regulates the activity of PO, there is also some evidence from *Drosophila* studies that other pathways may exert some level of control. For example, for melanization to occur, Spn27A has to be depleted which requires the activation of the Toll pathway: data suggest that one of the genes transcribed as a result of Toll activation may be responsible for Spn27A depletion (Ligoxygakis et al., 2002). It has also been observed that IMD pathway receptors PGRP-LC and PGRP-LE are able to activate melanization independent of infection.

PO functions only in the extracellular environment, but since it lacks a signal peptide to direct its secretion, PO is released to the hemolymph only by cell lysis. Melanization can occur at sterile wound sites (Galko et al., 2004), likely due to signals released during blood clotting which is accompanied by cell rupture (Bidla et al., 2007). In *Drosophila*, overexpression of PGRP-LE, a soluble receptor, leads to constitutive

melanization whereas a mutant PGRP-LE leads to blocked PO activation (Takehana et al., 2002, Takehana et al., 2004).

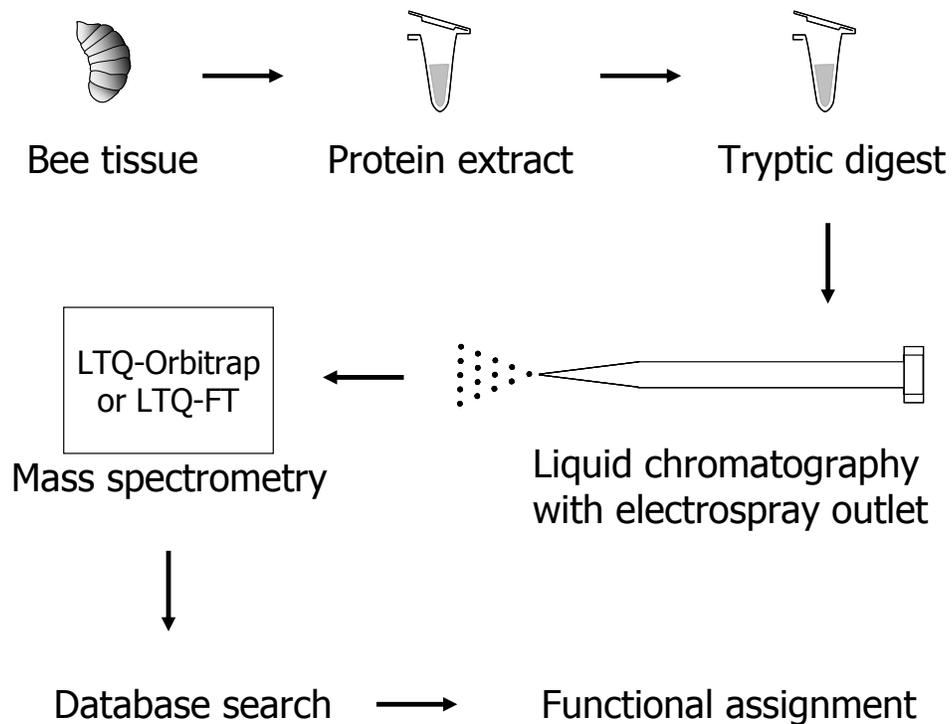
## **1.4 Experimental approach: mass spectrometry-based proteomics**

Despite the importance of bee research, a range of technical challenges deter the accessibility, speed, and ease of its progress. Many of the highly controlled experiments done on other species that may normally take place in regular laboratory settings, such as in culture dishes and reaction tubes, cannot be performed with bees at this time. Being eusocial organisms whose behaviour and biochemistry are largely affected by hivemates and the external environment, they cannot easily be raised in isolation and be expected to exhibit normal phenotypes, especially regarding their social conduct. It is also impossible to ensure that sampled bees, even from a single hive, are of the same genetic background, since a queen's stored sperm is usually derived from one to two dozen males. Artificial insemination from a single drone is possible (De Wilde et al., 1951, Laidlaw, 1949, Taber et al., 1960) but is a highly-specialized technique. Likewise, they are difficult to in-breed for even a few generations and a genetically homogenous line does not exist. Tissue culture of immortalized cells conveniently avoids the constraints associated with live animal work, but currently there are no honey bee cell lines. Primary cells can be maintained for several months (Bergem et al., 2006) but their biology outside of the whole-organism context is unknown. Unfortunately, the lack of previous biochemical/molecular biological studies on bees means that many of the tools we take for granted in other systems (e.g., antibodies, expression vectors) are not available, neither from commercial nor academic sources.

Mass spectrometry (MS)-based proteomics, followed by determination of protein function by experimental or bioinformatic means, can help us understand the biological status of a given sample. It circumvents the need for tools such as antibodies and cell lines which are not available in bee research. The draft honey bee genome sequence was announced in 2006 (Consortium, 2006), from which an official gene set and the corresponding protein set were produced in 2007, (Elsik et al., 2007) with updates downloadable from Beebase (<http://www.beebase.org>). Earlier unofficial protein databases were available, however, avoiding the need for *de novo* sequencing in

previous proteomics papers. The honey bee proteomics field is merely three to four years old, with the first paper published in 2005 (Peiren et al., 2005). The field is quickly diversifying its use of various mass spectrometry technologies, such as applying primarily two-dimensional gel electrophoresis (2DGE) followed by matrix-assisted laser desorption ionization (MALDI). However, there are increasing uses of gel-free separations coupled to nanoelectrospray ionization. Much of the MS work presented in this thesis involves in-solution tryptic digest of honey bee samples, followed by separation of peptides using liquid chromatography prior to MS. This section 1.4 discusses various aspects of MS-based proteomics that are relevant to the results included in this thesis, and is diagrammatically represented in Figure 1.6.

**Figure 1.6 General experimental scheme used for MS-based proteomics**



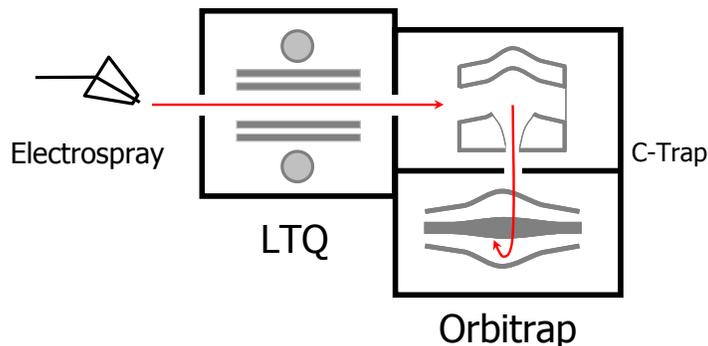
This is the simplified scheme of the steps involved in analyzing honey bee tissue samples by MS.

### 1.4.1 Tandem mass spectrometry

Tandem MS, known otherwise as MS/MS or MS<sup>n</sup> (where n equals 2 in this thesis and in most proteomics papers), refers to the acquisition of two sequential mass spectra

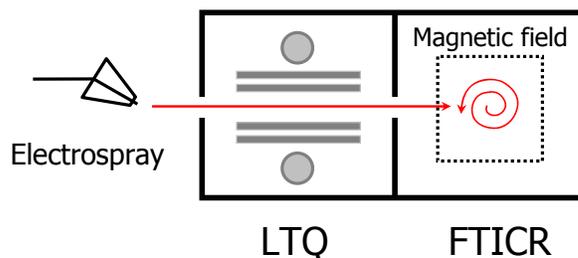
to identify a peptide. By hydrolyzing a protein sample with trypsin, peptides ending in the basic residues of lysine or arginine are produced. Together with their amino-termini, peptides with at least two positive charges in low pH environments are produced (more in the case of missed tryptic sites or histidines within the peptide). They are subjected to liquid chromatography, which will not be discussed here in detail. Briefly, the main purpose of this step is to separate complex peptide mixtures by various properties, so they appear at the outlet at different times. This reduces the effective complexity of the sample. Charged peptides, once promoted to a gaseous phase by an electrospray device, can then be manipulated in electric and/or magnetic fields, depending on the instrument used. There are many types of MS instruments and in this thesis the Linear trapping quadrupole (LTQ)-Orbitrap (Figure 1.7) and the LTQ-Fourier transform (LTQ-FT, Figure 1.8) ion cyclotron resonance (FTICR) were primarily used and will be discussed below. The Orbitrap and FTICR were chosen largely for their ability to measure peptide ion masses with high accuracy: both routinely measure to an accuracy of less than 2 parts-per-million (ppm), with the FTICR reaching sub-ppm levels. Furthermore, their high resolution of 100,000 for the Orbitrap and 500,000 for the FTICR generates fewer overlapping peaks, thus producing information-rich mass spectra compared to the lower resolution, such as the LTQ. High mass accuracy and high resolution ion spectra are crucial for accurate peptide identification and relative quantitation. In the work described in this thesis, the Orbitrap or the FTICR detects the masses of the incoming ions (parent ions), and those of interest are isolated in the LTQ portion of the instrument for fragmentation and mass detection (product ions), information from which the sequence of the peptide can be deduced.

**Figure 1.7 A simplified schematic of the LTQ-Orbitrap**



Ions exit from the electro spray outlet and enter the LTQ-Orbitrap mass spectrometer, following the path of the red arrow. The LTQ is discussed in Section 1.4.1.3 and the Orbitrap is discussed in Section 1.4.1.1. This diagram was adapted and simplified from (Scigelova et al., 2006).

**Figure 1.8 A simplified schematic of the LTQ-FT**



Ions exit from the electro spray outlet and enter the LTQ-FT mass spectrometer, following the path of the red arrow. The LTQ is discussed in Section 1.4.1.3 and the FT portion of the instrument is discussed in Section 1.4.1.2. This diagram was adapted and simplified from (Schrader et al., 2004)

#### 1.4.1.1 Orbitrap

The Orbitrap mass analyzer was invented by Alexander Makarov, first described in 2000 (Makarov, 2000) and finalized in 2005 (Hu et al., 2005). With the high mass resolution of up to 150000, ions of different mass-to-charge ratios are differentiated by an electric field, generated between a barrel-shaped outer electrode and a spindle-like inner electrode. Ions of various masses, condensed into a tight packet by what is called a C-Trap, named for its shape, are injected into this electric field. Their straight path is altered by their attraction to the inner electrode, but is balanced by its centrifugal force,

resulting in circular movement about the spindle. Ions of a given mass-to-charge ratio move together, not statically around one point of the spindle, but oscillating back and forth along it at a rate that is inversely proportional to the square root of the mass-to-charge ratio. This oscillation induces current on the outer electrodes, which are actually split along the axial direction. This is commonly known as the image current, which is recorded, and after Fourier transformation, yields frequency values from which mass-to-charge ratios can be inferred. Precursor and product ions can be measured in the Orbitrap, but in this thesis, the former is measured while the latter is evaluated in the LTQ.

#### 1.4.1.2 Fourier transform ion cyclotron resonance

Ion cyclotron resonance MS was invented by Alan Marshall and Melvin Comisarow in 1974 (Comisarow et al., 1974). Within a high magnetic field (largest existing instrument has a field strength of 14 T) and controlled application of excitation frequency, an ion's path of travel within this field can be affected, measured, and converted into mass output. This is a high resolution instrument due to the stability of the magnetic field; and the mass accuracy is positively correlated with magnetic field strength (Comisarow et al., 1976, Marshall et al., 2002).

When an ion enters the cell, its straight path becomes distorted by the magnetic field, causing it to move in a circular motion without colliding with the walls due to the trapping plates of opposing charge. The orbital of ions with any given mass-to-charge ratio can be excited to take up a higher orbit, using a pulse of radio frequency applied across the excitation plates of the cell. Their movement induces an alternating current between the detector plates, whose frequency is also the cyclotron frequency of the moving ion and whose intensity is proportional to the number of ions. Excitatory radiofrequency can be applied to excite all of the ions with different mass-to-charge ratios, and their cyclotron frequencies are recorded in what is called a free induction decay spectrum. By applying Fourier transform, this data can be deconvoluted to give a frequency vs. intensity spectrum, and each frequency value can be used to calculate the mass-to-charge ratio of the ion, given that these two values are related in an inversely-proportional manner through a constant value dictated by the magnet field.

### 1.4.1.3 Linear trapping quadrupole

In the LTQ-Orbitrap and LTQ-FT, the LTQ portion is physically placed at the stream of incoming ions before the Orbitrap or FTICR portion of the instrument. Despite this configuration, in a typical mode of operation the Orbitrap and FTICR first detects the masses of incoming ions, i.e. the precursor peptide ions; those ions with the strongest signals are subsequently captured by the LTQ portion of the mass spectrometer for fragmentation and detection, i.e. the product ions.

In the LTQ, there are four parallel rods and two plates on either end. Opposing rod pairs are electrically connected with an alternating current, along with a direct current trapping field provided by the two end plates, are able to trap charged particles without collision into the rods and plates. From the parent ion spectrum obtained by the Orbitrap or FTICR, ions of a particular mass-to-charge ratio of interest can be selected for fragmentation in the LTQ, which provides sequence information. To accomplish this, the LTQ would expel all ions from the trap except for the one with the mass-to-charge ratio of interest by superimposing a specific direct current voltage that induces unstable trajectories in the unwanted ions, causing them to collide with the rods. The desired ions are bombarded by an inert gas such as nitrogen or argon in a process called collision induced dissociation that severs the covalent bonds of the peptide ions at various places. The breakage point is often at a peptide bond, which conveniently renders their masses extremely helpful in deducing the amino acid sequence.

## 1.4.2 Database searching

The output of a typical tandem mass spectrometry experiment is a long list of ion mass-to-charge ratios and their intensities collected over the duration of the chromatographic separation. This raw data is searched against a database of peptide masses, one that hypothetically should contain every peptide that is expected to be present in the sample. In studying a particular organism, it would be useful know its genome and expressed sequence tag libraries and all possible proteins that can be predicted from these, as well as the sequences of all the likely protein contaminants, e.g. human keratins. These sequences will be theoretically digested with the same enzyme applied to the samples; mass changes resulting from the modifications to the peptides must also be considered: these peptides and their masses would form a suitable

database with which the raw data can be searched against. Additionally, masses of the theoretical fragments of these peptides are important for making use of fragment spectra collected.

These searches are conducted by specialized software available from a number of non-profit and commercial sources. The method of search varies between programs, but the general concept is similar. The experimental data, including parent ion or fragment ion masses (or spectra) are compared to the theoretical ones. The precision of the match is often given a score or ranking order. It also calculates a rate of occurrence of randomized hits – a score that is often related to the size of the database – and from that, will help determine the minimum quality score that match must be before it is considered a likely hit.

### **1.4.3 Relative protein quantitation**

The basic premise of experimental design involves the comparison between a control and test sample. MS-based relative quantitation is highly suitable in this context. Each peptide, depending on its charge, size, structure, and surrounding environment behaves uniquely with regards to both chromatographic separation and ionization properties. Based on the reproducibility of these behaviours, the signal intensity of identical peptides in two samples may be compared in relative quantitation. The signal in such an experiment is a three-dimensional peak volume: mass-to-charge (Thomson), time, and intensity.

#### **1.4.3.1 Peptide correlation profiling/ion intensities**

In a well-designed experiment the control and test samples should be identical in all ways except for the variable of interest, so under any detection system only the differences caused by the variable would be revealed. In the case of peptide mixtures, they can be resolved by liquid chromatography and detected by MS, and the signal intensities of each given peptide at a given elution time can be compared between samples that are similar, such as the control and test samples. This method relies heavily on the high level of similarity between the samples and the reproducibility of chromatography and peptide ionization. Quantitation involves the assumption that a given peptide eluting at a given time in one sample will elute at the same time in another sample; their corresponding signals are used for relative quantitation. One

caveat of this method is the imperfect run-to-run reproducibility, a problem that can become more pronounced as samples become more different, due either to known reasons for unknown ones from operator error.

#### 1.4.3.2 Isotopic labeling

In an ideal situation where a paired control and test sample are compared, they are handled and processed in an exactly identical manner, even processed in the same reaction tubes and handled as though they are one. This is, to a degree, possible with labeling methods. For purposes of MS-based detection, peptides from each sample can be chemically derivatized with modifying groups that are chemically identical but that are isotopically different. In theory there are no limitations to what this reaction is, as long as it is compatible with downstream procedures, specific to known groups and the reaction efficiency is completely or very nearly quantitative. In this thesis, chemical dimethylation is used, where the amine groups located at the amino-termini and epsilon position of the lysine side chain is modified. To one sample in a control-test pair, the light form of the methyl group  $-C^1H_3$  would be added, and to the other, the deuterated methyl group  $-C^2H_2^1H$ . To this effect, any given peptide would receive two methyl groups, and two more if it contains a lysine. We have found this reaction to be more than 99% efficient. Identical peptides in a paired sample would therefore have either a nominally 4 or 8 Da difference, which can be differentiated by a sufficiently high resolution mass spectrometer. Paired samples, once labeled, can be combined and processed in the same chromatography run, with the isotopologues of any particular peptide eluting into the mass spectrometer at same time. This greatly reduces the issue of imperfect reproducibility of peptide separation during chromatography and also ensures that the peptide ions of a given sequence are ionized in the identical environment. Particular to the use of isotopic variants of hydrogen however, is what is known as the isotope effect, referring to the differences in dipoles between the carbon-hydrogen and carbon-deuterium, which causes a slight disparity in elution times for isotopologues of some peptides. This difference can be as little as zero to as much as 10 s over a 90 min separation for the dimethylation derivatization employed here.

#### 1.4.4 Functional assignment

After searching the raw MS data to identify peptide matches, it is easy to infer with high certainty from which protein they are derived, as long as an appropriately stringent criteria against is applied against false discoveries. Yet a list of proteins, or even their relative quantities in the case of isotopic labeling experiments, would essentially be meaningless without understanding the proteins' functions. In well-studied organisms such as human, mouse, and fruit fly, this information is obtainable with relative ease: the huge amount of data available and the ample number database curators makes this possible. In the case of the honey bee, the severe lack of experimental data on the proteins implies a heavy reliance on homology-based functional inferences. The Basic Local Alignment Search Tool (BLAST) is most commonly used to search a protein sequence of interest against a database of proteins. This database would ideally be well-curated with rich information, not only regarding protein function, but also protein domains, or references to relevant literature. The BLAST-based software BLAST2GO (Conesa et al., 2005) is also extremely helpful in assigning a highly systematic, cross-species set of vocabulary to describe proteins in their various roles, known as Gene Ontology (GO). Without these tools it would be impossible to understand the biological context and significance of the proteomics data that has been generated from bee samples.

## 1.5 References

- Aggarwal K, Silverman N. **Positive and negative regulation of the *Drosophila* immune response.** *BMB Rep* 2008, 41(4):267-277.
- Aizen MA, Garibaldi LA, Cunningham SA, Klein AM. **Long-Term Global Trends in Crop Yield and Production Reveal No Current Pollination Shortage but Increasing Pollinator Dependency.** *Curr Biol* 2008.
- Albert S, Bhattacharya D, Klaudiny J, Schmitzova J, Simuth J. **The family of major royal jelly proteins and its evolution.** *J Mol Evol* 1999, 49(2):290-297.
- Albert S, Klaudiny J, Simuth J. **Molecular characterization of MRJP3, highly polymorphic protein of honeybee (*Apis mellifera*) royal jelly.** *Insect Biochem Mol Biol* 1999, 29(5):427-434.
- Aliouane Y, Kacimi El Hassani A, Gary V, Armengaud C, Lambin M, Gauthier M. **Subchronic exposure of honeybees to sublethal doses of pesticides: effects on behavior.** *Environ Toxicol Chem* 2008:1.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. **Basic local alignment search tool.** *J Mol Biol* 1990, 215(3):403-410.
- Antunez K, Anido M, Evans JD, Zunino P. **Secreted and immunogenic proteins produced by the honeybee bacterial pathogen, *Paenibacillus* larvae.** *Vet Microbiol* 2009.
- Bailey L, Ball BV. **Honey Bee Pathology**, Second edn. London: Academic Press; 1991.
- Bamrick JF, rothenbuhler WC. **Resistance to American foulbrood in honey bees. IV. The relationship between larval age at inoculation and mortality in a resistant and susceptible line.** *J Insect Pathol* 1961, 3:381-390.
- Bamrick JF. **Resistance to American foulbrood in honey bees. V. Comparative pathogenesis in resistant and susceptible larvae.** *J Insect Pathology* 1964, 6:284-304.
- Bergem M, Norberg K, Aamodt RM. **Long-term maintenance of in vitro cultured honeybee (*Apis mellifera*) embryonic cells.** *BMC Dev Biol* 2006, 6:17.
- Beukeboom LW. **Sex determination in Hymenoptera: a need for genetic and molecular studies.** *Bioessays* 1995, 17(9):813-817.
- Bidla G, Dushay MS, Theopold U. **Crystal cell rupture after injury in *Drosophila* requires the JNK pathway, small GTPases and the TNF homolog Eiger.** *J Cell Sci* 2007, 120(Pt 7):1209-1215.
- Bilikova K, Wu G, Simuth J. **Isolation of a peptide fraction from honeybee royal jelly as a potential antifoulbrood factor.** *Apidologie* 2001, 32:275-283.
- Bilikova K, Hanes J, Nordhoff E, Saenger W, Klaudiny J, Simuth J. **Apisimin, a new serine-valine-rich peptide from honeybee (*Apis mellifera* L.) royal jelly: purification and molecular characterization.** *FEBS Lett* 2002, 528(1-3):125-129.
- Bitondi MMG, Nascimento AM, Cunha AD, Guidugli KR, Nunes FMF, Simoes ZLP. **Characterization and expression of the Hex 110 gene encoding a glutamine-rich hexamerin in the honey bee, *Apis mellifera*.** *Arch Insect Biochem Physiol* 2006, 63:57-72.
- Blandin S, Levashina EA. **Thioester-containing proteins and insect immunity.** *Mol Immunol* 2004, 40(12):903-908.

- Brødsgaard CJ, Ritter W, Hansen H. **Response of in vitro reared honey bee larvae to various doses of Paenibacillus larvae spores.** *Apidologie* 1998, 29:569-578.
- Brødsgaard CJ, Ritter W, Hansen H, Brodsgaard HF. **Interactions among Varroa jacobsoni mites, acute paralysis virus, and Paenibacillus larvae larvae and their influence on mortality of larval honeybees in vitro.** *Apidologie* 2000, 31:543-554.
- Bulet P, Charlet M, Hetru C. **Innate Immunity.** In. Edited by Ezekowitz RA, Hoffmann J. Totowa, NJ: Humana Press; 2003: 89-107.
- Bulet P, Stocklin R. **Insect antimicrobial peptides: structures, properties, and gene regulation.** *Protein and Peptide Letters* 2005, 12:3-11.
- Casteels-Josson K, Capaci T, Casteels P, Tempst P. **Apidaecin multi-peptide precursor structure: a putative mechanism for amplification of the insect antibacterial response.** *Embo J* 1993, 12(4):1569-1578.
- Casteels P, Ampe C, Jacobs F, Vaeck M, Tempst P. **Apidaecins: antibacterial peptides from honeybees.** *Embo J* 1989, 8(8):2387-2391.
- Casteels P, Ampe C, Riviere L, Van Damme J, Elicone C, Fleming M, Jacobs F, Tempst P. **Isolation and characterization of abaecin, a major antibacterial response peptide in the honeybee (Apis mellifera).** *Eur J Biochem* 1990, 187(2):381-386.
- Casteels P, Ampe C, Jacobs F, Tempst P. **Functional and chemical characterization of Hymenoptaecin, an antibacterial polypeptide that is infection-inducible in the honeybee (Apis mellifera).** *J Biol Chem* 1993, 268(10):7044-7054.
- Cerenius L, Soderhall K. **The prophenoloxidase-activating system in invertebrates.** *Immunol Rev* 2004, 198:116-126.
- Chan QW, Foster LJ. **Apis mellifera proteomics: where will the future bee?** *Current Proteomics* 2009, 6(2):70-83.
- Chen Y, Evans J, Feldlaufer M. **Horizontal and vertical transmission of viruses in the honey bee, Apis mellifera.** *J Invertebr Pathol* 2006, 92(3):152-159.
- Chen YP, Siede R. **Honey bee viruses.** *Adv Virus Res* 2007, 70:33-80.
- Clark CC, Gordon RJ. **Encyclopedia of Complementary Health Practice**, 1st edn: Springer Publishing Company; 1999.
- Comisarow MB, Marshall AG. **Fourier transform ion cyclotron resonance spectroscopy.** *Chemical Physics Letters* 1974, 25(2):282-283.
- Comisarow MB, Marshall AG. **Theory of Fourier transform ion cyclotron resonance mass spectroscopy. I. Fundamental equations and low-pressure line shape.** *J Chem Physics* 1976, 64(1):110-119.
- 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, 21(18):3674-3676.
- The Honey Bee Genome Sequencing Consortium. **Insights into social insects from the genome of the honeybee Apis mellifera.** *Nature* 2006, 443(7114):931-949.
- Cook AB. **The Bee in Greek Mythology.** *The Journal of Hellenic Studies* 1895, 15:1-24.
- Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan PL, Briese T, Hornig M, Geiser DM *et al.* **A metagenomic survey of microbes in honey bee colony collapse disorder.** *Science* 2007, 318(5848):283-287.

- Crailsheim K. **Interadult feeding of jelly in honeybee (*Apis mellifera* L.) colonies.** *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology* 2004, 161(1):55-60.
- Cunha AD, Nascimento AM, Guidugli KR, Simoes ZL, Bitondi MM. **Molecular cloning and expression of a hexamerin cDNA from the honey bee, *Apis mellifera*.** *J Insect Physiol* 2005, 51(10):1135-1147.
- Dancer BN, Chantawannakul P. **The Proteases of American Foulbrood Scales.** *J Invertebr Pathol* 1997, 70(2):79-87.
- Danty E, Arnold G, Burmester T, Huet JC, Huet D, Pernollet JC, Masson C. **Identification and developmental profiles of hexamerins in antenna and hemolymph of the honeybee, *Apis mellifera*.** *Insect Biochem Mol Biol* 1998, 28(5-6):387-397.
- Davidson EW. **Ultrastructure of American foulbrood disease pathogenesis in larvae of the worker honey bee, *Apis mellifera*.** *J Invertebr Pathol* 1973, 21:53-61.
- De Gregorio E, Han SJ, Lee WJ, Baek MJ, Osaki T, Kawabata S, Lee BL, Iwanaga S, Lemaitre B, Brey PT. **An immune-responsive Serpin regulates the melanization cascade in *Drosophila*.** *Dev Cell* 2002, 3(4):581-592.
- de Oliveira VTP, Cruz-Landim C. **Morphology and function of insect fat body cells: a review.** *Biociencias* 2003, 11(2):195-205.
- De Wilde J, Drijver AA. **Artificial insemination of queen bees.** *Acta Physiol Pharmacol Neerl* 1951, 2(2):299-301.
- Diamond G, Beckloff N, Weinberg A, Kisich KO. **The roles of antimicrobial peptides in innate host defense.** *Curr Pharm Des* 2009, 15(21):2377-2392.
- Drapeau MD, Albert S, Kucharski R, Prusko C, Maleszka R. **Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees.** *Genome Res* 2006, 16(11):1385-1394.
- Dzierzon J. **Gutachten über die von Herrn Direktor Stöhr im ersten und zweiten Kapitel des General-Gutachtens aufgestellten Fragen.** *Eichstädter Bienenzeitung* 1845, 1:109-113, 119-121.
- Elango N, Hunt BG, Goodisman MA, Yi SV. **DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*.** *Proc Natl Acad Sci U S A* 2009, 106(27):11206-11211.
- Elsik CG, Mackey AJ, Reese JT, Milshina NV, Roos DS, Weinstock GM. **Creating a honey bee consensus gene set.** *Genome Biol* 2007, 8(1):R13.
- Evans JD, Wheeler DE. **Expression profiles during honeybee caste determination.** *Genome Biol* 2001, 2(1):RESEARCH0001.
- Evans JD. **Transcriptional immune responses by honey bee larvae during invasion by the bacterial pathogen, *Paenibacillus larvae*.** *J Invertebr Pathol* 2004, 85(2):105-111.
- Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, Kanost M, Thompson GJ, Zou Z, Hultmark D. **Immune pathways and defence mechanisms in honey bees *Apis mellifera*.** *Insect Mol Biol* 2006, 15(5):645-656.
- Fujiwara S, Imai J, Fujiwara M, Yaeshima T, Kawashima T, Kobayashi K. **A potent antibacterial protein in royal jelly. Purification and determination of the primary structure of royalisin.** *J Biol Chem* 1990, 265(19):11333-11337.
- Furusawa T, Rakwal R, Nam HW, Shibato J, Agrawal GK, Kim YS, Ogawa Y, Yoshida Y, Kouzuma Y, Masuo Y *et al.* **Comprehensive royal jelly (RJ) proteomics**

- using one- and two-dimensional proteomics platforms reveals novel RJ proteins and potential phospho/glycoproteins.** *J Proteome Res* 2008, 7(8):3194-3229.
- Galko MJ, Krasnow MA. **Cellular and genetic analysis of wound healing in *Drosophila* larvae.** *PLoS Biol* 2004, 2(8):E239.
- Gandhe AS, John SH, Nagaraju J. **Noduler, a novel immune up-regulated protein mediates nodulation response in insects.** *J Immunol* 2007, 179(10):6943-6951.
- Garcia L, Saraiva Garcia CH, Calabria LK, Costa Nunes da Cruz G, Sanchez Puentes A, Bao SN, Fontes W, Ricart CA, Salmen Espindola F, Valle de Sousa M. **Proteomic Analysis of Honey Bee Brain upon Ontogenetic and Behavioral Development.** *J Proteome Res* 2009.
- Gasic S, Vucevic D, Vasilijic S, Antunovic M, Chinou I, Colic M. **Evaluation of the immunomodulatory activities of royal jelly components in vitro.** *Immunopharmacol Immunotoxicol* 2007, 29(3-4):521-536.
- Gempe T, Hasselmann M, Schiott M, Hause G, Otte M, Beye M. **Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway.** *PLoS Biol* 2009, 7(10):e1000222.
- Glinski Z, Jarosz J. **Further evidence for cell-free immunity in the honeybee, *Apis mellifera*.** *Apiacta* 1993, 28:69-78.
- Govan VA, Brozel V, Allsopp MH, Davison S. **A PCR detection method for rapid identification of *Melissococcus pluton* in honeybee larvae.** *Appl Environ Microbiol* 1998, 64(5):1983-1985.
- Gregorc A, Bowen ID. **Histopathological and histochemical changes in honeybee larvae (*Apis mellifera* L.) after infection with *Bacillus* larvae, the causative agent of American foulbrood disease.** *Cell Biol Int* 1998, 22(2):137-144.
- Guo H, Saiga A, Sato M, Miyazawa I, Shibata M, Takahata Y, Morimatsu F. **Royal jelly supplementation improves lipoprotein metabolism in humans.** *J Nutr Sci Vitaminol (Tokyo)* 2007, 53(4):345-348.
- Hagenguth H, Remboldt H. **Identification of juvenile hormone 3 as the only JH homolog in all developmental stages of the honey bee.** *Zeitschrift fur Naturforschung* 1978, 33c:847-850.
- Higes M, Martin-Hernandez R, Botias C, Bailon EG, Gonzalez-Porto AV, Barrios L, Del Nozal MJ, Bernal JL, Jimenez JJ, Palencia PG *et al.* **How natural infection by *Nosema ceranae* causes honeybee colony collapse.** *Environ Microbiol* 2008, 10(10):2659-2669.
- Hoage TR, W.C. R. **Larval honey bee response to various doses of *Bacillus* larvae spores.** *J Econ Entomol* 1966, 59:42-45.
- Holmes MM, Rosen GJ, Jordan CL, de Vries GJ, Goldman BD, Forger NG. **Social control of brain morphology in a eusocial mammal.** *Proc Natl Acad Sci U S A* 2007, 104(25):10548-10552.
- Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. **The Orbitrap: a new mass spectrometer.** *J Mass Spectrom* 2005, 40(4):430-443.
- Klaudiny J, Albert S, Bachanova K, Kopernicky J, Simuth J. **Two structurally different defensin genes, one of them encoding a novel defensin isoform, are expressed in honeybee *Apis mellifera*.** *Insect Biochem Mol Biol* 2005, 35(1):11-22.

- Kocks C, Cho JH, Nehme N, Ulvila J, Pearson AM, Meister M, Strom C, Conto SL, Hetru C, Stuart LM *et al.* **Eater, a transmembrane protein mediating phagocytosis of bacterial pathogens in *Drosophila*.** *Cell* 2005, 123(2):335-346.
- Kucharski R, Maleszka R, Hayward DC, Ball EE. **A royal jelly protein is expressed in a subset of Kenyon cells in the mushroom bodies of the honey bee brain.** *Naturwissenschaften* 1998, 85(7):343-346.
- Kucharski R, Maleszka J, Foret S, Maleszka R. **Nutritional control of reproductive status in honeybees via DNA methylation.** *Science* 2008, 319(5871):1827-1830.
- Kurucz E, Markus R, Zsomboki J, Folkl-Medzihradzky K, Darula Z, Vilmos P, Udvardy A, Krausz I, Lukacsovich T, Gateff E *et al.* **Nimrod, a putative phagocytosis receptor with EGF repeats in *Drosophila* plasmatocytes.** *Curr Biol* 2007, 17(7):649-654.
- Kuwabara M. **Ueber die regulation im weisellosen Volke der Honigbiene, besonders die Bestimmung des neuen Weisels.** *J Fac Sci Hokkaido Univ* 1947, 9(6):359-381.
- Laidlaw HH, Jr. **Development of precision instruments for artificial insemination of queen bees.** *J Econ Entomol* 1949, 42(2):254-261.
- Lapidge KL, Oldroyd BP, Spivak M. **Seven suggestive quantitative trait loci influence hygienic behavior of honey bees.** *Naturwissenschaften* 2002, 89(12):565-568.
- Lavine MD, Strand MR. **Insect hemocytes and their role in immunity.** *Insect Biochem Mol Biol* 2002, 32(10):1295-1309.
- Lavine MD, Strand MR. **Haemocytes from *Pseudoplasia includens* express multiple alpha and beta integrin subunits.** *Insect Mol Biol* 2003, 12(5):441-452.
- Lemaitre B, Hoffmann J. **The host defense of *Drosophila melanogaster*.** *Annu Rev Immunol* 2007, 25:697-743.
- Levashina EA, Moita LF, Blandin S, Vriend G, Lagueux M, Kafatos FC. **Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*.** *Cell* 2001, 104(5):709-718.
- Levenbook L, Bauer A. **The fate of larval storage protein calliphorin during adult development of *Calliphora vicina*.** *Insect Biochem* 1984, 14:77-86.
- Li J, Li H, Zhang Z, Pan Y. **Identification of the proteome complement of high royal jelly producing bees (*Apis mellifera*) during worker larval development.** *Apidologie* 2007, 38(6):545-557.
- Li WF, Ma GX, Zhou XX. **Apidaecin-type peptides: biodiversity, structure-function relationships and mode of action.** *Peptides* 2006, 27(9):2350-2359.
- Ligoxygakis P, Pelte N, Ji C, Leclerc V, Duvic B, Belvin M, Jiang H, Hoffmann JA, Reichhart JM. **A serpin mutant links Toll activation to melanization in the host defence of *Drosophila*.** *Embo J* 2002, 21(23):6330-6337.
- Lindauer M. **Ein Beitrag zur Frage der Arbeitsteilung im Bienenstaat.** *Bee World* 1952, 34:63-73, 85-90.
- Lineburg B. **The feeding of honeybee larvae.** *Bull US Dep Agric* 1924, 1222:25-37.
- Makarov A. **Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.** *Anal Chem* 2000, 72(6):1156-1162.

- Marmaras VJ, Lampropoulou M. **Regulators and signalling in insect haemocyte immunity.** *Cell Signal* 2009, 21(2):186-195.
- Marshall AG, Hendrickson CL. **Fourier transform ion cyclotron resonance detection: principles and experimental configurations.** *Int J Mass Spectrometry* 2002, 215(1-3):59-75.
- Martins JR, Nunes FM, Simoes ZL, Bitondi MM. **A honeybee storage protein gene, hex 70a, expressed in developing gonads and nutritionally regulated in adult fat body.** *J Insect Physiol* 2008, 54(5):867-877.
- McMullan JB, Brown MJ. **Brood-cell size does not influence the susceptibility of honey bees (*Apis mellifera*) to infestation by tracheal mites (*Acarapis woodi*).** *Exp Appl Acarol* 2006, 39(3-4):273-280.
- Michener CD, Grimaldi DA. **The oldest fossil bee: apoid history, evolutionary stasis, and the antiquity of social behavior.** *Proceedings of the National Academy of Sciences* 1988, 85:6424-6426.
- Miller JS, Nguyen T, Stanley-Samuels DW. **Eicosanoids mediate insect nodulation responses to bacterial infections.** *Proc Natl Acad Sci U S A* 1994, 91(26):12418-12422.
- Mundo MA, Padilla-Zakour OI, Worobo RW. **Growth inhibition of foodborne pathogens and food spoilage organisms by select raw honeys.** *Int J Food Microbiol* 2004, 97(1):1-8.
- Nelson JA. **Morphology of the honeybee larva.** *J Agricult Res* 1924, 28(12):1167-1229.
- Oldroyd BP. **What's killing American honey bees?** *PLoS Biol* 2007, 5(6):e168.
- Pal S, Wu LP. **Lessons from the fly: pattern recognition in *Drosophila melanogaster*.** *Adv Exp Med Biol* 2009, 653:162-174.
- 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.** *Biochim Biophys Acta* 2005, 1752(1):1-5.
- Pernal SF. **CAPA Statement on Honey Bees Losses in Canada (2009).** In.: Canadian Association of Professional Apiculturalists; 2009.
- Pili-Floury S, Leulier F, Takahashi K, Saigo K, Samain E, Ueda R, Lemaitre B. **In vivo RNA interference analysis reveals an unexpected role for GGBP1 in the defense against Gram-positive bacterial infection in *Drosophila* adults.** *J Biol Chem* 2004, 279(13):12848-12853.
- Riddiford LM, Hiruma K, Lan Q. **Regulation and role of nuclear receptors during larval molting and metamorphosis of Lepidoptera.** *American Zoologist* 1999, 39:736-746.
- Sammataro D, Gerson U, Needham G. **Parasitic mites of honey bees: life history, implications, and impact.** *Annu Rev Entomol* 2000, 45:519-548.
- Sanford MT. **Diseases and Pests of the Honey Bee.** In. Edited by Entomology and Nematology Department FCES, Institute of Food and Agricultural Science, University of Florida.; 2003.
- Scarselli R, Donadio E, Giuffrida MG, Fortunato D, Conti A, Balestreri E, Felicioli R, Pinzauti M, Sabatini AG, Felicioli A. **Towards royal jelly proteome.** *Proteomics* 2005, 5(3):769-776.
- Schmitzova J, Kloudiny J, Albert S, Schroder W, Schreckengost W, Hanes J, Judova J, Simuth J. **A family of major royal jelly proteins of the honeybee *Apis mellifera* L.** *Cell Mol Life Sci* 1998, 54(9):1020-1030.

- Schneider I. **Cell lines derived from late embryonic stages of *Drosophila melanogaster***. *J Embryol Exp Morphol* 1972, 27(2):353-365.
- Schrader W, Klein HW. **Liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry (LC-FTICR MS): an early overview**. *Anal Bioanal Chem* 2004, 379(7-8):1013-1024.
- Scigelova M, Makarov A. **Orbitrap mass analyzer--overview and applications in proteomics**. *Proteomics* 2006, 6 Suppl 2:16-21.
- Seeley TD, Tarpy DR. **Queen promiscuity lowers disease within honeybee colonies**. *Proc Biol Sci* 2007, 274(1606):67-72.
- Shimanuki H. **Bacteria**. In: *Honey bee pests, predators, and diseases*. Edited by Morse RA, Flottum K. Medina: A.I. Root Co.; 1997: 35-54.
- Shuel RW, Dixon SE. **An artificial diet for laboratory rearing of honeybees**. *Journal of Apicultural Research* 1986, 25(1):35-43.
- Snodgrass RE. **Anatomy and physiology of the honeybee**, 1st edn. New York: McGraw-Hill; 1925.
- Stokstad E. **Entomology. The case of the empty hives**. *Science* 2007, 316(5827):970-972.
- Stokstad E. **Genomics. Puzzling decline of U.S. bees linked to virus from Australia**. *Science* 2007, 317(5843):1304-1305.
- Sulaimon SS, Kitchell BE. **The biology of melanocytes**. *Veterinary Dermatology* 2003, 14(2):57-65.
- Sultana A, Nabi AH, Nasir UM, Maruyama H, Suzuki KM, Mishima S, Suzuki F. **A dipeptide YY derived from royal jelly proteins inhibits renin activity**. *Int J Mol Med* 2008, 21(6):677-681.
- Taber S, 3rd, Blum MS. **Preservation of honey bee semen**. *Science* 1960, 131:1734-1735.
- Takehana A, Katsuyama T, Yano T, Oshima Y, Takada H, Aigaki T, Kurata S. **Overexpression of a pattern-recognition receptor, peptidoglycan-recognition protein-LE, activates imd/relish-mediated antibacterial defense and the prophenoloxidase cascade in *Drosophila* larvae**. *Proc Natl Acad Sci U S A* 2002, 99(21):13705-13710.
- Takehana A, Yano T, Mita S, Kotani A, Oshima Y, Kurata S. **Peptidoglycan recognition protein (PGRP)-LE and PGRP-LC act synergistically in *Drosophila* immunity**. *Embo J* 2004, 23(23):4690-4700.
- Tang H, Kambris Z, Lemaitre B, Hashimoto C. **Two proteases defining a melanization cascade in the immune system of *Drosophila***. *J Biol Chem* 2006, 281(38):28097-28104.
- Tang H, Kambris Z, Lemaitre B, Hashimoto C. **A serpin that regulates immune melanization in the respiratory system of *Drosophila***. *Dev Cell* 2008, 15(4):617-626.
- Wedenig M, Riessberger-Galle U, Crailsheim K. **A substance in honey bee larvae inhibits the growth of *Paenibacillus* larvae**. *Apidologie* 2003, 34:43-51.
- White GF. **The bacteria of the apiary with special reference to bee disease**. In: Edited by Entomology Bo, vol. 14: USDA; 1906: 1-50.
- Wilson EO, Holldobler B. **Eusociality: origin and consequences**. *Proc Natl Acad Sci U S A* 2005, 102(38):13367-13371.

- Winston ML. **The Biology of the Honey Bee**. In. Cambridge, MA: Harvard University Press; 1987: 50.
- Winston ML. **The Biology of the Honey Bee**. In. Cambridge, MA: Harvard University Press; 1987: 97.
- Winston ML. **The Biology of the Honey Bee**. In. Cambridge, MA: Harvard University Press; 1987: 67.
- Winston ML. **The Biology of the Honey Bee**. In. Cambridge, MA: Harvard University Press; 1987: 69.
- Winston ML. **The Biology of the Honey Bee**. Cambridge, MA: Harvard University Press; 1987.
- Yue D, Nordhoff M, Wieler LH, Genersch E. **Fluorescence in situ hybridization (FISH) analysis of the interactions between honeybee larvae and Paenibacillus larvae, the causative agent of American foulbrood of honeybees (Apis mellifera)**. *Environ Microbiol* 2008, 10(6):1612-1620.
- Zakaria ME. **The cellular immunity responses in the haemolymph of honey bee workers infected by american foulbrood disease (AFB)**. *Journal of Applied Science Research* 2007, 3(1):56-63.
- Zavalova L, Lukyanov S, Baskova I, Snezhkov E, Akopov S, Berezhnoy S, Bogdanova E, Barsova E, Sverdlov ED. **Genes from the medicinal leech (Hirudo medicinalis) coding for unusual enzymes that specifically cleave endo-epsilon (gamma-Glu)-Lys isopeptide bonds and help to dissolve blood clots**. *Mol Gen Genet* 1996, 253(1-2):20-25.
- Zavalova LL, Baskova IP, Lukyanov SA, Sass AV, Snezhkov EV, Akopov SB, Artamonova, II, Archipova VS, Nesmeyanov VA, Kozlov DG *et al*. **Destabilase from the medicinal leech is a representative of a novel family of lysozymes**. *Biochim Biophys Acta* 2000, 1478(1):69-77.

## 2 QUANTITATIVE COMPARISON OF CASTE DIFFERENCES IN HONEY BEE HEMOLYMPH<sup>2</sup>

### 2.1 Abstract

The honey bee, *Apis mellifera*, is an invaluable partner in agriculture around the world both for its production of honey and, more importantly, for its role in pollination. Honey bees are largely unexplored at the molecular level despite a long and distinguished career as a model organism for understanding social behavior. Like other eusocial insects, honey bees can be divided into several castes: the queen (fertile female), workers (sterile females), and drones (males). Each caste has different energetic and metabolic requirements, and each differs in its susceptibility to pathogens, many of which have evolved to take advantage of the close social network inside a colony. Hemolymph, arthropods' equivalent to blood, distributes nutrients throughout the bee, and the immune components contained within it form one of the primary lines of defense against invading microorganisms. In this study we have applied qualitative and quantitative proteomics to gain a better understanding of honey bee hemolymph and how it varies among the castes and during development. We found large differences in hemolymph protein composition, especially between larval and adult stage bees and between male and female castes but even between adult workers and queens. We also provide experimental evidence for the expression of several unannotated honey bee genes and for the detection of biomarkers of a viral infection. Our data provide an initial molecular picture of honey bee hemolymph, to a greater depth than previous studies in other insects, and will pave the way for future biochemical studies of innate immunity in this animal.

---

<sup>2</sup> A version of this chapter has been accepted for publication. Chan, Q.W., Howes, C.G. and Foster, L.J. (2006) Quantitative comparison of caste differences in honeybee hemolymph. *Mol Cell Proteomics*. 5(12):2252-2262.

## 2.2 Introduction

In higher organisms, blood is responsible for supplying nutrients to tissues and organs. This fluid is a complex mixture of whole cells, proteins, lipids, carbohydrates, nucleic acids and hormones, as well as degradation products of all of these. Strictly speaking, blood is a term specific for animals with a closed circulatory system so in arthropods, which have an open circulation, the analogous fluid is known as hemolymph. Key proteins in all adult hemolymph include: transferrin (Kucharski et al., 2003), an iron transporter, apolipoprotein (Robbs et al., 1985, Shipman et al., 1987), a lipid transporter, and vitellogenin, a major nutrient storage protein in females thought to serve as a nutrient reserve and lipid carrier (Barchuk et al., 2002, Wheeler et al., 1990). The fluid also contains components of the innate immune system, such as macrophage-like cells (hemocytes) (Lavine et al., 2002), antimicrobial peptides (Hoffmann, 2003), and prophenoloxidase for the encapsulation of pathogens (Lourenco et al., 2005). The innate immune system of insects holds many similarities to innate immunity in mammals, yet the study of hemolymph has lagged far behind mammalian serum, largely due to the enormous effort and resource investment into the search for disease biomarkers in humans. In the last four years only a handful of studies have used mass spectrometry to characterize hemolymph peptides and proteins in insects. *Drosophila melanogaster* is the best studied among them as researchers have worked to elucidate the hemolymph proteomes of healthy flies (Guedes Sde et al., 2003, Karlsson et al., 2004, Vierstraete et al., 2003), post-injury (Vierstraete et al., 2004) and immune-challenged flies (de Morais Guedes et al., 2005, Levy et al., 2004, Vierstraete et al., 2004, Vierstraete et al., 2005). Mass spectrometry-based hemolymph studies have also been conducted for species as diverse as *Helicoverpa armigera* (moth) (Zhao et al., 2006), *Bombyx mori* (silkworm) (Wang et al., 2004), *Amblyomma hebraeum* (tick) (Lai et al., 2004), and *Anopheles gambiae* (mosquito) (Han et al., 1999, Paskewitz et al., 2005).

Through its immense impact on flower pollination *Apis mellifera* (honey bee) has arguably the biggest impact on humanity of any member of Class Insecta, even more than *A. gambiae*, the malarial vector. A study from Cornell University estimated that the yearly value of pollination by honey bees for the United States alone is 14.6 billion dollars (Morse et al., 2000). Domestic consumption and export of honey is profitable

industry, worth more than 250 million dollars per year and is on the rise (2003 statistics, US Department of Agriculture) (2005). While the social behavior of honey bees has been well studied, little is known about the biochemistry of honey bees. On a systems biology level, recent transcriptome analyses (Evans et al., 2001, Kucharski et al., 2005, Whitfield et al., 2002, Whitfield et al., 2003) are the first inroads to studying the expression patterns of numerous low-abundance proteins in *A. mellifera*, but have been applied specifically to brain or whole body and not to hemolymph.

The honey bee genome sequencing and annotation effort is nearing completion making effective proteomic analysis of this animal possible. Partial proteomes of RJ and pollen (Scarselli et al., 2005) and bee venom (Peiren et al., 2005) have been analyzed mainly by 2DGE followed by MALDI-time of flight (TOF), resulting in 3 and 9 unique protein identifications respectively. Biochemical analysis of honey bee hemolymph over the past forty years has focused on amino acid content, sugar metabolism and the correlation of juvenile hormone to vitellogenin protein content, but beyond these two polypeptides there is very limited knowledge about the protein content of this fluid. The overall aim of this study was to expand our knowledge of honey bee hemolymph proteins beyond the very small number that are already known as a prelude to future biochemical studies of innate immunity in bees. Furthermore, we measured caste- and stage-specific protein differences in this critical fluid, providing an initial molecular understanding of the susceptibility of specific stages or castes to major honey bee diseases.

## **2.3 Experimental Procedures**

### **2.3.1 Materials**

All salts were of analytical grade or better and were obtained from Sigma-Aldrich (St. Louis, MO) unless otherwise indicated. All solvents were of HPLC grade and were obtained from Fisher Scientific (Hampton, NH). The following materials were obtained as indicated: Endopeptidase LysC - Wako Chemicals (Osaka, Japan), porcine modified trypsin – Promega (Nepean, ON), dithiothreitol and iodoacetamide – Sigma-Aldrich, loose ReproSil-Pur 120 C18-AQ 3  $\mu\text{m}$  – Dr. Maisch (Ammerbuch-Entringen, Germany), pre-packed 15 cm x 75  $\mu\text{m}$  inner diameter 3  $\mu\text{m}$  C18 column – Dionex (Sunnyvale, CA),

96-well full skirt PCR plates – Axygen (Union City, CA), fused silica capillary tubing – Polymicro (Phoenix, AZ), 5  $\mu$ L Microcap pipettes for hemolymph collection - Drummond (Broomall, PA), soft forceps for holding bees – BioQuip (Rancho Dominguez, CA), Protease Inhibitor Cocktail – Roche (Basel, Switzerland), precast 4-12%, 1 mm thick NuPAGE Novex BisTris Gels – Invitrogen (Carlsbad, CA).

### **2.3.2 Hemolymph collection**

Drones, hive workers, worker larvae and three queens were collected from colonies maintained at the University of British Columbia, Canada. An additional queen was acquired from a feral colony in Vancouver, Canada. Samples were taken in early autumn and spring. Additional worker larvae samples were collected at the Agriculture Canada Research Station in Beaverlodge, Alberta. For collection of hemolymph from all adult castes, the bee was held between soft forceps, trapping its wings against the thorax while a glass disposable 5  $\mu$ L Microcap pipette was inserted dorsally into the bee between the second and third abdominal terga from the thorax. Clear and slightly yellow hemolymph was drawn out by capillary action. If cloudy yellow intestinal contents were taken, the sample was discarded. Hemolymph was dispensed into a 1.5 mL microfuge tube containing 100  $\mu$ L of PBS with EDTA-free Protease Inhibitor Cocktail kept chilled on ice. For collection of larval hemolymph, five day-old (estimated) larvae were taken and washed in 10 mL PBS to reduce contamination from worker jelly. A small incision was then made two-thirds of the way down one side of the larva and milky white hemolymph was drawn as described above. Inspection of larval hemolymph by light microscopy revealed that the cloudiness of the liquid was due to a high number of hemocytes and prohemocytes. The microfuge tubes were spun at 16,100 x g in a microcentrifuge (Eppendorf) for 10 min at 4°C to pellet these cells and other debris, which were discarded. For protein identification, hemolymph from 3-6 workers or drones was pooled, while for quantitation, 3 samples of 2 bees from each caste were analyzed independently. Hemolymph from the queens was not pooled at all. The protein concentration of clarified hemolymph was assayed using Coomassie Plus Protein Assay Reagent (Pierce) according to the manufacturer's instructions and stored at -20°C.

### **2.3.3 Electrophoresis**

Hemolymph proteins from adult queens, drones, workers, and worker larvae were resolved on precast 4-12% NuPAGE gels in reducing conditions with MES buffer according to manufacturer's instructions (30 or 60 µg per lane). Blue-silver stain (Candiano et al., 2004) was used to visualize protein bands and then each lane was cut into between twenty and thirty slices of approximately equal staining.

### **2.3.4 Proteolytic digestion**

For in-solution digestion of hemolymph samples, proteins were precipitated in ethanol/sodium acetate (Foster et al., 2003) and resuspended in 6 M urea, 2 M thiourea, 20 mM Tris-HCl, pH 8.0 (urea/thiourea solution was deionized prior to buffering and then stored in aliquots at -80°C). In-gel (Shevchenko et al., 1996) and in-solution (Foster et al., 2003) digestions were performed essentially as described. After acidification of peptides with 1% trifluoroacetic acid, 0.5% acetic acid and 3% acetonitrile (Sample Buffer), a volume corresponding to 2 µg of total protein was purified and concentrated on STAGE-tips (Rappsilber et al., 2003), eluted in 80% acetonitrile, 0.5% acetic acid, dried in a vacuum concentrator (Eppendorf) and resuspended in Sample Buffer.

### **2.3.5 Liquid chromatography tandem mass spectrometry**

Peptide samples from in-gel digestions were separated by HPLC using a C18 column (Dionex) coupled to a Q-TOF hybrid instrument (QSTAR, Pulsar i, Applied Biosystems). Larval samples collected in Beaverlodge were analyzed on a linear trapping quadrupole-Fourier transform mass spectrometer (LTQ-FT, Thermo Electron, Bremen, Germany) and all quantitative analysis was performed on an LTQ-Orbitrap (Thermo Electron). The QSTAR was on-line coupled to a Dionex UltiMate capillary flow HPLC using a nanospray ionization source (Proxeon Biosystems, Odense, Denmark). Precolumns were only used with the QSTAR and the analytical column used with the QSTAR was 15 cm long, 75 µm inner diameter fritted fused silica preppacked with 3 µm C18 beads (Dionex UltiMate). For QSTAR analyses buffer A consisted of 0.1% formic acid and buffer B consisted of 0.1% formic acid in acetonitrile. Gradients were run from 2% B to 30% B over 45 minutes and then ramped up steeply to 80% B for ten minutes

to wash the column before reconditioning with 2% B. Data-dependent acquisition settings were set as described (de Hoog et al., 2004). The LTQ-FT and LTQ-Orbitrap systems were on-line coupled to Agilent 1100 Series nanoflow HPLCs using nanospray ionization sources (Proxeon Biosystems, Odense, Denmark) holding columns packed into 15 cm long, 75  $\mu\text{m}$  inner diameter fused silica emitters (8  $\mu\text{m}$  diameter opening, pulled on a P-2000 laser puller from Sutter Instruments) using 3  $\mu\text{m}$  diameter ReproSil Pur C18 beads. For both instruments buffer A consisted of 0.5% acetic acid and buffer B consisted of 0.5% acetic acid and 80% acetonitrile. Gradients were run from 6% B to 30% B over 60 minutes, then 30% B to 80% B in the next 10 minutes, held at 80% B for five minutes and then dropped to 6% B for another 15 minutes to recondition the column. The LTQ-FT was set to acquire a full range scan at 25,000 resolution in the FT, from which the three most intense multiply-charged ions per cycle were isolated for fragmentation in the LTQ. At the same time selected ion monitoring scans in the FT were carried out on each of the same three precursor ions exactly as described (Olsen et al., 2004). The LTQ-Orbitrap was set to acquire a full-range scan at 60,000 resolution from 350 to 1500 Th in the Orbitrap and to simultaneously fragment the top five peptide ions in each cycle in the LTQ. Since Orbitrap data were to be used for quantitation, blank gradients where buffer B was injected were interspersed between analytical gradients to eliminate carry-over.

### **2.3.6 Analysis of mass spectrometry data**

Centroided fragment peak lists were processed to Mascot generic format using the vendor-provided Extract\_MSN.exe (Thermo Electron, Bremen, Germany) for LTQ-FT/Orbitrap data or Mascot.dll (Applied Biosystems, Foster City, CA) for QSTAR data. Monoisotopic peak and charge states in LTQ-FT/Orbitrap data were corrected using DTA Supercharge, part of the MSQuant suite of software (<http://msquant.sourceforge.net>). Using Mascot v2.1 (Matrix Science, [www.matrixscience.com](http://www.matrixscience.com)), peak lists were searched against a protein database composed of predicted proteins from the Amel\_2.0 genome assembly (9759 sequences) compiled by the National Center for Biotechnology Information ([ftp://ftp.ncbi.nih.gov/genomes/Apis\\_mellifera/special\\_requests/NCBI-prot.fa.gz](ftp://ftp.ncbi.nih.gov/genomes/Apis_mellifera/special_requests/NCBI-prot.fa.gz)), as well as sequences for human keratins, porcine trypsin, and lysyl endopeptidase. The following criteria were used in the Mascot search for QSTAR data: trypsin cleavage specificity with up to one missed cleavage (Olsen et al., 2004), cysteine

carbamidomethyl fixed modification, no variable modifications,  $\pm 0.2$  Da peptide tolerance and MS/MS tolerance and ESI-QUAD-TOF fragmentation scoring. For LTQ-Orbitrap and LTQ-FT spectra, peptide tolerance and MS/MS tolerance were set at 10 ppm and 0.8 Da, respectively, and the scoring scheme used was ESI-TRAP. MSQuant v1.4.0a17 was used for parsing Mascot files and iterative mass recalibration. Another Mascot search was conducted against the MSDB database (Viridae taxonomy) for potential honey bee disease biomarkers using the same conditions as above. For data from in gel-digested samples, additional searches were made with X!Tandem (Craig et al., 2004), using the same criteria as used for Mascot searches with the Quad-TOF (100ppm) as the predefined scoring method.

After removing peptides from keratins, trypsin and LysC, peptides from all honey bee samples were compiled, redundancies removed and then BLASTed (Altschul et al., 1997) (Word size, 2; scoring matrix, PAM30; maximum expectation value, 20000) against the same honey bee protein database as used above to arrive at a final, non-redundant protein list (Foster et al., 2006). In-house scripts were used to extract protein matches from the BLAST output file with a minimum identity of 99.9% and to find the percent sequence coverage of each protein. At least 2 peptides over 7 residues in length and with a Mascot score greater than 25 or an X!Tandem expectation value less than 0.1 were required to consider a protein hit significant. To estimate the frequency of false-positive identifications using these criteria, all spectra were searched again by Mascot against a database where all honey bee proteins sequences were reversed. In order to identify unannotated peptides in the honey bee genome, all fragment spectra that did not match to a peptide in the initial Mascot searches described above were collected and used again to search against all six translated open reading frames in the Amel\_2.0 genome assembly (228567597 base pairs, [ftp://ftp.ensembl.org/pub/current\\_apis\\_mellifera/data/fasta/dna/Apis\\_mellifera.AMEL2.0.apr.dna.contig.fa.gz](ftp://ftp.ensembl.org/pub/current_apis_mellifera/data/fasta/dna/Apis_mellifera.AMEL2.0.apr.dna.contig.fa.gz)). Only those spectral matches with scores greater than 33 (the 99% confidence limit calculated by Mascot) were considered further. To access mapped peptides, go to the Ensembl mainpage for honey bees<sup>3</sup> ([http://www.ensembl.org/Apis\\_mellifera/index.html](http://www.ensembl.org/Apis_mellifera/index.html)) and in the "Region" box, type in a

---

<sup>3</sup> While fully functional at the time of publication, Ensembl no longer supports the honey bee genome nor any of the associated tools.

region of choice in the genome, which will be displayed as "ContigView". Under the "detailed view" window, click on "DAS Sources" in the title bar and select "Manage sources...". On the left-side of the page, click on "Add Data Source". In the "DAS Server URL" drop-down menu, select "http://das.ensembl.org/das/", then in the box directly below, enter "hydraeuf\_00001563" for DAS source name and click "Next". Choose "Next" and "Finish" on the following 2 pages. Upon viewing the genome again in "ContigView", the newly added "hydraeuf\_00001563" data can be selected in the "DAS Sources" drop-down menu. Once the page reloads, mapped peptides will be displayed in its own separate track titled "hydraeuf\_00001563".

MSQuant was used to correlate elution times between parallel analyses of in-solution-digested hemolymph samples and then to calculate peptide ion volumes (in Th\*s) as described (Foster et al., 2005) with manual validation of each value. In order to ensure the reliability of measured values, only proteins having at least three quantifiable peptides in each of three experiments were considered.

### **2.3.7 Gene ontology**

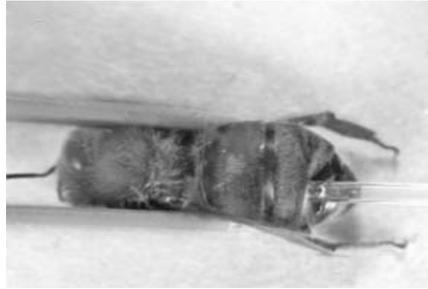
GO (Ashburner et al., 2000) assignments were made using Blast2GO (Conesa et al., 2005). The BLASTp searches were done against the nr database, with an expectation value maximum of 1e-3 and HSP length cutoff of 33. Annotation was made using the following criteria: pre-eValue-Hit-Filter, 1e-6; pre-Similarity-Hit-Filter, 15; Annotation Cutoff, 55; GO Weight, 5. Directed acyclic graphs were constructed using a sequence filter of 7, score alpha of 0.6, and node score filter of 0. From these graphs, proteins were categorized under Molecular Function and Biological Process using terms from level 3 of the graph.

## **2.4 Results**

### **2.4.1 Hemolymph extraction**

The most straightforward and well-established procedure for extracting hemolymph from adult bees is to insert a small glass capillary under the second abdominal segment (Figure 2.1) (Lin et al., 1999).

**Figure 2.1 Hemolymph extraction from an adult honey bee**



An adult worker was held between soft forceps while a glass capillary was inserted between the second and the third abdominal segment.

This method appeared to yield hemolymph relatively uncontaminated by environmental proteins (see below). On the other hand, the equivalent procedure on larvae too often resulted in fecal contamination, especially in older larvae, so an incision was made on the left-hand side of the larvae and the resulting hemolymph was collected in a glass capillary (Figure 2.2).

**Figure 2.2 Hemolymph extraction from a honey bee larva**



A mature larva was pierced by a glass capillary, which was inserted to draw hemolymph.

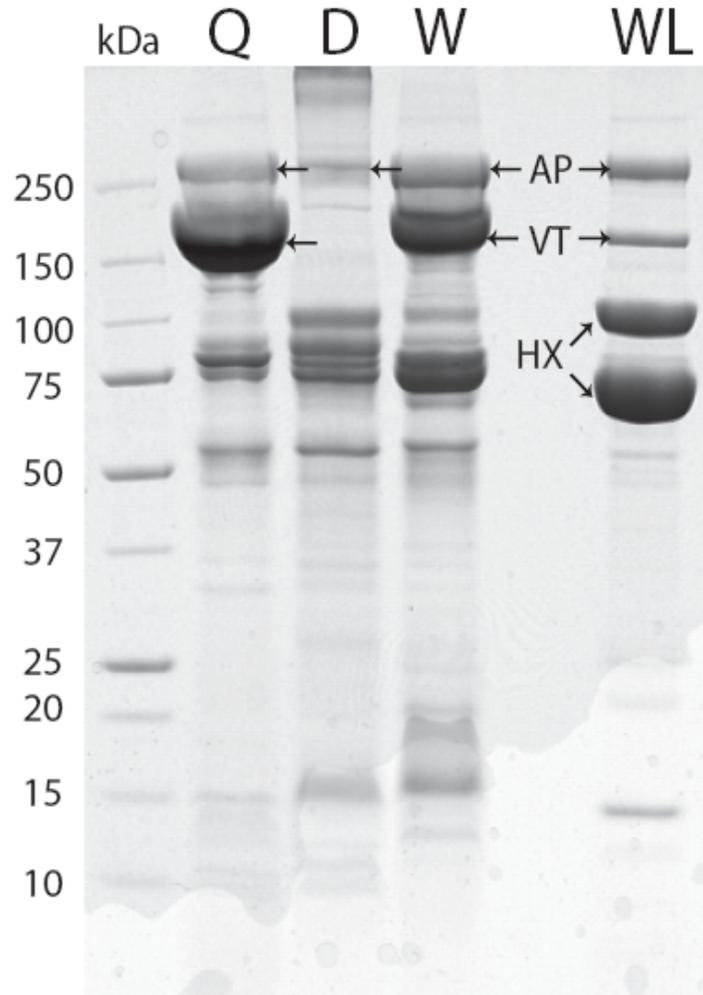
Initial hemolymph extracts from worker larvae were heavily contaminated with MRJPs to the extent that even vitellogenin, the protein that should be most abundant, was difficult to detect (results not shown). MRJPs were likely present because bee larvae are immersed in jelly from the time the egg hatches until pupation. Several variations of washing conditions were explored but simply washing the larvae in PBS for 2 minutes effectively reduced the MRJP contamination to levels seen in adult workers and drones (see Experimental Procedures). Adult workers and drones likely retain some MRJPs on their exterior from their larval and pupal stages and may acquire some through their diet (Schafer et al., 2006) but interestingly no MRJPs were detected in queen samples. This

may reflect the constant grooming of the queen by attendant workers (Winston, 1987) that would clear away any residual proteins. Alternatively, the longer lifespan of the queen may simply allow more time for such proteins to be removed. Finally, since these proteins were found on worker larvae it suggests that “major royal jelly proteins” are more generic than their name suggests, since “royal jelly” implies it is fed only to queen larvae.

#### **2.4.2 Qualitative analysis of caste and stage differences in hemolymph proteins**

Our initial hypothesis when starting these experiments was that we would only find very subtle differences in protein profiles of different castes, the reason being that sera from male and female humans are essentially indistinguishable by normal proteomic methods until one drills far down to the abundance levels of hormones. Straightforward gel electrophoresis of hemolymph proteins from different honey bee castes and worker larvae disproved this hypothesis immediately (Figure 2.3).

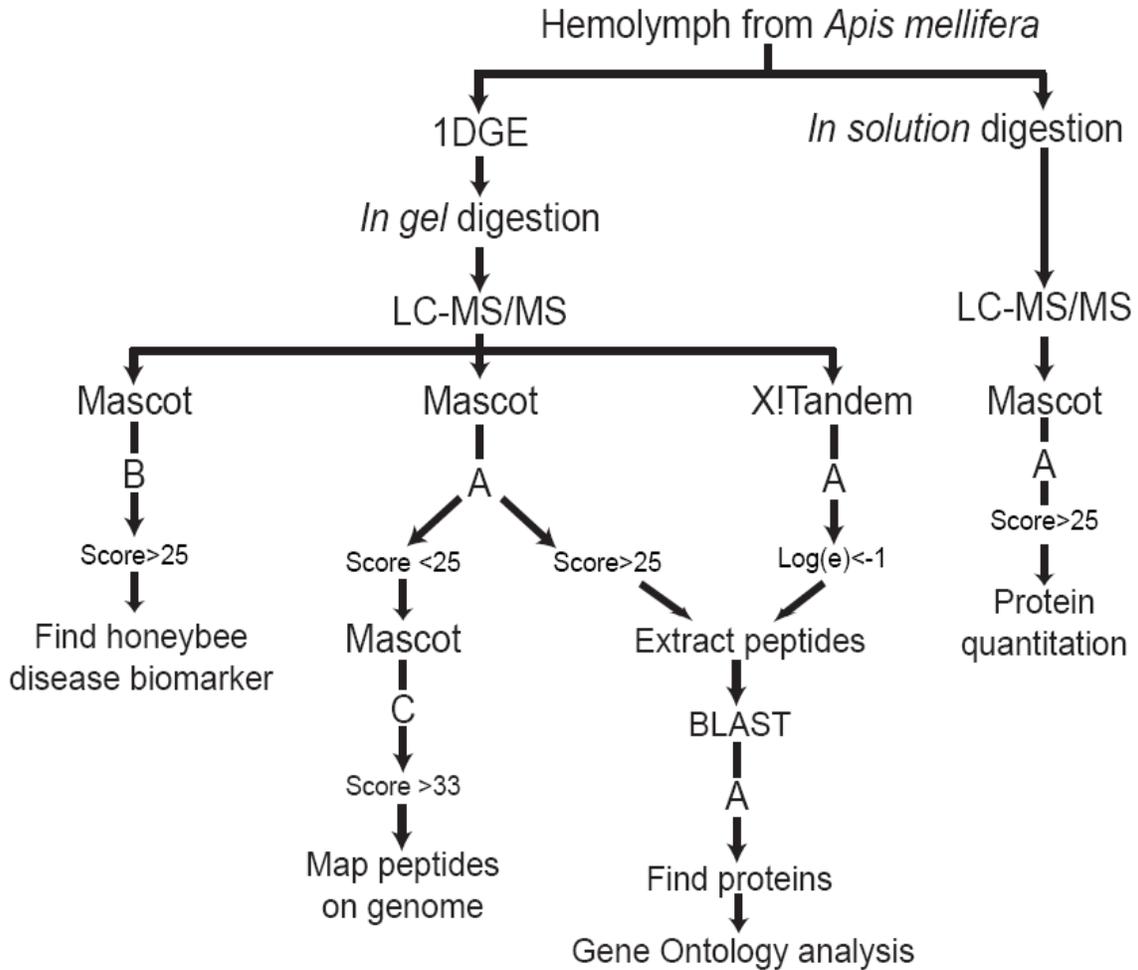
**Figure 2.3 Polyacrylamide gel electrophoresis of honey bee hemolymph**



Equal amounts of protein (30  $\mu\text{g}$ ) of hemolymph was separated on a 4-12% polyacrylamide gel and stained with Blue-silver (c) as described in Experimental Procedures. From left to right: molecular weight markers (kDa), queen (Q), drone (D), worker (W), and worker larvae (WL). Among the proteins of highest abundance are vitellogenin (VT), apolipoprotein precursor (AP), and hexamerins (HX), which are indicated by arrows adjacent to their abbreviations.

Intensely stained bands were likely to be the main previously characterized proteins in insect hemolymph, vitellogenin, apolipoprotein, and hexamerins, but the entire lane of each sample was excised and divided into slices for analysis by LC-MS/MS, in parallel with in-solution digests of each sample. The data were analyzed using two search engines and a cascade of successive searches against various databases, as summarized in Figure 2.4.

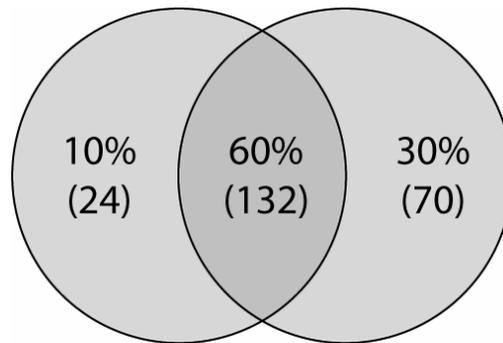
Figure 2.4 Schematic of data analysis procedures following hemolymph digestion



Results were searched against 3 different databases: A, proteins predicted from Amel\_2.0 genome assembly (9759 sequences); B, MSDB database from Mascot, limited to viruses only; C, Amel\_2.0 genome assembly.

All data obtained from the QSTAR were searched against the latest honey bee protein sequences using both Mascot and X!Tandem to obtain greater coverage. It has been reported by others (Kapp et al., 2005) that Mascot reported more protein hits than X!Tandem (Figure 2.5) for a given set of spectra, but the two methods were largely congruent (Whitten et al., 2005).

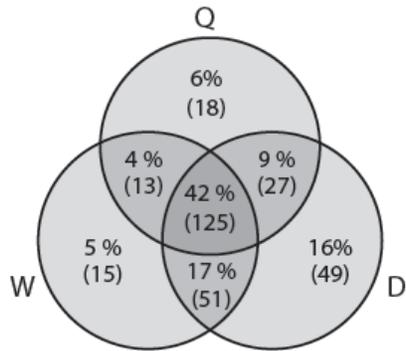
**Figure 2.5 X!Tandem versus Mascot**



Fragment spectra acquired on the QSTAR were searched against the *A. mellifera* protein database by both programs, using a peptide significance cutoff of ion score >25 and "require bold red" for Mascot, and  $\log(e) < -1$  for X!Tandem. A minimum of 2 peptides per protein from either search was required for a significant hit.

By visual inspection, those spectra assigned by only one of the search engines (see Experimental Procedures) appeared to be good matches so the reasons they were not picked up by the other engine were not clear. In total we identified 3657 unique peptides (Supplementary Table 1 in Appendix A) satisfying the criteria for significance (see Experimental Methods). After excluding peptides for contaminant keratins and digestion enzymes, as well as matches to highly identical proteins, 324 unique *A. mellifera* proteins were considered identified. Using the reversed database method (Steen et al., 2004), the probability of false-positive protein hits using these criteria was calculated at 0.5%, or 1.7 ~ 2 proteins. The sequence coverage of each protein hit (Supplementary Table 2 in Appendix A) gives a semi-quantitative measure of the level of each protein in hemolymph (Finehout et al., 2003, Ishihama et al., 2005) and in agreement with band intensities from 1DGE, proteins with the greatest coverage are vitellogenin, apolipoporphin, and hexamerins. The overall overlap between proteins expressed in adult queens, drones, and workers (Figure 2.6) was only 42% and fewer proteins were found in hemolymph from female castes (183 for queens, 204 for workers) than from drones (252).

**Figure 2.6 Number of proteins in adult hemolymph**



The number of proteins exceeding the significance criteria (see "Experimental Procedures") identified in each of queen (Q), drone (D), and worker (W) is shown.

All proteomic data described here can be found in on-line supplementary material and on our website at <http://foster.nce.ubc.ca/bee/><sup>4</sup>.

### 2.4.3 Honey bee biomarkers

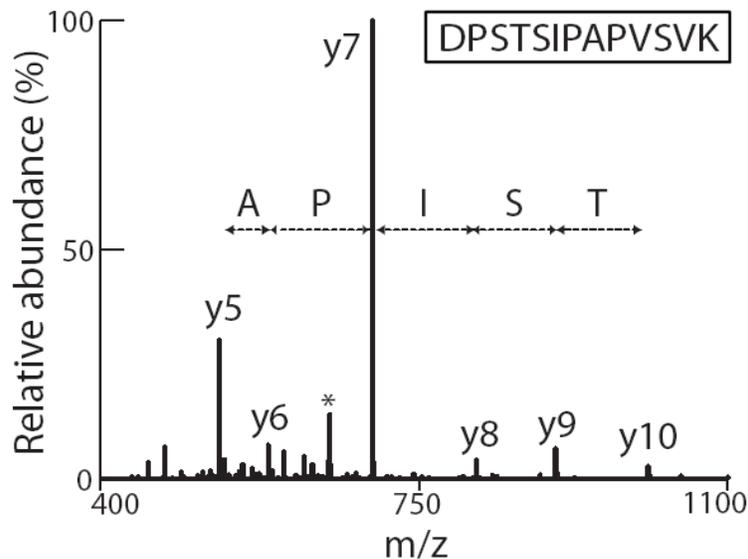
Proteomics holds great promise for identifying indicator proteins that would allow the easy diagnosis of a disease state. To date efforts towards identifying such biomarkers in mammalian systems have fallen flat (Duncan et al., 2005) but there is no reason to expect that biomarkers might not be found in other species. Towards this end, fragment spectra from hemolymph samples were also searched against all viral sequences and one drone sample returned a very strong match for the DWV polyprotein (AAP49283), a virus that can infect the honey bee asymptotically. Although the protein has a nominal mass of 332 kDa, peptides were detected across nearly the entire length of the gel lane, consistent with reports of autocleavage (Palmenberg, 1987). Seven peptides meeting our acceptance criteria (Experimental Procedures and Supplementary Table 3 in Appendix A) were fragmented, accounting for 2% sequence coverage, one of which (Figure 2.7) was critical to differentiate DWV from a virus with 95% identity and an identical length – the varroa destructor virus (VDV) (AAP51418 (Ongus et al., 2004)). VDV infects a honey bee mite *Varroa jacobsoni* (Yue et al., 2005)

---

<sup>4</sup> This web address is not shown in the original published manuscript. It has been updated to reflect the current location of the database.

and not honey bees directly so it was unlikely that VDV would be in the hemolymph sample but this discriminating peptide solidifies this conclusion.

**Figure 2.7** Fragment spectrum of a peptide derived from a DMV polyprotein



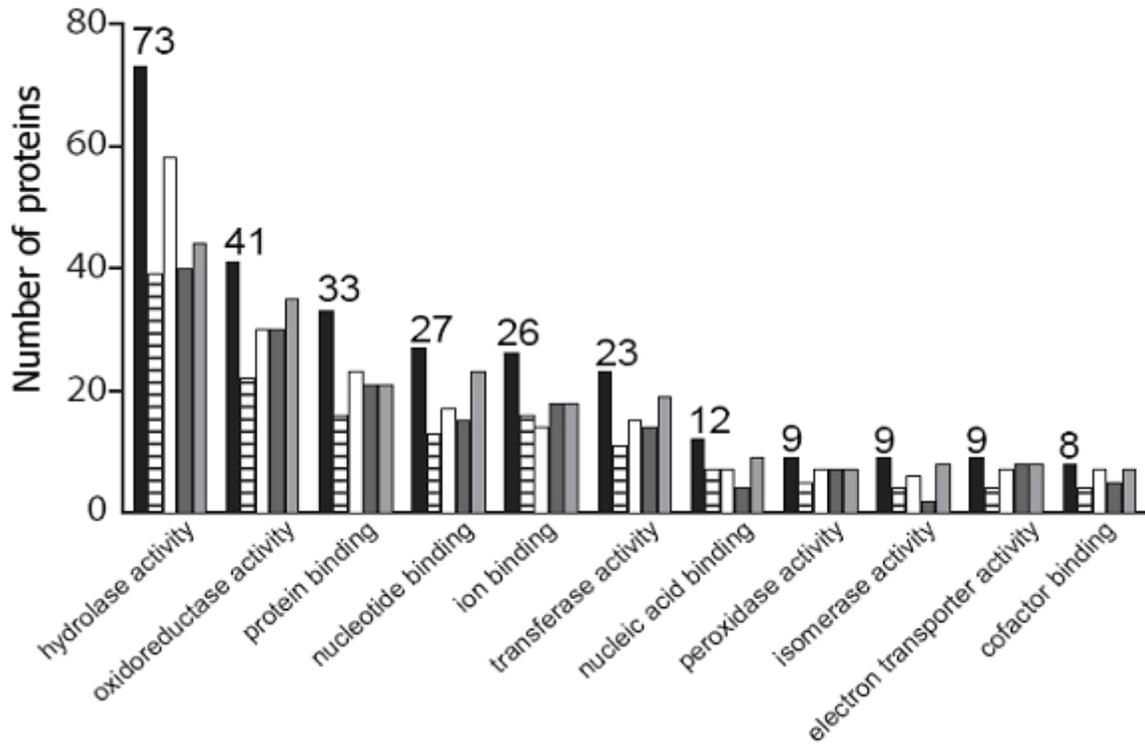
Relative abundance (y axis) of several y ions are shown of the fragment spectrum of peptide sequence DPSTSIAPVSVK. The remaining parent ion peak is marked with an asterisk. The sequencing of this peptide enabled the identification of DWV polyprotein instead of the highly identical VDV polyprotein.

#### 2.4.4 Gene ontology

The assignment of putative function to proteins by GO helps the interpretation of large proteomics datasets, especially for an organism with limited available knowledge about individual proteins such as *A. mellifera*. GO terms for hemolymph proteins were assigned using Blast2GO, resulting in 73% of the proteins being assigned at least one GO term. This effort helped rationalize the presence of proteins with specialized function with relation to caste-specific behaviors, growth and development, and immune functions (Supplementary Table 4 in Appendix A).

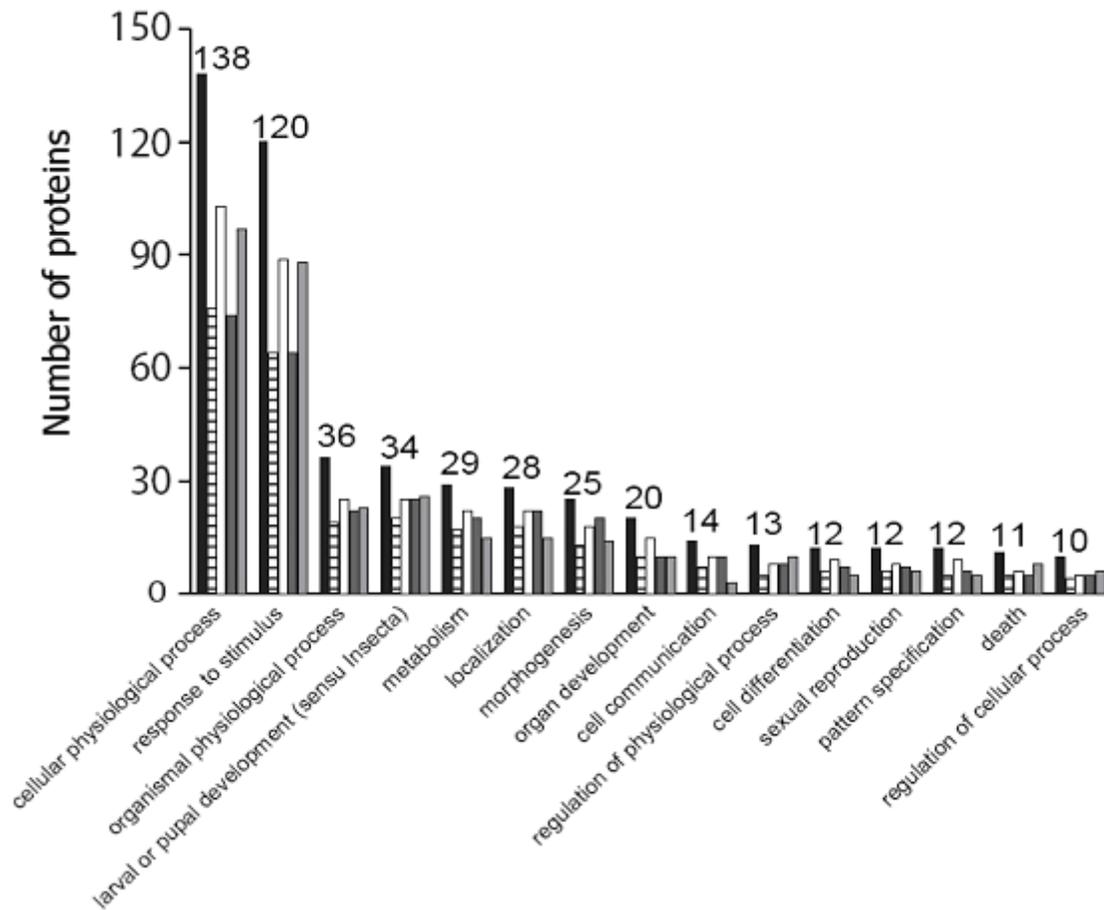
We used Blast2GO to organize the GO term assignment(s) of each hemolymph protein into a directed acyclic graph, which groups GO terms that have highly-specific meanings under terms with broader meanings. We selected terms at the third level of these graphs and plotted the number of proteins under each Molecular Function term (Figure 2.8) and Biological Activity term (Figure 2.9). Details can be found in Supplementary Table 5 and Supplementary Table 6 in Appendix A, respectively.

Figure 2.8 Protein categorization by Molecular Function GO terms



Hemolymph proteins were sorted by third level terms of Molecular Function. Bars represent the number of proteins in each term found in each type of bee: all bees (black, with number of proteins for that term), queen (horizontal lines), drone (white), worker adult (dark gray), and larval (light gray).

Figure 2.9 Protein categorization by Biological Activity GO terms



Hemolymph proteins were sorted by third level terms of Biological Activity. Bars represent the number of proteins in each term found in each type of bee: all bees (black, with number of proteins for that term), queen (horizontal lines), drone (white), worker adult (dark gray), and larvae (light gray).

Because GO assignments to Cellular Component could not be made for the majority of proteins, this ontology was not explored further. Approximately 62% of all proteins could be incorporated into at least one of these graphs. Third-level terms that are most represented are cellular physiological process (GO:0050875) and metabolism (GO:0008152), though more than three-quarters of the proteins in each are found under both terms. Adult drones and worker larvae appear to have an excess of proteins with these biological functions, but this may be due to the overall greater protein variety detected in these bees. Drones also seemed to have many more proteins with hydrolase activity (GO:0016787). The reduced quantity of major hemolymph proteins such as vitellogenin and apolipoprotein in drones effectively narrowed the dynamic range

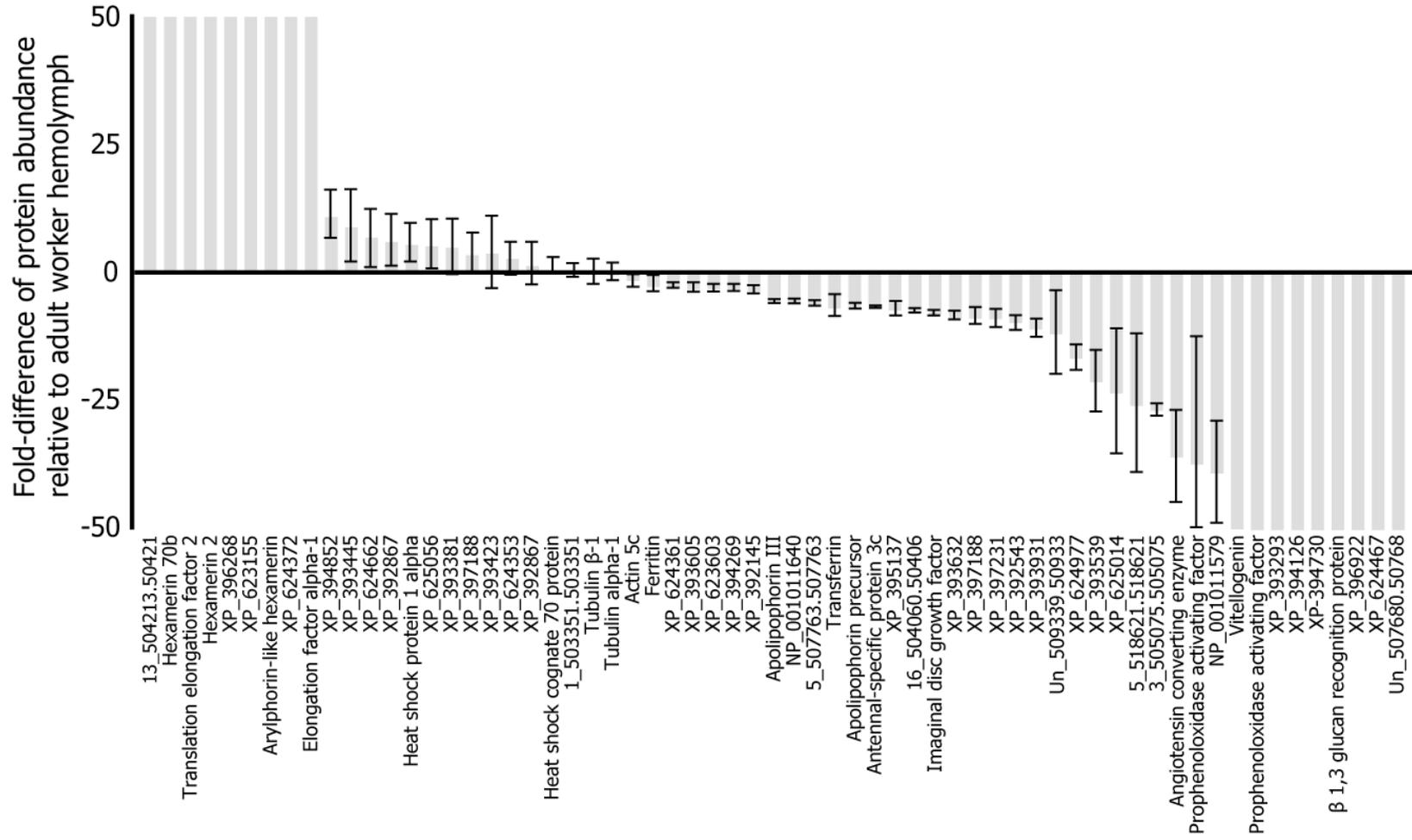
of protein concentration, which likely resulted in increased sensitivity to lower-abundance proteins.

#### **2.4.5 Quantitative differences in hemolymph**

Electrophoretic separation of hemolymph proteins revealed some striking differences between castes and adults versus larvae. For proteins that are too low in abundance to be visualized by gel-staining methods, sequence coverage can serve as a useful estimation of their levels. For example, we were able to see interesting differences in the expression of proteins associated with pheromone response between castes. The antennal-specific protein 3c (NP\_001011583) is an odorant-binding protein (OBP) that we identified in all bees but with far greater sequence coverage in drones and workers (48%) compared to queens (8.5%). Another OBP (XP\_624854) also appeared to be more abundant in drones (22% sequence coverage) and workers (20%) over queens (5%). Also, we found that the honey bee antibacterial peptide hymenoptaecin (NP\_001011615, (Casteels et al., 1993)), was expressed in lower amounts in the larvae compared to the adults (sequence coverage: larvae – 10% versus adult queens and workers – 30%, drones – 16%).

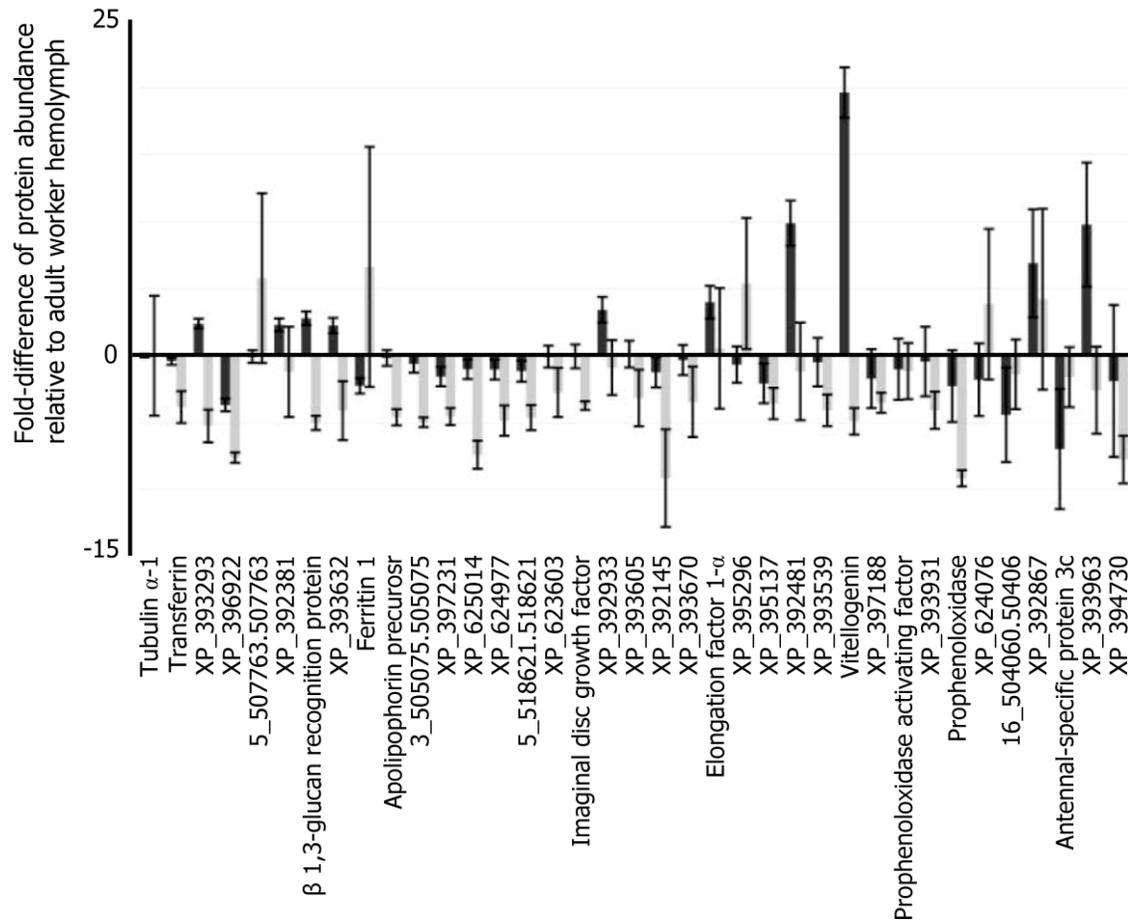
However, in order to more rigorously quantify differences in protein abundance between castes and stages, in-solution digests of hemolymph were analyzed using the extremely high resolving power and accuracy of an LTQ-Orbitrap system. Relative peptide ion intensities in parallel LC-MS/MS analyses were the basis for quantitation, as we have reported previously (Foster et al., 2005, Foster et al., 2006). Normalization of results to a housekeeping protein, such as tubulin  $\alpha$ -1 used here, was critical to control for potentially biased quantitation in a complex protein mixture with one or two dominant proteins. We were able to compare relative expression levels for 63 worker adult and larva proteins (Figure 2.10) and 37 proteins in the adult castes (Figure 2.11) using this method.

Figure 2.10 Quantitation of protein expression comparing worker larvae and adults



Relative quantities of proteins (x-axis) are shown, where a positive value indicates higher expression in larvae hemolymph, and a negative value denotes higher expression in adult hemolymph.

Figure 2.11 Quantitation of protein expression comparing adult castes



Relative quantities of proteins (x-axis) are shown, where protein levels in the hemolymph of drones (light bars) and queens (dark bars) are compared against workers. A negative value indicates higher expression in workers, and a positive value denotes higher expression in either drones or queens (depending on bar colour).

The data show that vitellogenin is expressed 20 times more highly in queens than workers, 5 times less in drones than workers, and that adult workers have at least 50-fold more vitellogenin than larvae. These values are in general agreement with a Northern blot study on vitellogenin mRNA in honey bee abdomen (Piulachs et al., 2003). All hexamerins (NP\_001011600, XP\_392869, XP\_624041) were more than 50-fold overexpressed in worker larvae than worker adult, which is similar to results from a Coomassie-stained SDS-polyacrylamide gel electrophoresis of hemolymph from workers of different ages (Cunha et al., 2005). The angiotensin converting enzyme (XP\_393561) involved in a highly conserved pathway to control blood pressure (Bernstein, 2002), was in 36-fold excess in worker adults than larvae. This method of protein quantitation was particularly effective at comparing protein levels in different developmental stages within the worker caste, likely because they share a similar complement of proteins. We were also able to quantify a number of immunity-related proteins, such as prophenoloxidase activating factor (XP\_623150) and prophenoloxidase itself (NP\_001011627), which are more abundant in adult workers compared to larvae by about 30- and at least 50-fold, respectively. The  $\beta$ -1,3-glucan recognition protein (XP\_395368) was also found to be at least 50 times greater in adult workers over larvae.

#### **2.4.6 Unassigned spectra**

The *A. mellifera* genome was only sequenced a year before this study was conducted and there are relatively few groups working on the gene annotation of this organism. The annotation that has been done has largely been based on homology to *D. melanogaster*, combined with modeling against honey bee ESTs, insect cDNAs, and some *ab initio* predictions. Despite these efforts, there are likely to be hundreds if not thousands of unannotated genes left to be identified in the honey bee genome. To this end we used all spectra that Mascot was not able to match to any peptides in the existing protein database and used them to search the genomic sequences directly. This resulted in 958 previously unpredicted peptides (Supplementary Table 7), which we have now mapped onto the genome and have been made available to the community through Ensembl's Distributed Annotation System (DAS, [http://www.ensembl.org/info/data/external\\_data/das/index.html](http://www.ensembl.org/info/data/external_data/das/index.html)), accessible from

Ensembl's honey bee website<sup>5</sup> (see Experimental Procedures for instructions [http://www.ensembl.org/Apis\\_mellifera/index.html](http://www.ensembl.org/Apis_mellifera/index.html)).

## 2.5 Discussion

Honey bees and ants are the two classic model systems for investigating social behavior, but fruit flies, and to a lesser extent mosquitoes, have been the insects of choice for molecular studies. The vast majority of genes in *A. mellifera* are only predicted, meaning that most of the gene products have never been directly observed. The data presented here provides solid experimental validation for a large number of these proteins and goes further to identify other peptides outside of annotated open reading frames that appear to be expressed.

Currently, there are only a handful of honey bee proteins with reports of expression levels during development or in different castes and the proteomic data presented here supports these previous findings. Vitellogenin, a largely female-specific glucolipoprotein (Amdam et al., 2003), was presented as an intensely stained 180 kDa protein band in adult queens and workers and worker larvae (Barchuk et al., 2002, Wheeler et al., 1990) but was also detected by mass spectrometry in drones (Guidugli et al., 2005, Trenczek et al., 1989). Lower levels of vitellogenin in worker larvae relative to the adult female castes agree with previous reports of its expression (Guidugli et al., 2005). On the other hand, hexamerins are a family of nutrient storage proteins that make up the bulk of larval hemolymph, (Danty et al., 1998, Telfer et al., 1991) and the queen was the only caste to retain hexamerin 70b expression during adulthood in quantities comparable to larva, also confirming previous studies (Danty et al., 1998). Apolipophorin was highly expressed to approximately equal levels in all castes. The extremely high levels of vitellogenin, apolipophorin, and hexamerins in females is a considerable hindrance to effective MS analysis of less abundant proteins, analogous to the problems posed by albumin in human serum proteomics (Colantonio et al., 2005). In support of this, the protein variety detected in samples of hemolymph from females was always lowest. From a behavioral standpoint, the number of tasks accomplished by females far exceeds the male honey bees, whose sole purpose is mating, therefore

---

<sup>5</sup> This feature no longer exists.

females might express a greater variety of proteins compared to males. Consequently, the lower protein diversity detected in females compared to the males is most likely rationalized by the problem of insufficient dynamic range of mass spectrometers, and not due to a biological phenomenon. Despite the presence of a few highly abundant proteins that significantly reduced the number of protein hits, 324 proteins in total were identified in this study, which is one of the largest to date for *A. mellifera*. Clearly though, more work is needed to bring proteomic research for this species on par with *Drosophila*, where about 150 hemolymph proteins have been identified from healthy *Drosophila* larvae alone (Guedes Sde et al., 2003, Karlsson et al., 2004, Vierstraete et al., 2003).

GO term mappings allowed putative assignment of immunity-related functions to *A. mellifera* hemolymph proteins, for which there are only a few studies for the major proteins. For example, drone-specific protein (AmeLGuN\_WGA1241\_2.510469.510469.p) with predicted hydrolase activity (GO:0016787) is a homolog of prostasin, a protein found in human seminal fluid (Yu et al., 1995) and the spermatophore of *Tenebrio molitor* (mealworm beetle) (Takiguchi et al., 1992), an organ involved in the transfer of sperm to females. Queen hemolymph was expected to include proteins related to reproduction that are not present in other caste members. One such protein (XP\_395884) was predicted to be involved in chromatin silencing (GO:0006342) and oocyte maturation (GO:0001556). Both adult queen and worker larva hemolymph contained a protein (XP\_392479) with a role in oocyte microtubule cytoskeleton polarization (GO:0048129). As expected, worker larvae hemolymph contained far more enzymes involved in making simple biomolecules than adult workers, including those needed for the synthesis of thymidylate (XP\_624530, GO:0006233, GO:0006235), steroids (XP\_397214, GO:0006694), fatty acids (XP\_396268, GO:0006633), arginine and glutamine (XP\_393888, GO:0006541, GO:0006526). Perhaps a better indication of rapid tissue growth is the presence of an elongation factor (XP\_623682, GO:0006414) sequenced only in larva hemolymph. Synthesis of these basic building blocks of life would be required throughout all stages of development so these proteins may exist in adult bees as well, but just at lower levels.

Sequence coverage is directly correlated with peptide fragmentation efficiency, which in turn depends on the amino acid sequence of the peptides so this method is, at

best, a semiquantitative measure of abundance (Ishihama et al., 2005). As long as sampling of different conditions is consistent though, then sequence coverage is a valid method for comparing the relative abundance of proteins in different samples, especially since it is available to all tandem mass spectrometers without any addition preparative work to the samples or complex data analysis software. Relative peptide ion intensities, on the other hand, infer protein quantities by directly comparing identical peptides for a given protein found between samples, relying on high mass accuracy of a parent peptide ion obtained using the LTQ-Orbitrap and highly reproducible LC to find the elution peak for a peptide in a parallel analysis even if it is not sequenced. The precision of this method suffers somewhat since the ionization environment for a given peptide in two different samples will not be identical and so the MS detector response may vary. Therefore, we only considered proteins for quantitation by this method when at least three peptides were observed in each analysis, ignoring all other proteins that did not meet this criteria.

Proteomics-based biomarker hunts in mammalian systems have yet to produce any major breakthroughs (Duncan et al., 2005) so we were surprised at the relative ease with which we were able to identify a *bona fide* 'biomarker' in one drone hemolymph sample, and doubly so since the bees had no outward indications that they carried DWV (i.e., their wings were not deformed). Several highly similar viruses in the DWV family can infect bees but the seven peptides identified in the DWV polypeptide allowed us to diagnose this virus specifically. This positive result suggests that hemolymph might be a useful fluid for the diagnosis of other honey bee pathogens with much larger impact on apiculture, such as American Foulbrood. In reality, mass spectrometers are far too expensive and their use too specialized for them to become commonplace tools for even large apiaries but they could be used in large testing centers such as those run by various governmental departments of agriculture.

An initial goal of this study was to identify the molecular components of the honey bee immune system, which is expected to be highly similar to *D. melanogaster* since both insects are under the Superorder Endopterygota. Indeed, honey bee hemolymph was found to contain at least two proteins of the prophenoloxidase pathway (prophenoloxidase activating factor and prophenoloxidase), which results in melanin synthesis that leads to black pigmentation, and is a crucial part of the encapsulation

response against microorganisms and parasites (Zufelato et al., 2004), a defense mechanism common to insects and arthropods. Lower levels of prophenoloxidase in larva reported here may be rationalized by the incomplete development of its immune system and may be at least partially responsible for the complete lack of black coloration in all larval tissues. A homolog of the *Drosophila*  $\beta$ -1,3-glucan recognition protein, a cell membrane protein anchored by a glycosylphosphatidylinositol group (Gobert et al., 2003, Kim et al., 2000), responsible for binding glucan on the surface of yeast, bacteria, and fungi (Rosenberger, 1976), was also found in *A. mellifera* (XP\_395368). Furthermore, two proteins (AmelGUn\_WGA635\_2.509004.509004.p 509004, XP\_395941) were found to be homologous to the *Drosophila* peptidoglycan recognition protein, where its binding to the surface of Gram-positive bacteria is necessary for Toll activation (Michel et al., 2001). Neither protein could be quantitated, but interestingly, the latter protein was not identified at all in worker larvae hemolymph. Hymenoptaecin (NP\_001011615), an antibacterial polypeptide (Casteels et al., 1993), is expressed in lower amounts in larvae compared to the adults. One protein in our dataset,  $\alpha$ -2-macroglobulin (XP\_392454), showed homology to *Drosophila* thioester-containing protein (TEP)-2 (CAB87808) (Lagueux et al., 2000) and to ENSANGP00000019522 (AAG00600) of *A. gambiae* (Levashina et al., 2001), which is a known component of the insect complement system (Blandin et al., 2004). The honey bee  $\alpha$ -2-macroglobulin is a large protein (1777 residues) but was only identified by three peptides, hinting that it is probably not very abundant. This suggests that other members of the TEP family may exist in the hemolymph but are difficult to detect. In general, it is apparent that adult workers have more immunity-associated proteins compared to its larvae, which is unsurprising given the developmental stage differences and this suggests a likely cause for the vulnerability of larvae to pathogens that do not infect adults, such as *Ascospaera apis* (chalkbrood) and *Paenibacillus larvae* (American foulbrood). To confirm such speculations, proteomics could be used to compare hemolymph from healthy versus immune-challenged honey bees. Similar studies have already been done in *D. melanogaster* using a variety of bacteria, fungal, and other challenges such as lipopolysaccharide and peptidoglycan (de Morais Guedes et al., 2005, Levy et al., 2004, Vierstraete et al., 2004, Vierstraete et al., 2005), in *B. mori* (silkworm) after LPS challenge (Wang et al., 2004), and in *A. gambiae* in response to LPS, *Escherichia coli*, and several small proteins (Han et al., 1999).

## 2.6 Conclusions

Now that the honey bee genome has been sequenced an avalanche of data similar to that seen in other model systems will likely follow for bees. Here, our efforts to identify the protein repertoire of hemolymph are one of the first direct applications of this vast and largely unexplored genomic data. Through quantitative analysis of different castes and life stages our data poses several plausible and testable hypotheses to explain behavior at a molecular level. The current study provides an in-depth view of the composition of honey bee hemolymph that will open new avenues for biochemical analysis of this most beneficial insect.

## 2.7 References

- MSQuant [<http://msquant.sourceforge.net>]
- Honey - National Agricultural Statistics Service, Agricultural Statistics Board, U.S. Department of Agriculture  
[<http://usda.mannlib.cornell.edu/reports/nassr/other/zho-bb/hony0205.txt>]
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, 25(17):3389-3402.
- Amdam GV, Norberg K, Hagen A, Omholt SW. **Social exploitation of vitellogenin.** *Proc Natl Acad Sci U S A* 2003, 100(4):1799-1802.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al.* **Gene ontology: tool for the unification of biology.** The Gene Ontology Consortium. *Nat Genet* 2000, 25(1):25-29.
- Barchuk AR, Bitondi MM, Simoes ZL. **Effects of juvenile hormone and ecdysone on the timing of vitellogenin appearance in hemolymph of queen and worker pupae of *Apis mellifera*.** *J Insect Sci* 2002, 2:1.
- Bernstein KE. **Two ACEs and a heart.** *Nature* 2002, 417(6891):799-802.
- Blandin S, Levashina EA. **Thioester-containing proteins and insect immunity.** *Mol Immunol* 2004, 40(12):903-908.
- Candiano G, Bruschi M, Musante L, Santucci L, Ghiggeri GM, Carnemolla B, Orecchia P, Zardi L, Righetti PG. **Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis.** *Electrophoresis* 2004, 25(9):1327-1333.
- Casteels P, Ampe C, Jacobs F, Tempst P. **Functional and chemical characterization of Hymenoptaecin, an antibacterial polypeptide that is infection-inducible in the honeybee (*Apis mellifera*).** *J Biol Chem* 1993, 268(10):7044-7054.
- Colantonio DA, Dunkinson C, Bovenkamp DE, Van Eyk JE. **Effective removal of albumin from serum.** *Proteomics* 2005, 5(15):3831-3835.
- 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, 21(18):3674-3676.
- Craig R, Beavis RC. **TANDEM: matching proteins with tandem mass spectra.** *Bioinformatics* 2004, 20(9):1466-1467.
- Cunha AD, Nascimento AM, Guidugli KR, Simoes ZL, Bitondi MM. **Molecular cloning and expression of a hexamerin cDNA from the honey bee, *Apis mellifera*.** *J Insect Physiol* 2005, 51(10):1135-1147.
- Danty E, Arnold G, Burmester T, Huet JC, Huet D, Pernollet JC, Masson C. **Identification and developmental profiles of hexamerins in antenna and hemolymph of the honeybee, *Apis mellifera*.** *Insect Biochem Mol Biol* 1998, 28(5-6):387-397.
- de Hoog CL, Foster LJ, Mann M. **RNA and RNA binding proteins participate in early stages of cell spreading through spreading initiation centers.** *Cell* 2004, 117(5):649-662.
- de Morais Guedes S, Vitorino R, Domingues R, Tomer K, Correia AJ, Amado F, Domingues P. **Proteomics of immune-challenged *Drosophila***

- melanogaster larvae hemolymph.** *Biochem Biophys Res Commun* 2005, 328(1):106-115.
- Duncan MW, Hunsucker SW. **Proteomics as a tool for clinically relevant biomarker discovery and validation.** *Exp Biol Med (Maywood)* 2005, 230(11):808-817.
- Evans JD, Wheeler DE. **Expression profiles during honeybee caste determination.** *Genome Biol* 2001, 2(1):RESEARCH0001.
- Finehout EJ, Lee KH. **Comparison of automated in-gel digest methods for femtomole level samples.** *Electrophoresis* 2003, 24(19-20):3508-3516.
- Foster LJ, De Hoog CL, Mann M. **Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors.** *Proc Natl Acad Sci U S A* 2003, 100(10):5813-5818.
- Foster LJ, Zeemann PA, Li C, Mann M, Jensen ON, Kasseem M. **Differential expression profiling of membrane proteins by quantitative proteomics in a human mesenchymal stem cell line undergoing osteoblast differentiation.** *Stem Cells* 2005, 23(9):1367-1377.
- Foster LJ, de Hoog CL, Zhang Y, Zhang Y, Xie X, Mootha VK, Mann M. **A Mammalian Organelle Map by Protein Correlation Profiling.** *Cell* 2006, 125(1):187-199.
- Gobert V, Gottar M, Matskevich AA, Rutschmann S, Royet J, Belvin M, Hoffmann JA, Ferrandon D. **Dual activation of the Drosophila toll pathway by two pattern recognition receptors.** *Science* 2003, 302(5653):2126-2130.
- Guedes Sde M, Vitorino R, Tomer K, Domingues MR, Correia AJ, Amado F, Domingues P. **Drosophila melanogaster larval hemolymph protein mapping.** *Biochem Biophys Res Commun* 2003, 312(3):545-554.
- Guidugli KR, Piulachs MD, Belles X, Lourenco AP, Simoes ZL. **Vitellogenin expression in queen ovaries and in larvae of both sexes of Apis mellifera.** *Arch Insect Biochem Physiol* 2005, 59(4):211-218.
- Han YS, Chun J, Schwartz A, Nelson S, Paskewitz SM. **Induction of mosquito hemolymph proteins in response to immune challenge and wounding.** *Dev Comp Immunol* 1999, 23(7-8):553-562.
- Hoffmann JA. **The immune response of Drosophila.** *Nature* 2003, 426(6962):33-38.
- Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. **Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.** *Mol Cell Proteomics* 2005, 4(9):1265-1272.
- Kapp EA, Schutz F, Connolly LM, Chakel JA, Meza JE, Miller CA, Fenyo D, Eng JK, Adkins JN, Omenn GS *et al.* **An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis.** *Proteomics* 2005, 5(13):3475-3490.
- Karlsson C, Korayem AM, Scherfer C, Loseva O, Dushay MS, Theopold U. **Proteomic analysis of the Drosophila larval hemolymph clot.** *J Biol Chem* 2004, 279(50):52033-52041.
- Kim YS, Ryu JH, Han SJ, Choi KH, Nam KB, Jang IH, Lemaitre B, Brey PT, Lee WJ. **Gram-negative bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and beta-1,3-glucan that mediates the signaling for the induction of innate immune genes in Drosophila melanogaster cells.** *J Biol Chem* 2000, 275(42):32721-32727.

- Kucharski R, Maleszka R. **Transcriptional profiling reveals multifunctional roles for transferrin in the honeybee, *Apis mellifera*.** *J Insect Sci* 2003, 3:27.
- Kucharski R, Maleszka R. **Microarray and real-time PCR analyses of gene expression in the honeybee brain following caffeine treatment.** *J Mol Neurosci* 2005, 27(3):269-276.
- Lagueux M, Perrodou E, Levashina EA, Capovilla M, Hoffmann JA. **Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of *Drosophila*.** *Proc Natl Acad Sci U S A* 2000, 97(21):11427-11432.
- Lai R, Lomas LO, Jonczyk J, Turner PC, Rees HH. **Two novel non-cationic defensin-like antimicrobial peptides from haemolymph of the female tick, *Amblyomma hebraeum*.** *Biochem J* 2004, 379(Pt 3):681-685.
- Lavine MD, Strand MR. **Insect hemocytes and their role in immunity.** *Insect Biochem Mol Biol* 2002, 32(10):1295-1309.
- Levashina EA, Moita LF, Blandin S, Vriend G, Lagueux M, Kafatos FC. **Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*.** *Cell* 2001, 104(5):709-718.
- Levy F, Rabel D, Charlet M, Bulet P, Hoffmann JA, Ehret-Sabatier L. **Peptidomic and proteomic analyses of the systemic immune response of *Drosophila*.** *Biochimie* 2004, 86(9-10):607-616.
- Lin H, Winston ML, Haunerland NH, Slessor KN. **Influence of age and population size on ovarian development, and of trophallaxis on ovarian development and vitellogenin titres of queenless worker honey bee (*Hymenoptera: Apidae*).** *Can Entomol* 1999, 131(5):695-706.
- Lourenco AP, Zufelato MS, Bitondi MM, Simoes ZL. **Molecular characterization of a cDNA encoding prophenoloxidase and its expression in *Apis mellifera*.** *Insect Biochem Mol Biol* 2005, 35(6):541-552.
- Michel T, Reichhart JM, Hoffmann JA, Royet J. ***Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein.** *Nature* 2001, 414(6865):756-759.
- Morse RA, Calderone NW. **The value of honey bees as pollinators of US crops in 2000.** *Bee culture* 2000, 128:1-15.
- Olsen JV, Ong SE, Mann M. **Trypsin cleaves exclusively C-terminal to arginine and lysine residues.** *Mol Cell Proteomics* 2004, 3(6):608-614.
- Ongus JR, Peters D, Bonmatin JM, Bengsch E, Vlak JM, van Oers MM. **Complete sequence of a picorna-like virus of the genus *Iflavirus* replicating in the mite *Varroa destructor*.** *J Gen Virol* 2004, 85(Pt 12):3747-3755.
- Palmenberg AC. **Picornaviral processing: some new ideas.** *J Cell Biochem* 1987, 33(3):191-198.
- Paskewitz SM, Shi L. **The hemolymph proteome of *Anopheles gambiae*.** *Insect Biochem Mol Biol* 2005, 35(8):815-824.
- 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.** *Biochim Biophys Acta* 2005, 1752(1):1-5.
- Piulachs MD, Guidugli KR, Barchuk AR, Cruz J, Simoes ZL, Belles X. **The vitellogenin of the honey bee, *Apis mellifera*: structural analysis of the cDNA and expression studies.** *Insect Biochem Mol Biol* 2003, 33(4):459-465.

- Rappsilber J, Ishihama Y, Mann M. **Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics.** *Anal Chem* 2003, 75(3):663-670.
- Robbs SL, Ryan RO, Schmidt JO, Keim PS, Law JH. **Lipophorin of the larval honeybee, *Apis mellifera* L.** *J Lipid Res* 1985, 26(2):241-247.
- Rosenberger R. **The filamentous fungi**, vol. 2. London: Edward Arnold; 1976.
- Scarselli R, Donadio E, Giuffrida MG, Fortunato D, Conti A, Balestreri E, Felicioli R, Pinzauti M, Sabatini AG, Felicioli A. **Towards royal jelly proteome.** *Proteomics* 2005, 5(3):769-776.
- Schafer MO, Dietemann V, Pirk CW, Neumann P, Crewe RM, Hepburn HR, Tautz J, Crailsheim K. **Individual versus social pathway to honeybee worker reproduction (*Apis mellifera*): pollen or jelly as protein source for oogenesis?** *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 2006, 192(7):761-768.
- Shevchenko A, Wilm M, Vorm O, Mann M. **Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels.** *Anal Chem* 1996, 68(5):850-858.
- Shipman BA, Ryan RO, Schmidt JO, Law JH. **Purification and properties of a very high density lipoprotein from the hemolymph of the honeybee *Apis mellifera*.** *Biochemistry* 1987, 26(7):1885-1889.
- Steen H, Mann M. **The ABC's (and XYZ's) of peptide sequencing.** *Nat Rev Mol Cell Biol* 2004, 5(9):699-711.
- Takiguchi M, Niimi T, Su ZH, Yaginuma T. **Trehalase from male accessory gland of an insect, *Tenebrio molitor*. cDNA sequencing and developmental profile of the gene expression.** *Biochem J* 1992, 288:19-22.
- Telfer WH, Kunkel JG. **The function and evolution of insect storage hexamers.** *Annu Rev Entomol* 1991, 36:205-228.
- Trenczek T, Zillikens A, Engels W. **Developmental patterns of vitellogenin haemolymph titre and rate of synthesis in adult drone honey bees (*Apis mellifera*).** *Journal of Insect Physiology* 1989, 35(6):475-481.
- Vierstraete E, Cerstiaens A, Baggerman G, Van den Bergh G, De Loof A, Schoofs L. **Proteomics in *Drosophila melanogaster*: first 2D database of larval hemolymph proteins.** *Biochem Biophys Res Commun* 2003, 304(4):831-838.
- Vierstraete E, Verleyen P, Baggerman G, D'Hertog W, Van den Bergh G, Arckens L, De Loof A, Schoofs L. **A proteomic approach for the analysis of instantly released wound and immune proteins in *Drosophila melanogaster* hemolymph.** *Proc Natl Acad Sci U S A* 2004, 101(2):470-475.
- Vierstraete E, Verleyen P, Sas F, Van den Bergh G, De Loof A, Arckens L, Schoofs L. **The instantly released *Drosophila* immune proteome is infection-specific.** *Biochem Biophys Res Commun* 2004, 317(4):1052-1060.
- Vierstraete E, Verleyen P, A DEL, Schoofs L. **Differential proteomics for studying *Drosophila* immunity.** *Ann N Y Acad Sci* 2005, 1040:504-507.
- Wang Y, Zhang P, Fujii H, Banno Y, Yamamoto K, Aso Y. **Proteomic studies of lipopolysaccharide-induced polypeptides in the silkworm, *Bombyx mori*.** *Biosci Biotechnol Biochem* 2004, 68(8):1821-1823.
- Wheeler DE, Kawooya JK. **Purification and characterization of honey bee vitellogenin.** *Arch Insect Biochem Physiol* 1990, 14(4):253-267.

- Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson HM, Soares MB, Robinson GE. **Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee.** *Genome Res* 2002, 12(4):555-566.
- Whitfield CW, Cziko AM, Robinson GE. **Gene expression profiles in the brain predict behavior in individual honey bees.** *Science* 2003, 302(5643):296-299.
- Whitten DA, Larson M, Ngowe C, Wilkerson CG, Phinney BS. **A comparison between two tandem mass spectrometry search engines - Mascot vs. X!Tandem.** In: *American Society for Mass Spectrometry: 2005; San Antonio, Texas, 2005.*
- Winston ML. **The Biology of the Honey Bee.** Cambridge, MA: Harvard University Press; 1987.
- Yu JX, Chao L, Chao J. **Molecular cloning, tissue-specific expression, and cellular localization of human prostatic acid phosphatase mRNA.** *J Biol Chem* 1995, 270(22):13483-13489.
- Yue C, Genersch E. **RT-PCR analysis of Deformed wing virus in honeybees (*Apis mellifera*) and mites (*Varroa destructor*).** *J Gen Virol* 2005, 86(Pt 12):3419-3424.
- Zhao XF, He HJ, Dong DJ, Wang JX. **Identification of differentially expressed proteins during larval molting of *Helicoverpa armigera*.** *J Proteome Res* 2006, 5(1):164-169.
- Zufelato MS, Lourenco AP, Simoes ZL, Jorge JA, Bitondi MM. **Phenoloxidase activity in *Apis mellifera* honey bee pupae, and ecdysteroid-dependent expression of the prophenoloxidase mRNA.** *Insect Biochem Mol Biol* 2004, 34(12):1257-1268.

## 3 PROTEIN PROFILES DURING HONEY BEE LARVAL DEVELOPMENT<sup>6</sup>

### 3.1 Abstract

The honey bee (*Apis mellifera*), besides its role in pollination and honey production, serves as a model for studying the biochemistry of development, metabolism, and immunity in a social organism. Here we use mass spectrometry-based quantitative proteomics to quantify nearly 800 proteins during the five- to six-day larval developmental stage, tracking their expression profiles. We report that honey bee larval growth is marked by an age-correlated increase of protein transporters and receptors, as well as protein nutrient stores, while opposite trends in protein translation activity and turnover were observed. Levels of immunity factors prophenoloxidase and apisimin are positively correlated with development, while others surprisingly were not significantly age-regulated, suggesting a molecular explanation for why bees are susceptible to major age-associated, bee-specific bacterial infections such as American Foulbrood or fungal diseases such as chalkbrood. Previously unreported findings include the reduction of antioxidant and G proteins in aging larvae. This data has allowed us to integrate disparate findings in previous studies to build a model of metabolism and maturity of the immune system during larval development. This publicly-accessible resource for protein expression trends will help generate new hypotheses in the increasingly important field of honey bee research.

### 3.2 Introduction

Honey bees (*Apis mellifera*) have been a subject of scientific research for more than 2300 years (Haldane, 1955), yet it is only in the past two decades that bee research has expanded beyond behavioral or social traits to a molecular level. With the publication of the honey bee genome in 2006 (2006), the basic information to enable proteome-level analyses of this organism is now available. Since then various groups

---

<sup>6</sup> A version of this chapter has been accepted for publication. Chan, Q.W. and Foster, L.J. (2008) Changes in protein expression during honey bee larval development. *Genome Biol.* 9(10):R156.

have published proteomic analyses of whole bees or individual organs/tissues (Chan et al., 2006, Scharlaken et al., 2008, Wolschin et al., 2007, Wolschin et al., 2007) but these studies have focused on adult animals. Larval development in honey bees is largely unexplored, despite its significance in caste determination (Patel et al., 2007) and in the pathogenesis of certain economically significant honey bee diseases such as American and European Foulbrood.

The larval development of the honey bee, which follows a three-day period as an egg, is five to six days in duration and precedes the pupal (metamorphosis) and adult stages. Apart from an astounding increase in size, larval growth is relatively unremarkable at the macroscopic level (Nelson, 1924). However, female bees differentiate into workers or queens (caste differentiation) in response to diet very early in larval development and the acquisition of immunity to certain diseases during this five- to six-day period suggests complex molecular biological changes are taking place.

Insect development has been studied mainly using the fruit fly as the model system. *Drosophila* embryogenesis has historically attracted far more attention than any other growth stage, due to its value for studying the mechanism of spatial regulation of transcription and translation. With the exception of the economically important silkworm *Bombyx mori*, research on larval development has been slow. For honey bees, the lack of published works is evident: the article entitled "Morphology of the Honey bee Larva" published by Nelson in 1924 (Nelson, 1924) still remains today as one of the most cited resources on this subject. Here we have used mass spectrometry-based proteomics to profile the changing abundance of individual proteins over the first five days of the worker larval stage and used these data, with the help of sequence-based function prediction, to build a framework for the developmental processes going on in the maturing larva.

### **3.3 Experimental Procedures**

#### **3.3.1 Reagents**

All salts and chemicals were of analytical grade or better and were obtained from Sigma-Aldrich unless otherwise indicated. All solvents were of HPLC grade and were obtained from ThermoFisher Scientific. The following materials were obtained as

indicated: endopeptidase Lys-C, Wako Chemicals (Osaka, Japan); porcine modified trypsin, Promega (Nepean, Ontario, Canada); loose ReproSil-Pur 120 C18-AQ 3  $\mu\text{m}$ , Dr. Maisch (Ammerbuch-Entringen, Germany); 96-well full skirt PCR plates, Axygen (Union City, CA); fused silica capillary tubing, Polymicro (Phoenix, AZ); 5- $\mu\text{l}$  Microcap pipettes for hemolymph collection, Drummond (Broomall, PA); soft forceps for holding bees, BioQuip (Rancho Dominguez, CA); protease inhibitor mixture, Roche Applied Science; precast 4–12%, 1-mm-thick NuPAGE Novex BisTris2 Gels, Invitrogen.

### **3.3.2 Obtaining larvae of known ages**

Honey bee (*A. mellifera ligustica*) larvae were obtained from colonies kept at the University of British Columbia, Vancouver, BC, Canada. Samples were collected in the summer and early autumn. To acquire larvae of known ages, a queen was isolated on an empty frame of dark comb bracketed by two frames approximately 50% filled with honey and pollen for 16 hours inside a nucleus colony with several hundred worker bees. The brood frame with newly-laid eggs was then replaced into the original hive, along with the queen, workers and two supporting frames. The queen was separated from the newly-laid eggs using a queen excluder to prevent additional eggs from being deposited. Three days after reintroducing the eggs into the colonies, larvae were collected for five consecutive days. In this system the maximum error in larval age would be 16 h. Empirical testing with shorter times did not yield enough eggs/larvae to sample the same population over all five days of development. Before proceeding with protein collection, all larvae were washed three times in phosphate buffered saline (PBS) to reduce royal jelly contamination.

### **3.3.3 Protein collection**

For one- to three-day old larvae (day of eclosion will be defined as day one hereinafter), hemolymph was collected by piercing the larval skin, taking care not to cause organ damage by avoiding deep cuts. For four and five-day old larvae, hemolymph was collected by inserting a disposable 5  $\mu\text{l}$  glass Microcap pipette two-thirds of the way down one side of the larva, drawing liquid by capillary action. All hemolymph samples were centrifuged for 10 minutes at 16,100 relative centrifugal force (r.c.f.) at 4°C to pellet cells and debris, which were added to the tissue samples. The tissues were homogenized by a bead mill using a tungsten bead in each 2 ml self-

locking tube (Eppendorf) at 30 hertz for 5 minutes in 50  $\mu$ l of phosphate buffered saline containing a protease inhibitor cocktail tablet solution (Roche) at eight times the suggested concentration. Lysis buffer (100  $\mu$ l of 1% NP-40, 150mM NaCl, 20mM Tris-HCl pH 7.5) was added before the sample was homogenized by ten strokes through a syringe tipped with a 25 G needle. The sample was clarified for ten minutes at 16,100 r.c.f. at 4°C and the pelleted debris was discarded. The Coomassie Plus Protein Assay (Pierce) was used to determine protein concentrations of the tissue lysates and the clarified hemolymph according to the manufacturer's instructions before they were stored at -20°C until used.

### **3.3.4 Denaturing protein gel electrophoresis**

Tissue and hemolymph proteins were resolved on precast (Invitrogen) 4-12% NuPAGE gels (30  $\mu$ g/lane) in reducing conditions with MES buffer according to manufacturer's instructions. Blue-silver stain (Candiano et al., 2004) was used to visualize protein bands.

### **3.3.5 Sample preparation for mass spectrometry analysis**

Larval tissue or hemolymph protein were aliquoted to provide 20  $\mu$ g per sample before they were precipitated using the ethanol/acetate method as described (Foster et al., 2003). The insoluble proteins were pelleted and temporarily stored at 4°C after a ten-minute centrifugation at 16,100 r.c.f.. The ethanol supernatant was vacuum-dried, solubilized in Sample Buffer (3% acetonitrile, 1% trifluoroacetic acid, 0.5% acetic acid), and purified using the C8 flavor of STop And Go Extraction (STAGE) tips (Rappsilber et al., 2003) to remove contaminants such lipids, nucleic acids, and protease inhibitors. Bound proteins were eluted using 100% acetonitrile and vacuum-dried before adding 0.5  $\mu$ l of 1.5 M Tris-HCl, pH 8.8. The bulk protein pellet and C8 purified proteins were digested using LysC and trypsin as described (Foster et al., 2003). Peptides were desalted using C18 STAGE tips and the eluted solution was dried by vacuum centrifugation. For proteome profiling by relative quantitation, binary analysis between timepoints was performed by chemical dimethylation of peptides from different timepoints using either light ( $\text{CH}_2\text{O}$ ) or heavy ( $\text{CD}_2\text{O}$ ) isotopologues of formaldehyde (Hsu et al., 2003). For both the hemolymph and tissue samples, three-day old larvae was used as a reference for all other timepoints, such that its peptides were always

labeled with the opposing form of formaldehyde from days one, two, four, and five before mixing the differentially labeled samples. Samples were fractionated on C18-SCX-C18 STAGE tips using 10-step ammonium acetate elution gradient (Ishihama et al., 2006) and dried peptide samples were resuspended in 1% trifluoroacetic acid, 3% acetonitrile, 0.5% acetic acid prior to analysis on a linear trapping quadrupole-Orbitrap hybrid mass spectrometer (ThermoFisher Scientific) as described in (Chan et al., 2006).

### **3.3.6 Raw data processing**

Following LC-MS analysis, peak lists were extracted from the raw data using Extract\_MSN.exe (ThermoFisher Scientific) and DTA Supercharge as described (Kwok et al., 2008). Results were searched using Mascot (v2.2) against a database containing the protein sequences of: Honey Bee Official Gene Set 1 (Elsik et al., 2007), common exogenous contaminants (human & sheep keratins) and additives (porcine trypsin, lysyl endopeptidase C), the polyprotein of the deformed wing virus (common and often asymptomatic (Yue et al., 2005)), and the reversed sequences of all of the above as a decoy for reporting false discovery rates. The following Mascot parameters were used: trypsin (allowing up to one missed cleavage) or no enzyme specificity (in separate searches); carbamidomethyl as a fixed modification, variable modifications of dimethylation by both hydrogen isotopes at the peptides' N-termini and lysine  $\epsilon$ -amino groups, 10 parts-per-million (ppm) peptide tolerance; 0.8 Da MS/MS tolerance, and ESI-Trap fragmentation characteristics. Results were saved in Peptide Summary format with the "Require Bold Red" option checked, applying a score cutoff corresponding to  $p < 0.05$ , which is 27 where results were limited to tryptic peptides, and 47 where no enzymes were specified. Since each sample was fractionated, generated files were combined using the in-house script Pickletrimmer.pl. MSQuant was used to semi-automatically extract chromatographic peak volumes in both the light and heavy isotopologues of each detected peptide. Only peptides with an absolute calibrated mass error of  $< 5$  ppm were considered further. For protein quantitation, the file was parsed (in-house script: QC\_msqfa.pl) to obtain natural logarithm (Ln)-transformed heavy/light peptide volume ratios which were median-normalized before they were averaged to calculate a relative protein ratio of day-3 larvae / day-x ( $x=1, 2, 4, 5$ ). From the 3 biological replicates of each binary comparison, proteins quantified with at least 2 quantified peptides from 2 or more replicates were averaged. Proteins whose relative quantities could be tracked for

at least 4 of 5 days in either the tissue or hemolymph were considered to be profiled. For protein identification, the above peptides and unquantified sequences were extracted from MSQuant outputs. After removing redundant entries (in-house script: QC\_remduplicate.pl), each was matched to their respective protein (in-house script: finalist.pl), excluding hits that were verified by equal to or less than 2 peptides of at least 6 or more residues. The false discovery rate was estimated by dividing the sequence-reversed proteins that failed to be eliminated after applying the above criteria.

### **3.3.7 Automated protein annotation to Gene Ontology terms**

All identified proteins were matched to Gene Ontology (Ashburner et al., 2000) terms using BLAST2GO (Conesa et al., 2005), following their standard procedure of performing BLAST searches for each protein (BLASTp, nr database, HSP cutoff length 33, report 20 hits, maximum e-Value 1e-10), followed by mapping and annotation (e-Value hit filter 1e-10, annotation cutoff 55, GO weight 5, HSP-hit coverage cutoff 20). After generating a directed acyclic graph (sequence filter 2, score alpha 0.6, node score filter 0) of Molecular Function terms (not shown), which groups specific terms into broader categories, ontologies on the third level of this graph were further analyzed by statistical testing (see below). The term "protein binding" (GO:0005515) was omitted because this included the most number of proteins, most of which belonged under another more informative term.

### **3.3.8 Semi-automated protein annotation and manual categorization**

Protein descriptions were taken from several sources or tools, all of which are sequence homology-based derivations. Official protein names given in the Official Gene Set 1 (Elsik et al., 2007) were used if the name was informative. BLAST2GO-derived descriptions were used where protein function was not clear from the official name (for example "hypothetical protein"). If an appropriate name was still not derived, searches against the Conserved Domain Database (NCBI) were performed and considered matched for e-Values <1e-10. As a final measure for matching a protein with a functional name, proteins descriptions were copied from Blast2seq (Tatusova et al., 1999) results (accessed via BLink in NCBI) if matches had >25% sequence identity and e-Value of <1e-10 over the aligned region. If none of these steps provided useful

information, the protein was labeled and categorized with “unknown function.” Proteins with descriptions but did not fit under a specific category were classified as “uncategorized” (Supplementary Table 11 in Appendix A). Proteins that were not manually annotated were marked with “N/A” in column 3 of Supplementary Table 7 in Appendix A.

### **3.3.9 Statistical analysis**

To each class of manually assigned proteins, a pairwise, 2-tailed t-test was performed using each protein in that class by taking the relative ratio in d1 and comparing to d5. Groups with  $p < 0.05$  were considered to be temporally-regulated, and their directionality of regulation was calculated by averaging the slopes of individual proteins within a group using d1 and d5 timepoints. Third-level Gene Ontology Molecular Function terms were analyzed in the same manner, except all the proteins considered were quantified over all 5 days tested in at least one of the tissue or hemolymph data. To individual proteins, the same criteria for significance was used, taking values from each biological replicate as a data point in a pairwise comparison between the earliest and latest day the protein was quantified. We also performed average linkage clustering of the proteins expression levels for proteins that were quantified over at least 4 days in either the tissue or hemolymph, using Cluster and visualized by Treeview (Eisen et al., 1998). The grouping sizes ranged from 2 to 55 proteins. To normalize this variation, the number of proteins in a given class is reported as a percentage of the total class size (percent enrichment, using in-house script QC\_nodeenrichment.pl). Only nodes that included at least 50% of all the proteins in that class and had a Pearson’s correlation coefficient of greater than 0.8 were considered to be within the same cluster. Protein families with 3 or fewer members were included as part of the tree diagram, but were not considered for whether they formed a significant cluster.

### **3.3.10 Comparison to Drosophila**

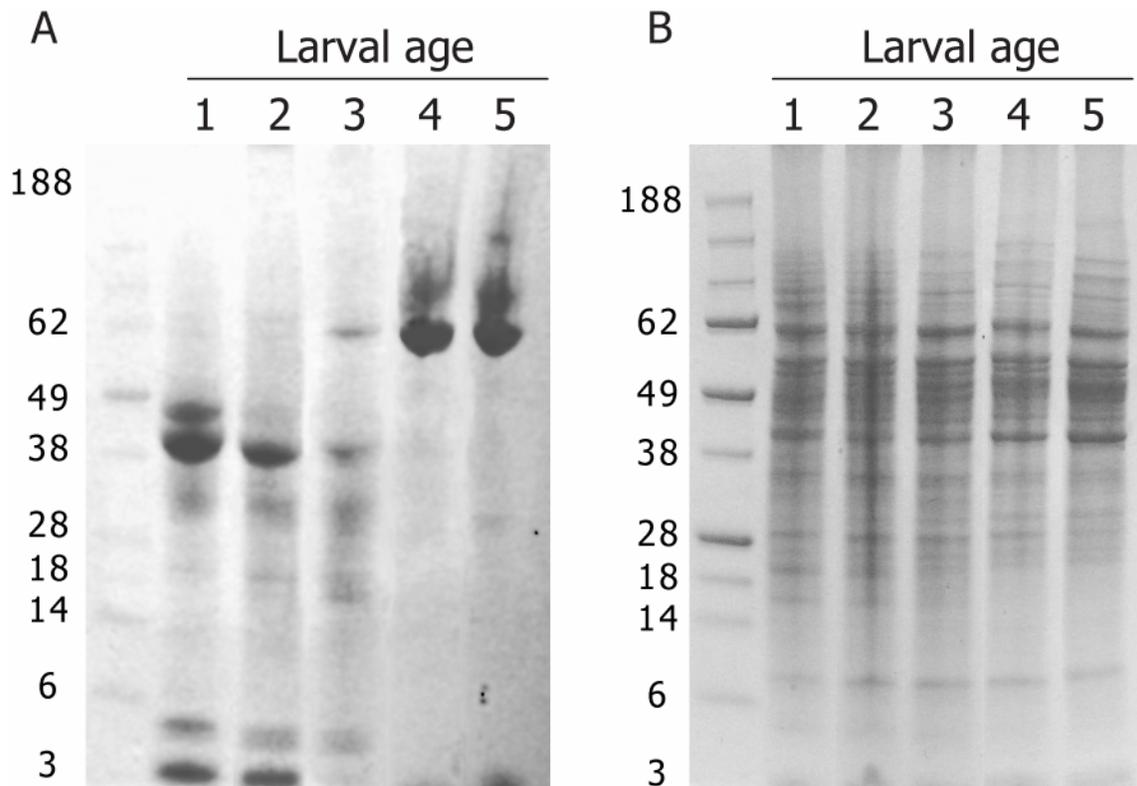
Proteomic profiles resulting from this work were compared to the transcriptomic profiles of previously published *Drosophila* homologs (Arbeitman et al., 2002) for the timepoints matching most closely to days 1 to 4 of the honey bee larval stage (h=24, 49, 72, 96) for genes that were significantly regulated over this period (Fruit fly larval stage is shorter than bees’ by 1 to 2 days). BLASTp was used to find homologs in the Honey

Bee Official Gene Set 1 which were defined as matches having E-values lower than  $1e-10$  with at least 25% identity within the aligned region. Timepoints of the *Drosophila* dataset were normalized to the h=72 timepoint and ln transformed. To compare the expression trend between the two organisms, the slope of the line-of-best-fit for proteins (bees) or genes (flies) was calculated: expression trends with slopes that differed in signage or had an absolute difference of greater than 0.75 were considered to be dissimilar. Slopes whose absolute value of Pearson's correlation coefficient was less than 0.5 were considered insignificant and therefore not considered. In instances where a significant slope could be calculated for a protein in both the tissue and hemolymph samples, the slopes were averaged.

### **3.4 Results**

In order to obtain suitably-aged larval samples for proteomic profiling of the first five days of development, for each experiment we isolated an open-mated, laying queen on an empty frame of brood comb for a short period of time to allow her to lay several hundred eggs (see Experimental Procedures). The frame and queen were then separated by a queen excluder and workers were allowed to tend the brood. Starting on the day the eggs hatched (Day 1, roughly corresponding to first instar) larvae were collected every day for five days. Hemolymph was separated from the remaining tissues (termed 'solid tissues' henceforth) prior to protein extraction (see Experimental Procedures) and equal amounts of protein from each age were resolved on a reducing sodium dodecylsulfate polyacrylamide gel (Figure 3.1).

**Figure 3.1 Polyacrylamide gel electrophoresis of honey bee larvae proteins**



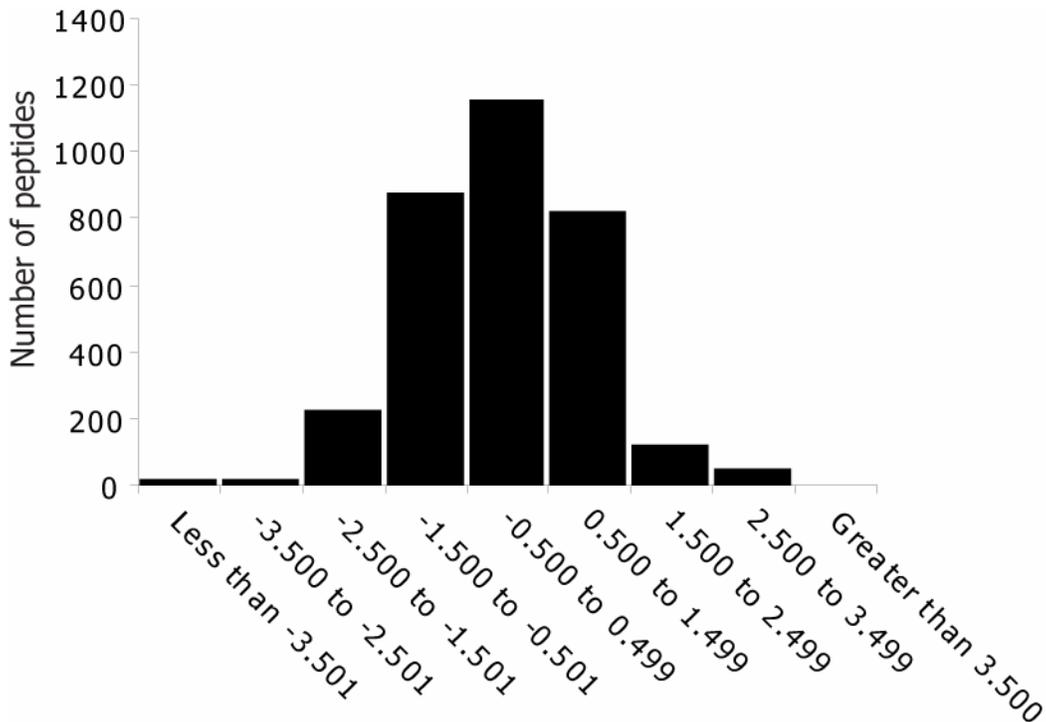
Electrophoretic banding patterns of proteins from larval hemolymph (A) and solid tissue (B) are shown for the first five days of the larval developmental stage.

The protein composition of solid tissues was grossly consistent across all ages, but varied drastically in the hemolymph. Hemolymph from 1 to 3-day old larvae show a staining pattern distinct from that of 4 to 5-day old larvae. These differences may be partially attributed to slight variations in collection methods for young and old larvae but it is more likely that these represent real biological changes occurring as the late larvae prepare for pupation. Most notably, a 70kDa hexamerin band emerges from day 3 and beyond and accounts for the majority of the protein in the hemolymph, an observation that has been made numerous times by other researchers (Cunha et al., 2005, Danty et al., 1998, Korochkina et al., 1997). A second observation that argues against these dramatic changes around day 3 being simply an artifact of sample collection is the absence of the major protein bands in the hemolymph gel in the solid tissue gel, and vice versa.

As a means for identifying and quantifying the expression profiles of proteins in developing larvae, we used a quantitative proteomics approach employing stable isotope

labeling and liquid chromatography-tandem mass spectrometry. The labeling method we used employs deuterated and hydrogenated forms of formaldehyde to reductively dimethylate primary amines in peptides but since there are only two labeling conditions possible in this schema we compared the expression of protein from days 1, 2, 4 and 5 larvae versus day 3 in order to generate an expression profile spanning the whole development period. Three biological replicates of each tissue type were analyzed, which resulted in the detection of 12,421 non-redundant peptides (Supplementary Table 17 in Appendix A). After applying the cutoff criteria (see Experimental Procedures), 1333 proteins were identified (Supplementary Table 8 in Appendix A) with an estimated false discovery rate of 0.97% (see Experimental Procedures), thus providing experimental evidence for 12.7% of the 10517 genes in the predicted honey bee gene set. In general the peptide ratios showed no labeling bias and were approximately normally distributed (Figure 3.2).

**Figure 3.2 Histogram of heavy/light peptide ratios within a sample**



The peptide ratios within an experiment are roughly normally distributed and show no labeling bias. Using replicate number 1 of a day 1 versus day 3 solid tissue quantitation data as an example, the peptide ratios are displayed as a histogram, sorted into natural-log unit bins (bin size = 1).

Among these, 790 were quantified in 2 or more days by averaging the intensity ratio from at least 2 of the 3 replicates (if more than 5 peptides were quantified, the top 5 most intense peptides were selected): 378 (48%) of them matched this criteria in both the tissue and hemolymph, 309 (39%) were specific to solid tissue and 103 (13%) were specific to hemolymph. An example of using peptide ratios to derive relative protein expression profile is shown in Table 3.1 for the odorant binding protein 14 [GenBank:94158822].

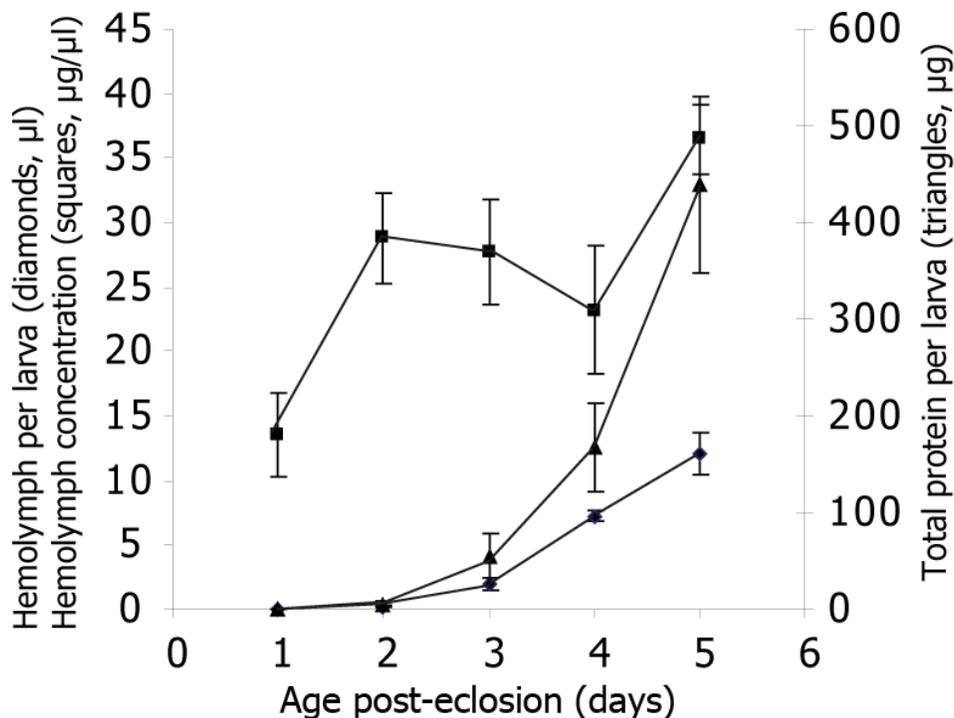
**Table 3.1 Using peptide ratios to derive relative protein expression: an example**

Sample	Larval Ages	Replicate (1, 2, or 3)	Ln(ratio) of Top 5 Most Intense Peptides					Average Ln(ratio)	
			1	2	3	4	5	Peptides	Protein
Hemolymph	1, 3	1	-1.26	-1.41	-1.66	-1.82	-2.25	-1.68	-3.01
		2	-3.29	-3.39	-3.55	-3.68	-3.74	-3.53	
		3	-3.76	-3.82	-3.82	-3.82	-3.82	-3.81	
	2, 3	1	-0.68	-0.93	-1.28	-1.32	-1.77	-1.20	-1.35
		2	-0.85	-0.97	-1.06	-1.13	-1.62	-1.12	
		3	-1.57	-1.59	-1.69	-2.11	N/A	-1.74	
	4, 3	1	3.91	3.91	3.91	3.91	N/A	3.91	2.22
		2	-0.10	0.53	0.91	1.94	3.10	1.28	
		3	0.67	1.56	1.60	1.69	1.84	1.47	
	5, 3	1	3.91	3.91	3.91	3.91	3.91	3.91	3.09
		2	2.16	2.32	2.50	3.04	3.04	2.61	
		3	2.05	2.55	2.57	2.78	3.79	2.75	
Solid Tissue	1, 3	1	-1.29	-1.37	-1.57	-1.80	-1.80	-1.57	-0.53
		2	-1.68	-2.16	-2.16	N/A	N/A	-2.00	
		3	1.47	1.83	2.59	N/A	N/A	1.96	
	2, 3	1	-0.55	-0.58	-0.78	-0.91	-0.93	-0.75	-0.72
		2	-0.34	-0.42	-0.55	-1.04	-1.42	-0.75	
		3	-0.46	-0.66	-0.86	N/A	N/A	-0.66	
	4, 3	1	-0.39	-0.40	-0.69	-0.70	-0.76	-0.59	-0.37
		2	-0.06	0.24	0.38	0.70	0.81	0.41	
		3	-0.65	-0.95	-1.02	-1.07	N/A	-0.92	
	5, 3	1	-0.02	-0.05	-0.27	-0.31	-0.40	-0.21	0.76
		2	0.37	0.67	0.90	1.47	N/A	0.85	
		3	0.80	1.53	1.67	2.04	2.08	1.62	

This example refers to odorant binding protein 14 [GenBank:94158822]. Hemolymph and solid tissue proteins were measured for 5 consecutive days during larval development and reported relative to the third day in natural log numbers "Ln(ratio)". The Ln(ratio) for top 5 most intense peptides for a given protein is averaged, first within the replicate (1, 2, or 3), then across the replicates to arrive at the average protein ratio in natural log.

A major strength of this method is the ability to track the changing abundances of hundreds of proteins during development. Those whose levels can be traced for at least 4 out of 5 days in either the tissue or hemolymph were considered to have an informative profile, a total of 522 proteins. Approximately equal numbers of tissue proteins showed an expression trend either positively or negatively correlated with age, but the latter was more common for hemolymph proteins, as might be expected from the high dynamic range of hemolymph as shown in Figure 3.1. It is crucial to note that the decreasing trend likely does not reflect an absolute reduction in expression levels of most proteins, but is rather a phenomenon of analyzing equal amounts of protein between two samples with a very large difference in absolute protein amounts caused primarily by drastic increases in secreted hexamerins. Consequently, lower abundance proteins become harder to detect in this background. Although the protein concentration in hemolymph changes only slightly beyond 1 day after hatching, the total volume and thus absolute protein content increases exponentially with age (Figure 3.3).

**Figure 3.3 Developmental changes of protein content in larval hemolymph**



The left axis denotes the volume of hemolymph per larva (diamonds,  $\mu\text{l}$ ) or hemolymph protein concentration (squares,  $\mu\text{g}/\mu\text{l}$ ), while the right axis describes the mass of total protein per larva (triangles,  $\mu\text{g}$ ). Measurements were made by pooling 5 to 120 larvae ( $n=3$  separate pools) depending on age (x-axis, in days) and size.

There is no direct functional information available for more than 99% of honey bee proteins so to derive some functional insight from the data acquired here we used BLAST2GO (Conesa et al., 2005) to systematically predict function based on sequence similarity (Supplementary Table 9 in Appendix A). After grouping specific Molecular Function ontologies into broader categories until they converged under one term (Supplementary Table 10 in Appendix A), the third-level terms were analyzed in detail. To find whether a given function term was developmentally regulated, an average expression profile was generated using data from proteins belonging under each term and tested for significance at the  $p < 0.05$  level (see Experimental Procedures). The slope between d1 and d5 was calculated to approximate the directionality and strength of temporal correlation. In the 34 terms considered, 11 of them had activity profiles that satisfied the significance criteria in at least one of either the solid tissue or hemolymph expression profiles (Table 3.2, details in Supplementary Table 11 in Appendix A).

**Table 3.2 Expression trends of proteins categorized under Gene Ontology terms**

Organ	GO Id Number	GO Term	Proteins considered	T-test of slope between day 1 and day 5	Slope
H	GO:0004857	enzyme inhibitor activity	6	1.9E-02	-0.19
T	GO:0004857	enzyme inhibitor activity	6	0.007	-0.41
T	GO:0004386	helicase activity	4	0.016	-0.37
T	GO:0016787	hydrolase activity	80	0.002	-0.10
H	GO:0003676	nucleic acid binding	18	7.2E-05	-0.24
T	GO:0003676	nucleic acid binding	38	7.8E-09	-0.33
H	GO:0000166	nucleotide binding	36	4.8E-03	-0.09
T	GO:0000166	nucleotide binding	72	0.022	-0.08
H	GO:0016491	oxidoreductase activity	19	1.8E-02	0.18
H	GO:0004871	signal transducer activity	3	4.7E-02	-0.22
H	GO:0003735	structural constituent of ribosome	18	1.5E-08	-0.41
T	GO:0003735	structural constituent of ribosome	35	8.4E-16	-0.34
T	GO:0022892	substrate-specific transporter activity	34	0.035	0.11
T	GO:0008135	translation factor activity, nucleic acid binding	10	0.037	-0.24
H	GO:0022857	transmembrane transporter activity	4	3.9E-02	-0.11

Organ	GO Id Number	GO Term	Proteins considered	T-test of slope between day 1 and day 5	Slope
T	GO:0022857	transmembrane transporter activity	27	0.010	0.12

Proteins were categorized under third-level Molecular Function terms and were evaluated as a group to assess whether their expression trends were age-regulated by performing paired t-tests comparing values from day 1 and day 5 of larval development, reporting the average slope between these 2 days if  $p < 0.05$  in either the solid tissue (T) or hemolymph (H). The total number of proteins belonging under a particular Gene Ontology term considered in the calculation is listed under "Proteins considered".

Terms "substrate-specific transporter activity" (GO:0022892) and "transmembrane transporter activity" (GO:0022857), both of which were tissue-specific activities, were very mildly positively-correlated with larval age. The majority were negatively-correlated with age, with the most statistically significant being "structural constituent of ribosome" (GO:0003735) and "nucleic acid binding" (GO:0003676). Others showing a similar trend include "enzyme inhibitor activity" (GO:0004857), "helicase activity" (GO:0004386), and "nucleotide binding" (GO:0000166). Terms that did not show regulation in either the tissue or hemolymph tended to be ones with non-specific participation in different pathways such as "transferase activity" (GO:0016740), "kinase regulator activity" (GO:0019207) and "cofactor binding" (GO:00048037).

With the current lack of a thoroughly curated protein function database for the honey bee, we manually assigned functional categories by employing a variety of available bioinformatic tools (see Methods and Supplementary Table 12 in Appendix A). This is necessary because certain major classes of honey bee proteins, such as hexamerins and odorant binding proteins, do not have high enough homology to proteins in other better annotated organisms and would thus be ignored. Furthermore, most proteins were assigned to multiple terms, or two very similar proteins were assigned to different but similar terms (GO: 0003676 nucleic acid binding and GO:0008135 translation factor activity, nucleic acid binding), which greatly complicates downstream hierarchical clustering and enrichment analysis. Groups that showed a significant temporal regulation (criteria nearly identical to the analysis of level 3 Molecular Function Gene Ontology terms) are shown in Table 3.3 Supplementary Table 13 in Appendix A).

**Table 3.3 Expression trends of manually annotated and categorized proteins**

Organ	Class	Proteins considered	T-test of slope between day 1 and day 5	Slope
T	adaptor	2	0.002	-0.43
T	aldo-keto reductase superfamily	3	0.041	-0.22
T	antioxidant	15	0.017	-0.20
T	ATP synthase	10	1.4E-04	0.21
H	carbohydrate metabolism	15	0.003	0.20
T	cuticle	7	0.036	0.19
T	electron transport chain	14	1.1E-05	0.22
H	energy storage	4	0.004	1.20
T	energy storage	5	0.028	0.56
T	kinases or phosphatases	2	0.044	0.24
H	pentose phosphate pathway	4	0.001	0.06
H	peptidase	15	0.045	0.14
H	proteasome	9	1.4E-04	-0.23
T	proteasome	18	8.6E-10	-0.32
T	protein folding	34	0.001	-0.17
T	Ras superfamily	10	0.009	-0.27
T	ribonucleoprotein	4	0.024	-0.43
H	ribosome	20	2.1E-09	-0.40
T	ribosome	38	4.7E-17	-0.34
T	TCA cycle	21	0.033	0.10
T	translation	14	0.015	-0.25
T	ubiquitination	3	0.021	-0.35
H	uncategorized	21	0.029	0.18

Proteins were categorized manually by function were evaluated as a group to assess whether their expression trends were age-regulated by performing t-tests paired t-tests to comparing values from day 1 and day 5. Significant ( $p < 0.05$ ) groups in either the solid tissue (T) or hemolymph (H) are shown.

A common protein expression pattern within a group was frequently observed. Ribosomal proteins levels in the tissue were consistently lowest in d2 and d5, but were overall decreasing in relative concentration with age ( $p < 1e-16$ ). Proteasome subunits and protein-folding chaperones exhibited the same overall trend ( $p < 1e-9$  and  $p < 0.005$  respectively). Energy storage proteins including apolipoproteins and hexamerins increased with age throughout the body but the trend was more dramatic in the hemolymph ( $p < 0.005$ ). There were no signs of temporal regulation of enzymes for fatty acid synthesis, beta oxidation, and carbohydrate metabolism. However, several groups of energy producing proteins showed varying degrees of positive correlation with time: tricarboxylic acid cycle proteins ( $p < 0.05$ ), ATP synthase subunits ( $p < 0.0005$ ), and

electron transport chain enzymes ( $p < 0.00005$ ). Surprisingly, we observed a decreased expression of antioxidant proteins, members of the Ras GTPase superfamily, and ubiquitylation enzymes in the solid tissues as development progressed ( $p < 0.05$ ,  $p < 0.01$ , and  $p < 0.05$  respectively). Many typically intracellular proteins, such as ribosomal proteins and proteasome subunits, were found in hemolymph as we have described previously (Chan et al., 2006) and as others have reported in other insects (Guedes Sde et al., 2003, Li et al., 2006).

We used hierarchical clustering to further analyze the 522 proteins that were profiled in either or both the solid tissue and hemolymph (Figure 3.4), followed by enrichment analysis according to manually assigned groupings. Clusters that satisfied the criteria (see Methods) for significant enrichment are shown in Table 3.4 (solid tissue) and Table 3.5 (hemolymph) (complete dataset shown in Supplementary Table 14 and Supplementary Table 15 in Appendix A, respectively).

**Table 3.4 Enrichment analysis of tissue protein functional classes following hierarchical clustering**

Node Number	Correlation	Proteins in this node	Protein class	Class total	Percent Class Enrichment
265	0.98	5	helicase	3	67
277	0.98	8	hormone synthesis	4	50
376	0.94	79	transcription	3	100
376	0.94	79	chromatin-associated protein	3	67
376	0.94	79	tRNA synthetase	3	67
376	0.94	79	pentose phosphate pathway	4	50
376	0.94	79	ubiquitylation	4	50
377	0.94	5	food	6	50
411	0.90	108	aldo-keto reductase superfamily	3	67
419	0.89	137	proteasome	24	67
419	0.89	137	antioxidant	16	50
419	0.89	137	protein receptor	4	50
421	0.89	29	ATP synthase	10	60
427	0.88	34	small molecule receptor	4	50
434	0.86	51	energy storage	5	80
434	0.86	51	beta-oxidation	8	50
438	0.83	7	cuticle	7	57
439	0.83	146	Ras superfamily	10	50

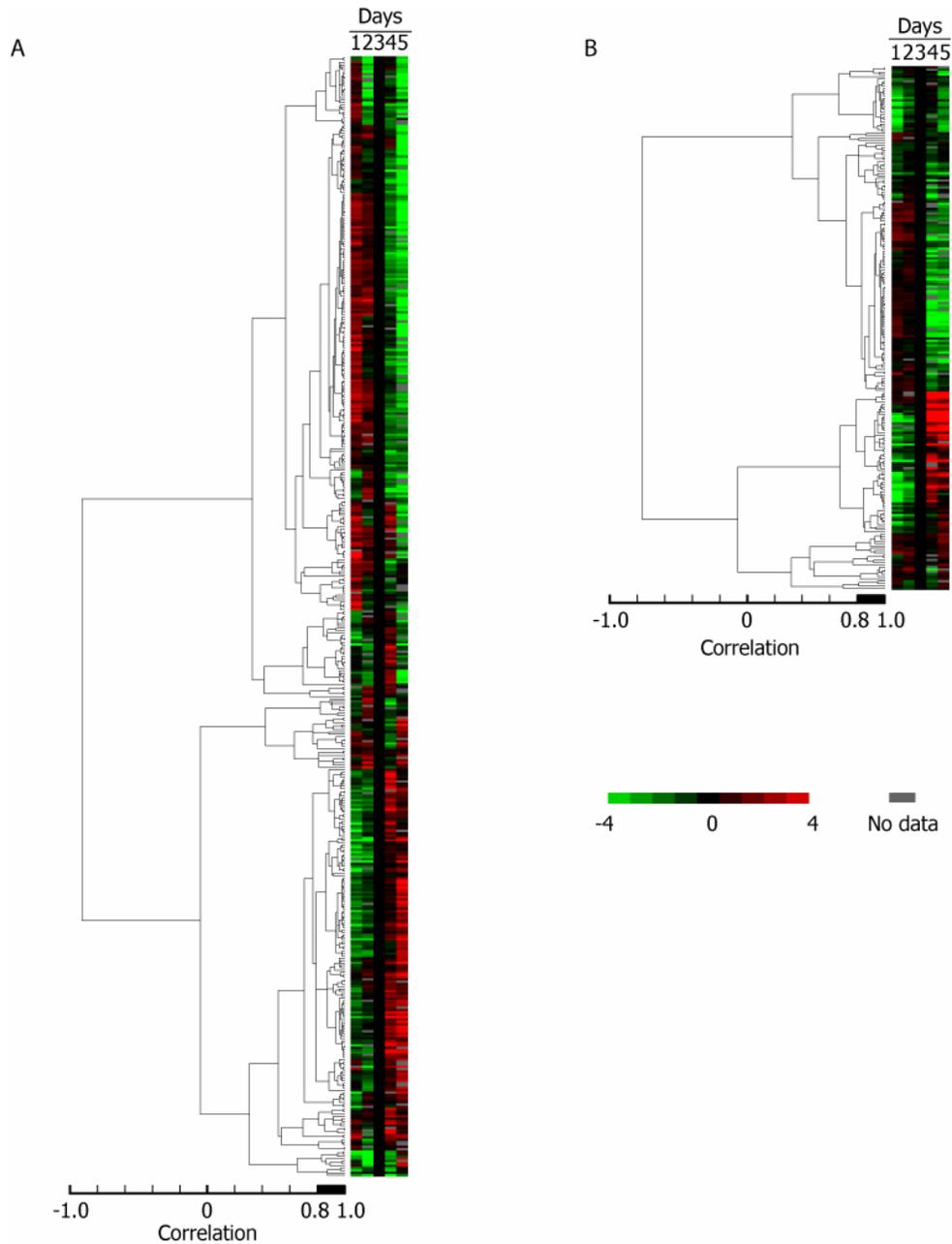
Functionally categorized proteins found in tissue samples during larval development, represented in Figure 3.4 (Panel A), are listed here if considered significant, defined by at least 50% enrichment in nodes with a correlation of  $> 0.8$ .

**Table 3.5 Enrichment analysis of hemolymph protein functional classes following hierarchical clustering**

Node Number	Correlation	Proteins in this node	Protein class	Class total	Percent Enrichment
128	0.97	35	translation	7	57
150	0.96	6	short-chain dehydrogenase family	4	50
180	0.93	21	small molecule receptor	4	50
183	0.92	23	food	8	63
183	0.92	23	glycolipid metabolism	3	67
185	0.92	9	ubiquitylation	4	50
190	0.89	63	amino acid metabolism	8	50
197	0.86	29	energy storage	4	100
200	0.84	81	proteasome	10	90
200	0.84	81	protein folding	20	60
200	0.84	81	ribosome	31	81
203	0.82	21	TCA cycle	4	75

Functionally categorized proteins found in hemolymph samples during larval development, represented in Figure 3.4 (Panel B), are listed here if considered significant, defined by at least 50% enrichment in nodes with a correlation of >0.8.

**Figure 3.4 Average-linkage clustering of proteins quantified in the honey bee larvae**



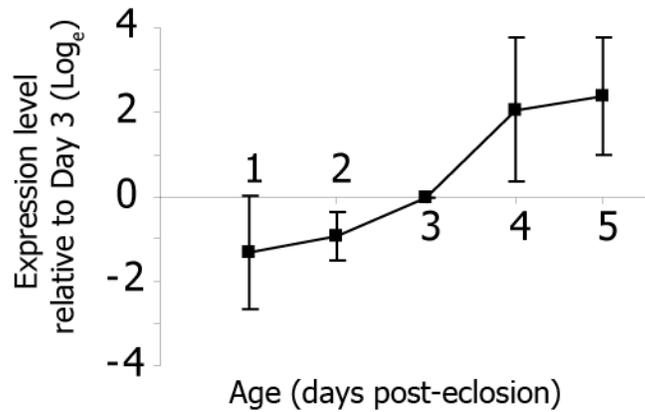
Proteins that were quantified in either or both the tissue (A) or hemolymph (B) for at least 4 out of 5 tested days were arranged by hierarchical clustering using software described in (Eisen et al., 1998). All expression values, shown relative to day 3 (=0, black), have been natural log-transformed (>0, red; <0 green; no data=grey). These proteins, which have been manually annotated with a function and category, are calculated for enrichment within a node (results in Table 3) if the node correlation value is >0.8 (see thick bar on scale).

Only a few functional classes of proteins were enriched in the same node, since expression profiles for some proteins exhibit biological variability that causes apparent

inconsistency with the timecourse. Transcription and chromatin-associated proteins as well tRNA synthetases – clearly related by their tasks – shared Node 376 (correlation 0.94) with pentose phosphate pathway and ubiquitylation enzymes. Energy storage and beta-oxidation proteins were both concentrated in Node 434 (correlation 0.86, solid tissue). Protein turnover machinery including ribosomes, protein folding, and proteasome were all enriched in Node 200 (correlation 0.84, hemolymph). Many of these clusters are also protein families already noted to show significant temporal regulation, such as energy storage proteins, ATP synthases, antioxidant proteins, and ubiquitylation enzymes. This indirectly suggests that suitable assignments were made during manual annotation and categorization, since their regulation pattern were grouped using independent methods.

Automated and semi-automated functional annotation and categorization effectively highlighted expression trends in large classes of proteins. With this approach, however, classes with only a few members or those where particular proteins have highly specialized function tended to fall below the significance threshold unless they were considered individually. In solid tissues, the levels of 86 proteins changed significantly ( $p < 0.05$ ) over the tested period, accounting for 13% of all the quantifiable proteins in solid tissues. For example, levels of neuropeptide Y receptor increased 46-fold from d3 to d5. In the hemolymph, 66 of 481 (14%) quantified proteins changed significantly during the larval stage ( $p < 0.05$ ). Most of these are intracellular proteins, yet the regulation of truly secreted proteins is frequently far more dramatic. An imaginal disc growth factor [GenBank:66514614] increased more than 13-fold from d1 to d5 (Figure 3.5).

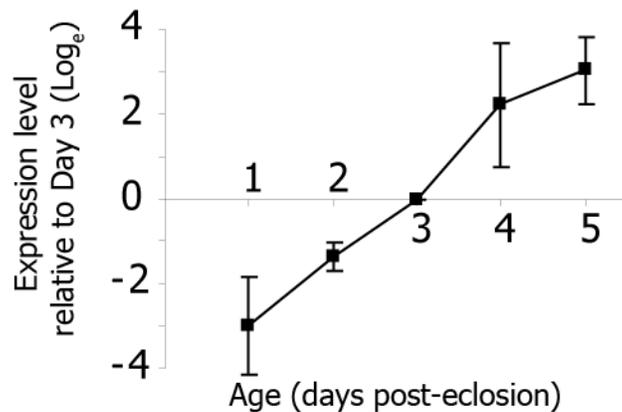
**Figure 3.5 Expression profile of imaginal disc growth factor during larval development**



Imaginal disc growth factor relative expression levels (y-axis, in natural log scale) over 5 days of larval growth (x-axis) is shown. Error bars represent one standard deviation.

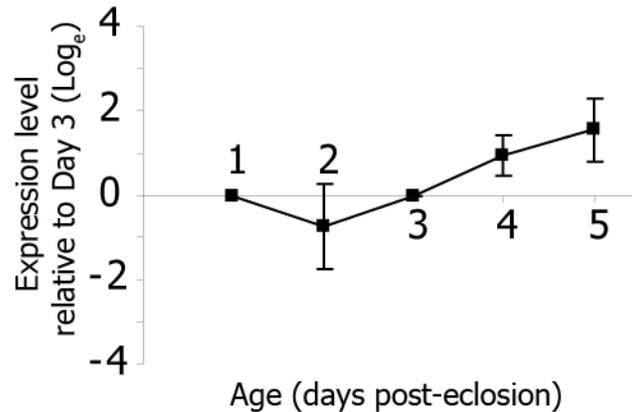
OBP14 [GenBank:94158822] levels changed in a similar fashion, with the former displaying a 40-fold change over 5 days (Figure 3.6). Antimicrobial peptide apisimin [GenBank:58585112] (Figure 3.7) and melanization enzyme proPO [GenBank:58585196] expression were also positively correlated with age.

**Figure 3.6 Expression profile of odorant binding protein 14 during larval development**



OBP14 relative expression levels (y-axis, in natural log scale) over 5 days of larval growth (x-axis) is shown. Error bars represent one standard deviation.

**Figure 3.7 Expression profile of apisimin during larval development**



Apisimin relative expression levels (y-axis, in natural log scale) over 5 days of larval growth (x-axis) is shown. Error bars represent one standard deviation.

To our knowledge this is the first proteome-level description of honey bee larval development so to gain additional insight, we compared our data with a previously reported developmental study of the fruit fly. While *Drosophila* and *Apis* are separated by 300 million years of evolution (2006), *Drosophila* is nonetheless the closest highly-studied phylogenetic neighbour of the bee. A whole body transcriptome study of the *D. melanogaster* life cycle was published in 2002 (Arbeitman et al., 2002), which included a list of genes that were significantly regulated during the larval period. After finding the protein homologs common to our study and the fruit fly larval dataset (34 in total), we calculated the slope of linear regression of expression trends for both organisms (the slope of the honey bee tissue and hemolymph profiles were averaged when needed, see Experimental Procedures). Slope values that have opposite signs or an absolute difference in slope of greater than 0.75 were considered dissimilar, amounting to 38% (13 of 34) of the proteins considered (Table 3.6, complete dataset with BLAST homolog search results in Supplementary Table 16 in Appendix A). The most extreme slope reported for both organisms is for the hexamerin 70b protein (1.5 for bees and 1.6 for flies).

**Table 3.6 Protein expression levels during larval development: honey bee versus fruit fly larvae**

Honey bee accession number	Protein description	Slope difference (Honey bee minus fruit fly)	Expression trend: different or same?
GenBank:48126476	translation --- initiation factor 3f	0.01	same
GenBank:110749015	short-chain dehydrogenase family --- oxidoreductase	0.22	same
GenBank:110756656	protein methylation --- arginine methyltransferase	0.10	same
GenBank:110759433	ribonucleoprotein --- ribonucleoprotein	0.10	same
GenBank:110761364	cytoskeleton --- alpha-actinin	0.19	same
GenBank:66547531	pentose phosphate pathway --- 6-phosphogluconate dehydrogenase	0.24	same
GenBank:66509442	peptidase --- dipeptidyl aminopeptidase	0.02	same
GenBank:110764347	amino acid metabolism --- enolase-phosphatase E1 (methionine salvage pathway)	0.50	same
GenBank:110762382	transcription --- spermidine synthase	0.51	same
GenBank:110763730	antioxidant --- glutathione S transferase	0.32	same
GenBank:66504249	uncategorized --- protein kinase c substrate	0.33	same
GenBank:110755309	protein receptor --- high density lipoprotein binding protein	0.33	same
GenBank:110750855	unknown function --- unknown function	0.36	same
GenBank:94158626	electron transport chain --- cytochrome p450	0.51	same
GenBank:58585148	energy storage --- hexamerin 70b	0.15	same
GenBank:48095159	peptidase --- serine protease	0.34	same
GenBank:66524124	peptidase --- carboxypeptidase B	0.07	same
GenBank:66509812	peptidase --- angiotensin converting enzyme	0.28	same
GenBank:110762229	peptidase --- chymotrypsin	0.47	same
GenBank:66510448	glycolipid metabolism --- beta-glucosidase (glucocerebrosidase)	0.10	same
GenBank:110766932	uncategorized --- mannosidase, lysosomal	0.50	same
GenBank:66513481	ubiquitination --- ubiquitin-activating enzyme E1	0.48	different
GenBank:66522467	uncategorized --- juvenile hormone inducible protein	0.95	different
GenBank:48104663	protein receptor --- protein kinase C receptor	0.60	different
GenBank:110758189	uncategorized --- carboxylesterase	0.94	different
GenBank:110756254	ribonucleoprotein --- ribonucleoprotein	0.67	different
GenBank:66522232	uncategorized --- isochorismatase	0.43	different
GenBank:66535270	uncategorized --- oxoacidtransferase	0.42	different
GenBank:66521459	membrane transporter --- porin	0.59	different
GenBank:110764660	helicase --- RNA helicase	0.56	different
GenBank:110762902	ATP synthase --- ATP synthase component	0.77	different
GenBank:66525867	small molecule carrier --- solute carrier	0.82	different

Honey bee accession number	Protein description	Slope difference (Honey bee minus fruit fly)	Expression trend: different or same?
GenBank:58531215	membrane transporter --- translocase, ATP	0.87	different
GenBank:110759569	apoptosis --- beta-hexosaminidase	1.14	different

Genes in a life-cycle transcriptomic analysis of *Drosophila melanogaster* (Arbeitman et al., 2002) were compared to honey bee larval proteomics data in this report by finding homologs common to these studies. Significant matches (see Methods for criteria) were assessed by comparing the slope values calculated between days 1 and 4: the protein was marked "same" if the sign of the slope was the same had a difference no greater than 0.75, or they were otherwise marked as "different".

### 3.5 Discussion

The data presented here, at the level of the whole proteome, documents the dramatic changes occurring in developing honey bee larvae. The most striking, by far, is the 1500-fold increase in weight over just 6 days (Snodgrass, 1925). In our proteomic analysis of the solid tissue, the most abundant organs are best represented, namely the fat body (accounts for 65% of the mass (Bishop, 1923)), followed by the midgut and larval tubules. The hemolymph fraction reflects the secretory activities of all these tissues and also the molecules associated with intercellular communication and regulation. The presence of intracellular proteins suggest that hemolymph plays a major role in clearing apoptotic cells, in line with observations of the equivalent connective tissue in mammals (i.e., blood) (Omenn et al., 2005). No dissection of specific larval organs was performed because many do not develop until the late stages, making direct comparisons of organ development by quantitative proteomics impossible.

We have found both automated (BLAST2GO) and semi-automated annotation (manual selection of descriptions provided by automated tools and manual categorization) to be very valuable for maximizing available information on an organism with otherwise very little functional annotation. While automated ontological methods were reliable and bias-free, outputs might be too generic (e.g., "ion binding" - GO:0043167), or failing to accurately represent several very important protein families of the honey bee (e.g., hexamerins and odorant binding proteins), highlighting the need for manual intervention. Cluster analysis is an indispensable tool for spotting expression trends, but given that the software for rigorous statistical enrichment analysis is designed specifically for popular model organisms such as mouse, worm, and yeast, the

descriptive statistical approach used here was nevertheless able to provide credible insights about larval developmental biology or led to conclusions confirmed by other information.

The major behaviour during the larval stage is feeding as it prepares itself for the subsequent pupal stage when no feeding occurs. Based on various data acquired over the past century, it has been proposed that the larval fat body undergoes two phases: beginning with a high rate of protein synthesis and poor uptake of hemolymph substances, followed by a phase of low cellular synthesis and improved uptake and storage of hemolymph proteins (de Oliveira et al., 2003). Our data now allows us to clarify this model and provide molecular-level detail of these changes. One of the most remarkable events in a growing larva is the substantial synthesis of hexamerins and lipoproteins in the fat body, followed by their appearance in the hemolymph near the end of this developmental stage (reviewed extensively in (Price, 1973)). While the age-dependent production of these abundant storage proteins is well known, here we provide evidence of a concomitant up-regulation of low copy transmembrane transporters (GO:0022857) that may facilitate the export, including a porin (GenBank:66521459). Paradoxically, this astounding rate of protein production and export is paired with an opposite trend in protein synthesis machinery and accessories, which had been suspected in two reports in the 1960's ((Martin et al., 1969, Ruegg, 1968)). Now we have evidence for these previous suggestions, including the clear age-associated decrease of more than 50 detected ribosomal subunits, coupled with an increase of two transcription repressors (although at  $p < 0.1$  these did not satisfy significance criteria) to support this former notion.

Fat accumulation is an important purpose of the rapid larval growth, clearly indicated by the size of the fat body tissue relative to the whole organism, as well as the buildup of lipophorins. Lipids in larval food is only 4% by weight (Antinelli et al., 2002), meaning that de novo synthesis must account for the bulk of stored fat. Fatty acid synthase [GenBank:66515350] was one of the most abundant proteins throughout the entire tested period based on absolute protein expression estimates (Lu et al., 2007), yet to our surprise we did not observe significant temporal regulation in the expression of this enzyme with age. It is worth noting that "fat body" is somewhat of a misnomer, given that it is involved in protein and glycogen storage, as well as fat (Coupland, 1957,

de Oliveira et al., 2003). To drive these endergonic biosynthetic processes, the demands for ATP must therefore be great. Not only do we observe significant age-associated increases in ATP synthase subunits, but also enzymes in energy-producing pathways such as the TCA cycle and the electron transport chain components. This may be attributed to an increase in mitochondria size or numbers; however, there are at least two reports that claim the number of mitochondria decreases as the larva approaches pupation in other insects (Benson, 1965, Walker, 1966).

Proteins with high copy number, including the many discussed above, are always the first to be investigated in any organism. The difficulties in studying proteins in honey bee larva have multiple sources: a) the abundant storage proteins broaden the dynamic concentration range, obscuring the rare proteins, b) the clean dissection of larval organs presents a technical challenge since the fat body is large and is difficult to remove, and finally c) the lack of available antibodies against even the most common proteins makes many conventional biochemistry experiments such as immunoprecipitation and Western blotting impossible. These reasons have especially hindered the study of fine larval organs such as the nervous system and low abundance proteins related to immunity or pathway regulation.

The ability of larvae to respond to external stimuli and internal regulatory cues increases with time, a trend that is clearly reflected in our data. For example, odorant-based communication has been observed in old larvae (Laurent et al., 2002, Le Conte et al., 2006). OBP14 [GenBank:94158822] was detected even on the first day after hatching, showing upregulation with age (Figure 3.6). This suggests that younger larvae may have the capability to bind certain odorant molecules, but whether that translates into pheromonal communication is entirely speculative. The positive temporal regulation of antimicrobial peptide apisimin [GenBank:58585112] (Figure 3.7) and the melanization enzyme prophenoloxidase [GenBank:58585196] in the hemolymph, which have clear roles in defense (Bilikova et al., 2002, Ling et al., 2005, Marmaras et al., 1996, Tang et al., 2006), matches the observed susceptibility to diseases such as foulbroods of the young larvae, suggesting that one or both of these may be the factor responsible for successful defense against foulbroods in older larvae. However, a C-lectin [GenBank:110750008] and a complement factor [GenBank:66508940] actually has no observable expression trends, indicating that they may have alternative roles different

from homology-based function predictions. The 46-fold increase of a neuropeptide Y receptor [GenBank:110764421], which controls appetite and fat storage is reasonable given the feeding activity of the larvae. An imaginal disc growth factor [GenBank:66514614] increased by 40-fold over the course of the experiment (Figure 3.5) presumably gears the larva for pupal development where specific limbs and organs will grow from imaginal discs containing highly differentiated cells.

Proteomics is generally a discovery-method and is thus an excellent mechanism for hypothesis generation. We were able to find several peculiar proteins supported by a number of high quality mass spectra but no plausible explanation for its presence or degree of age-dependent regulation. A protein annotated as "PREDICTED: similar to CG15040-PA" [GenBank:110749732] was consistently found only in the hemolymph of older larvae (up to 24-fold higher in 5-day larvae old compared to 3-day old), yet it has no likely homologs or discernable functional domains, bearing only a vague resemblance to a protein [GenBank:124512744] from *Plasmodium falciparum* 3D7, found by PSI-BLAST (Altschul et al., 1997, Altschul et al., 2005).

### **3.6 Conclusions**

To study honey bees, individual, environmental, and social factors must be considered. The larval developmental stage has shown to be a highly complex period of biochemical regulation. The proteomics data presented here was able to support a model for energy metabolism and storage, as well as reveal unexpected expression trends for proteins that respond to external and internal stimulus, such as pheromones, pathogens, and oxidants.

### 3.7 References

- MSQuant [<http://msquant.sourceforge.net>]
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, 25(17):3389-3402.
- Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schaffer AA, Yu YK. **Protein database searches using compositionally adjusted substitution matrices.** *Febs J* 2005, 272(20):5101-5109.
- Antinelli JF, Davico R, Rognone C, Faucon JP, Lizzani-Cuvelier L. **Application of solid/liquid extraction for the gravimetric determination of lipids in royal jelly.** *J Agric Food Chem* 2002, 50(8):2227-2230.
- Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP. **Gene expression during the life cycle of *Drosophila melanogaster*.** *Science* 2002, 297(5590):2270-2275.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al.* **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, 25(1):25-29.
- Benson KA. **An investigation of some developmental changes in the larval fat body of *Sarcophaga bullata* prior to metamorphosis.** University of Virginia; 1965.
- Bilikova K, Hanes J, Nordhoff E, Saenger W, Klaudiny J, Simuth J. **Apisimin, a new serine-valine-rich peptide from honeybee (*Apis mellifera* L.) royal jelly: purification and molecular characterization.** *FEBS Lett* 2002, 528(1-3):125-129.
- Bishop GH. **Cell metabolism in the insect fat body.** *J Morph* 1923, 37:533-553.
- Candiano G, Bruschi M, Musante L, Santucci L, Ghiggeri GM, Carnemolla B, Orecchia P, Zardi L, Righetti PG. **Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis.** *Electrophoresis* 2004, 25(9):1327-1333.
- Chan QW, Howes CG, Foster LJ. **Quantitative comparison of caste differences in honeybee hemolymph.** *Mol Cell Proteomics* 2006, 5(12):2252-2262.
- 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, 21(18):3674-3676.
- The Honey Bee Genome Sequencing Consortium. **Insights into social insects from the genome of the honeybee *Apis mellifera*.** *Nature* 2006, 443(7114):931-949.
- Coupland RE. **Observations on the normal histology and histochemistry of the fat body of the locust (*Schistocerca gregaria*).** *J Exp Biol* 1957, 34:290-296.
- Cunha AD, Nascimento AM, Guidugli KR, Simoes ZL, Bitondi MM. **Molecular cloning and expression of a hexamerin cDNA from the honey bee, *Apis mellifera*.** *J Insect Physiol* 2005, 51(10):1135-1147.
- Danty E, Arnold G, Burmester T, Huet JC, Huet D, Pernollet JC, Masson C. **Identification and developmental profiles of hexamerins in antenna and hemolymph of the honeybee, *Apis mellifera*.** *Insect Biochem Mol Biol* 1998, 28(5-6):387-397.

- de Oliveira VTP, Cruz-Landim C. **Morphology and function of insect fat body cells: a review.** *Biociencias* 2003, 11(2):195-205.
- Eisen MB, Spellman PT, Brown PO, Botstein D. **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci U S A* 1998, 95(25):14863-14868.
- Elsik CG, Mackey AJ, Reese JT, Milshina NV, Roos DS, Weinstock GM. **Creating a honey bee consensus gene set.** *Genome Biol* 2007, 8(1):R13.
- Foster LJ, De Hoog CL, Mann M. **Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors.** *Proc Natl Acad Sci U S A* 2003, 100(10):5813-5818.
- Guedes Sde M, Vitorino R, Tomer K, Domingues MR, Correia AJ, Amado F, Domingues P. **Drosophila melanogaster larval hemolymph protein mapping.** *Biochem Biophys Res Commun* 2003, 312(3):545-554.
- Haldane JBS. **Aristotle's account of bees' 'dances'.** *J Hellenic Studies* 1955, 75(1955):24-25.
- Hsu JL, Huang SY, Chow NH, Chen SH. **Stable-isotope dimethyl labeling for quantitative proteomics.** *Anal Chem* 2003, 75(24):6843-6852.
- Ishihama Y, Rappsilber J, Mann M. **Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics.** *J Proteome Res* 2006, 5(4):988-994.
- Korochkina SE, Gordadze AV, York JL, Benes H. **Mosquito hexamerins: characterization during larval development.** *Insect Mol Biol* 1997, 6(1):11-21.
- Kwok MC, Holopainen JM, Molday LL, Foster LJ, Molday RS. **Proteomics of photoreceptor outer segments identifies a subset of SNARE and rab proteins implicated in membrane vesicle trafficking and fusion.** *Mol Cell Proteomics* 2008.
- Laurent S, Masson C, Jakob I. **Whole-cell recording from honeybee olfactory receptor neurons: ionic currents, membrane excitability and odourant response in developing workerbee and drone.** *Eur J Neurosci* 2002, 15(7):1139-1152.
- Le Conte Y, Becard JM, Costagliola G, de Vaublanc G, El Maataoui M, Crauser D, Plettner E, Slessor KN. **Larval salivary glands are a source of primer and releaser pheromone in honey bee (*Apis mellifera* L.).** *Naturwissenschaften* 2006, 93(5):237-241.
- Li XH, Wu XF, Yue WF, Liu JM, Li GL, Miao YG. **Proteomic analysis of the silkworm (*Bombyx mori* L.) hemolymph during developmental stage.** *J Proteome Res* 2006, 5(10):2809-2814.
- Ling E, Yu XQ. **Prophenoloxidase binds to the surface of hemocytes and is involved in hemocyte melanization in *Manduca sexta*.** *Insect Biochem Mol Biol* 2005, 35(12):1356-1366.
- Lu P, Vogel C, Wang R, Yao X, Marcotte EM. **Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation.** *Nat Biotechnol* 2007, 25(1):117-124.
- Marmaras VJ, Charalambidis ND, Zervas CG. **Immune response in insects: the role of phenoloxidase in defense reactions in relation to melanization and sclerotization.** *Arch Insect Biochem Physiol* 1996, 31(2):119-133.

- Martin MD, Kinnear JF, Thomson JA. **Developmental changes in the late larva of *Calliphora stygia*. II. Protein synthesis.** *Aust J Biol Sci* 1969, 22:935-945.
- Nelson JA. **Morphology of the honeybee larva.** *J Agricult Res* 1924, 28(12):1167-1229.
- Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, Apweiler R, Haab BB, Simpson RJ, Eddes JS *et al.* **Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database.** *Proteomics* 2005, 5(13):3226-3245.
- Patel A, Fondrk MK, Kaftanoglu O, Emore C, Hunt G, Frederick K, Amdam GV. **The making of a queen: TOR pathway is a key player in diphenic caste development.** *PLoS ONE* 2007, 2(6):e509.
- Price GM. **Protein and nucleic acid metabolism in insect fat body.** *Biological Reviews* 1973, 48:333-375.
- Rappsilber J, Ishihama Y, Mann M. **Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics.** *Anal Chem* 2003, 75(3):663-670.
- Ruegg MK. **Untersuchungen zum Proteinstoffwechsel des Wildtyps und der Letalmutante (Itr) von *Drosophila melanogaster*.** *vergl Physiol* 1968, 60:275-307.
- Scharlaken B, de Graaf DC, Goossens K, Peelman LJ, Jacobs FJ. **Differential gene expression in the honeybee head after a bacterial challenge.** *Dev Comp Immunol* 2008, 32(8):883-889.
- Snodgrass RE. **The Feeding and Growth of the Larvae.** In: *Anatomy and Physiology of the Honeybee*. Edited by Piper CV. New York: McGraw-Hill; 1925: 172.
- Tang H, Kambris Z, Lemaitre B, Hashimoto C. **Two proteases defining a melanization cascade in the immune system of *Drosophila*.** *J Biol Chem* 2006, 281(38):28097-28104.
- Tatusova TA, Madden TL. **BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences.** *FEMS Microbiol Lett* 1999, 174(2):247-250.
- Walker PR. **An electromicroscope study of the fat body of the moth, *Philosamia*, during growth and metamorphosis.** *J Insect Physiol* 1966, 12:1009-1018.
- Wolschin F, Amdam GV. **Comparative proteomics reveal characteristics of life-history transitions in a social insect.** *Proteome Sci* 2007, 5:10.
- Wolschin F, Amdam GV. **Plasticity and robustness of protein patterns during reversible development in the honey bee (*Apis mellifera*).** *Anal Bioanal Chem* 2007, 389(4):1095-1100.
- Yue C, Genersch E. **RT-PCR analysis of Deformed wing virus in honeybees (*Apis mellifera*) and mites (*Varroa destructor*).** *J Gen Virol* 2005, 86(Pt 12):3419-3424.

## 4 THE INNATE IMMUNE AND SYSTEMIC RESPONSE IN HONEY BEES TO A BACTERIAL PATHOGEN, *PAENIBACILLUS LARVAE*<sup>7</sup>

### 4.1 Abstract

There is a major paradox in our understanding of honey bee immunity: the high population density of a bee hive implies a high rate of disease transmission among individuals, yet bees are predicted to express only two-thirds as many immunity genes as solitary insects, e.g., mosquito or fruit fly. This suggests that the immune response in bees is subdued in favor of social immunity, yet some specific immune factors are expressed in response to infection. To explore the response to infection more broadly, we employ mass spectrometry-based proteomics in a quantitative analysis of honey bee larvae infected with the bacterium *Paenibacillus larvae*. Newly-eclosed bee larvae, in the second stage of their life cycle, are susceptible to this infection, but become progressively more resistant with age. We used this host-pathogen system to probe not only the role of the immune system in responding to a highly evolved infection, but also what other mechanisms might be employed in response to infection. Using quantitative proteomics, we compared the hemolymph (insect blood) of five-day old healthy and infected honey bee larvae and found a strong up-regulation of some metabolic enzymes and chaperones, while royal jelly (food) and energy storage proteins were down-regulated. We also observed increased levels of the immune factors proPO, lysozyme and the antimicrobial peptide hymenoptaecin. Furthermore, mass spectrometry evidence suggests that healthy larvae have significant levels of catalytically inactive proPO in the hemolymph that is proteolytically activated upon infection. PO enzyme activity was undetectable in one or two-day-old larvae and increased dramatically thereafter, paralleling very closely the age-related ability of larvae to resist infection. We propose a model for the host response to infection where energy stores and metabolic enzymes

---

<sup>7</sup> A version of this chapter has been accepted for publication. Chan, Q.W., Melathopoulos, A.P., Pernal, S.F. and Foster, L.J. (2009) The innate immune and systemic response in honey bees to a bacterial pathogen, *Paenibacillus larvae*. BMC Genomics. 10:387.

are regulated in concert with direct defensive measures, such as the massive enhancement of PO activity.

## 4.2 Introduction

Honey bees, *Apis mellifera*, face a number of niche-specific pathogens such as the endospore-forming bacterium *Paenibacillus larvae*, the causative agent of American Foulbrood (Shimanuki, 1997). Honey bees are only susceptible to *P. larvae* during the first 48 h following eclosion (egg hatching), in their first and second instar developmental stages. It remains unclear why larvae acquire immunity against *P. larvae* after the third instar, whereas the ingestion of merely 10 spores can cause systemic infection and death in the previous instars (Brødsgaard et al., 1998). It was thought that *P. larvae* spores germinate in the larval midgut and enter the epithelium by phagocytosis (Davidson, 1973, Gregorc et al., 1998) but recent data suggest that the bacteria follow a paracellular route to breach the epithelial wall (Yue et al., 2008). The effectiveness of the antimicrobial peptide defensin against *P. larvae* was documented in growth inhibition assays (Bachanova et al., 2002, Bilikova et al., 2001) using fractionated royal jelly (honey bee food). However, Evans et al. found no changes in defensin gene expression in larvae fed *P. larvae* spores and, paradoxically, that abaecin (an AMP) gene expression was greatest in newly eclosed larvae (Evans, 2004, Evans et al., 2004), the most susceptible stage. More recently, the same group also showed that infection caused elevated expression of Toll-like receptor, MyD88 and I $\kappa$ B (Evans et al., 2006). Thus, even though they do not always respond as expected, honey bees have all the components of an innate immune system. Here we explore the response of this system to a physiologically relevant infection in a natural setting.

To this end, hemolymph (arthropod blood) is well-suited for studying insect immunity; it is especially relevant in the case of *P. larvae* as the bacterium contacts hemolymph as soon as it breaches the gut epithelium. This fluid contains antimicrobial factors produced largely by the fat body and, to a lesser extent, hemocytes. These cells can also respond to infectious particles by phagocytosing them or by autolysis, which is part of an encapsulation pathway used to inhibit growth of microorganisms. Likewise, as the connective tissue responsible for transporting various molecules throughout the body, it is also optimally suited for monitoring systemic changes in other pathways. Previously,

we have examined how hemolymph changes during normal larval development (Chan et al., 2008). In that study we observed that most immune factors were not significantly altered during development, with only the antimicrobial peptide apisimin and the monooxygenase prophenoloxidase showing any age-related changes in expression. Thus, based on our earlier work and that of others (Chan et al., 2008, Evans, 2004, Evans et al., 2004, Evans et al., 2006, Randolt et al., 2008), we expected that a *P. larvae* infection should induce, in hemolymph, elevated levels of at least some AMPs, as well as other antibacterial enzymes such as lysozyme and prophenoloxidase. To address these predictions, we use MS-based proteomics to measure changes in hemolymph protein levels in larvae challenged with *P. larvae*. Furthermore, we predict that a protein that is able to convey immunity to older larvae and adults must be expressed at extremely low levels in the susceptible early larval instars. Using a functional assay, we demonstrate how one potentially critical player in host defense, PO, correlates with larval resistance to infection.

## **4.3 Experimental Procedures**

### **4.3.1 Honey bees and infection experiments**

All infection experiments were conducted at Beaverlodge, AB, Canada during July and August of 2005. Three five-frame nucleus colonies ('nucs') were prepared with three frames of bees and open brood and with newly-mated sister queens. In each nuc, 100 by 100 mm patches of first instar larvae were selected and sprayed with 20 mL of one of the following: 1) a  $6.0E+06$  spores/mL suspension of spores isolated from naturally occurring AFB 'scale' collected in 2004 (PL-Scale), 2) a  $4.4E+06$  spores/mL suspension of spores from NRRL B-3650 (PL-Lab), a virulent laboratory strain of *P. larvae* (courtesy of Jay Evans), and 3) phosphate-buffered saline (PBS).

### **4.3.2 Sample collection and processing for MS**

Four days after infection larvae (estimated to be in late fourth or fifth instar) within each marked square (PL-Scale, PL-Lab and control) were extracted using soft forceps (Bioquip, Rancho Dominguez, CA) and bled as described (Chan et al., 2008). Hemolymph was processed as described for larvae in (Chan et al., 2006). In each case, we compared 20  $\mu$ g of protein from infected hemolymph with the control by differential

labeling of tryptic peptides using light ( $C^1H_2O$ ) and heavy ( $C^2H_2O$ ) isotopologs of formaldehyde prior to analysis by LC-MS/MS using a linear trapping quadrupole-OrbitrapXL exactly as described (Chan et al., 2008).

#### **4.3.3 MS data analysis**

Raw data processing to arrive at peptide ion volume ratios was performed exactly as described (Chan et al., 2008). Data for each infection method were pooled, and proteins with five or more quantified peptides from any one or all of the replicates were considered quantified. This approach operates on the underlying assumption that each peptide ratio is a technical replicate, which is different from most publications where averages are made at the protein, not the peptide level. The conventional method implies that a protein quantified from averaging over a large number of quantified peptides has the same statistical power than one quantified with the bare minimum. To circumvent this disadvantage, we took single peptides from three biological replicates as individual data points, which reasonably accounts for the greater statistical power afforded by well-detected proteins with many quantified peptides. The average level of protein regulation is represented by the median peptide ratio. We employed the two-tailed Wilcoxon matched-pairs signed-ranked test on these proteins using Analyse-It (v2.12, [www.analyse-it.com/](http://www.analyse-it.com/)), with peptides as data points to assess whether the expression level of each protein was significantly changed by infection at 95% and 99% confidence (McCornack, 1965). For proteins with more than 100 peptides, the 100 most intense  $[M+nH]^{n+}$  ions (heavy and light combined,  $n = 2$ ) were selected for analysis. The same test was used on peptides quantified in both infection methods to assess whether the two methods yielded the same effects on protein expression. Experimentally or bioinformatically-inferred evidence of protein functions and names discussed throughout this report is provided in Supplementary Table 21 in Appendix A.

#### **4.3.4 Hemolymph collection for PO activity assay**

We collected honey bee larvae and estimated their age in days by size. Animals at each age were pooled to collect at least 8  $\mu$ L of hemolymph per replicate for three replicates – the number of larvae required varied from approximately 150 for the very young larvae, and 2-5 for the oldest larvae tested (five days old); 18 fourth to fifth instar larvae were pooled and used for the fractionation experiment and processed as

described for larvae in (Chan et al., 2006). Protein concentrations were assayed by Coomassie Plus (except in the protein fractionation experiment, see below) and were normalized across all samples using Assay Buffer (20mM Tris-HCl, pH 8).

#### **4.3.5 Hemolymph fractionation**

Hemolymph was desalted using a mini Zeba column (Pierce) according to the manufacturer's instructions. Of the 75  $\mu$ L total volume, 25  $\mu$ L was reserved for MS analysis. The remainder was applied to a mini strong anion exchange column (Pierce), and washed with 100  $\mu$ L of Assay Buffer between step-elutions of a 10-step sodium chloride gradient prepared in Assay Buffer: 0.04, 0.08, 0.12, 0.16, 0.20, 0.24, 0.28, 0.32, 0.36, 0.40, 0.50, 1.0, 2.0 M. Protein concentrations of all the fractions, including the flow-through and desalted hemolymph, were estimated by absorbance at 280 nm. Fractions eluted from 0.28 M or higher salt had negligible amounts of protein and were not further analyzed. The protein concentration of the other fractions was equalized using Assay Buffer. For each fraction we then measured the PO activity using an enzyme assay (see below) and levels of each protein relative to the 0.16 M NaCl fraction. At least 2 peptides of PO were used for calculating the average PO level in each fraction. In cases where peptide ratios were above 50-fold and likely beyond the linear dynamic range of the ratio measurements, the high ratios were arbitrarily given the same value as the next highest ratio value below 50-fold.

#### **4.3.6 In solution phenoloxidase assay**

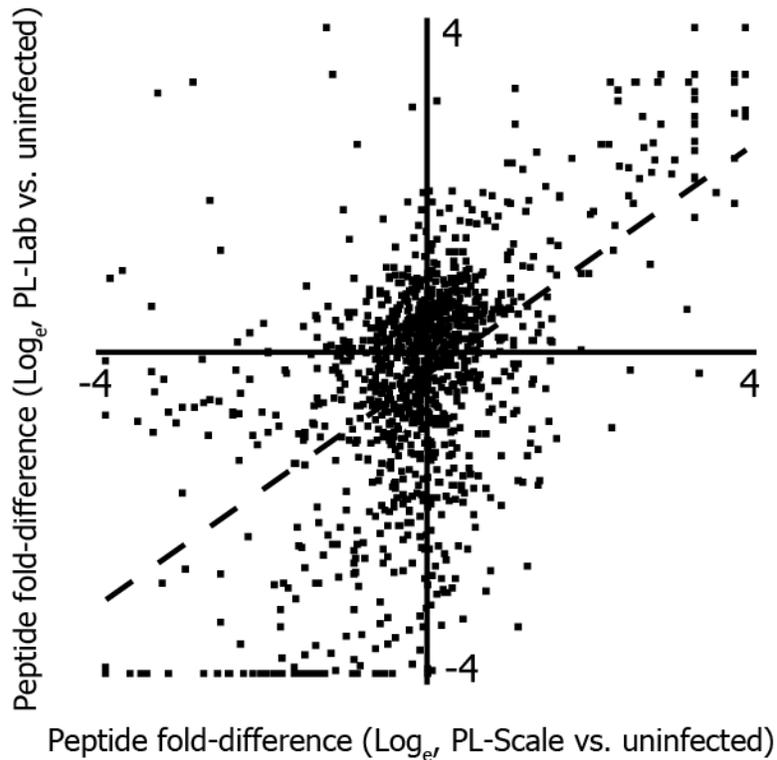
Conducted as described (Pye, 1974), substrate (8  $\mu$ L of 5 mM 4-methylcatechol (Sigma) and 8  $\mu$ L of 40 mM 4-hydroxyproline ethyl ester) was added to 8  $\mu$ L of hemolymph to start the reaction, except for the experiment with larvae of different ages where the hemolymph:substrate ratio was 4:1. Absorbance readings at 520 nm were taken immediately using a Nanodrop spectrophotometer (ND-1000, ThermoFisher Scientific) to calculate the initial reaction rate ( $\Delta A_{520}/\Delta T$ ). For larval aging experiments, instead of the rate, the maximum A520 value was recorded after the highest level was reached in approximately 40 min.

## 4.4 Results

### 4.4.1 Different strains of *P. larvae* produce equivalent outcomes

In order to test the response of worker larvae to *P. larvae* infection, we spray-inoculated a small section of comb containing one-day-old worker larvae with either (A) a homogenate of scale from natural infections of *P. larvae* (PL-Scale) or (B) a laboratory-cultured strain of *P. larvae* (NRRL B-3650, PL-Lab). We used two sources of *P. larvae* as we had no a priori knowledge regarding their relative pathogenicity. Four days post-infection we harvested hemolymph from 5-day larvae and compared the protein expression in the two infected conditions versus an uninfected control using a quantitative proteomics approach (Chan et al., 2008). Using an ultra-high accuracy/resolution LTQ-OrbitrapXL, we identified a combined total 331 proteins (Supplementary Table 18 in Appendix A) with an estimated false discovery rate of 0.30%. To ascertain whether the two infection methods had equivalent effects, we compared 1207 peptides that were quantified in both conditions (Supplementary Table 19 in Appendix A). We observed that protein regulation, as judged from peptide ratios, differed between bees infected with the PL-Lab strain compared to the PL-Scale strain (Figure 4.1,  $p < 0.0001$ ).

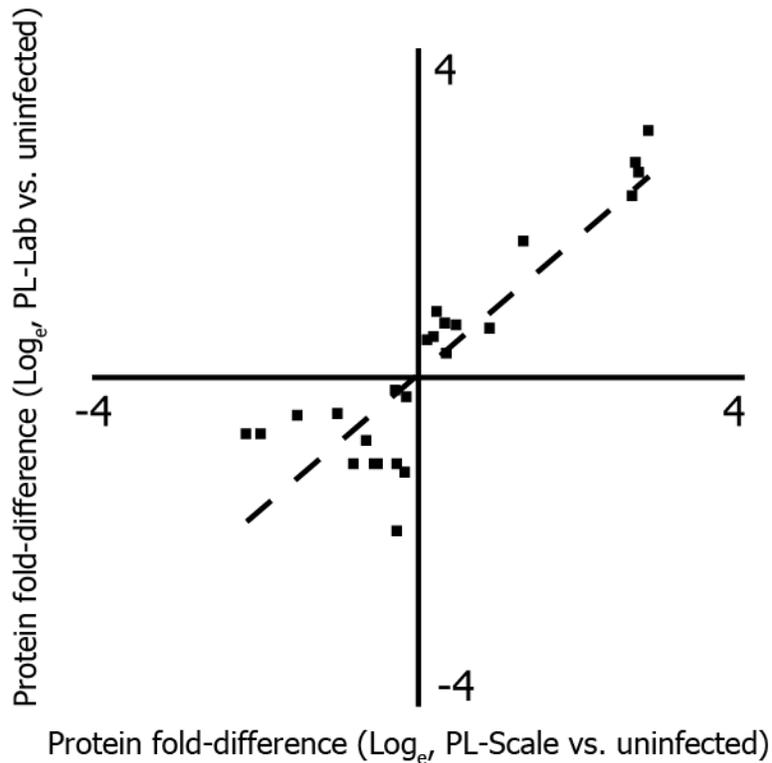
Figure 4.1 Comparing two infection methods based on peptide ratios



Peak volume ratios of the 1207 peptides concomitantly quantified two different infection methods (PL-Lab and PL-Scale), relative to uninfected sample, are shown. Values are in natural log. Linear regression is represented by the diagonal line: slope = 0.69, y-intercept = 0.28,  $R^2 = 0.29$ .

Peptides from hemolymph of larvae infected with PL-Lab appear to be more dramatically down-regulated than those from larvae infected with PL-Scale. However, in considering the 25 proteins quantified in both infections (Figure 4.2) that were significantly different from control ( $p < 0.05$ ), 40% (10) were higher in PL-Scale and 60% (15) were higher in PL-Lab. Taken together, there was not enough evidence to reject the null hypothesis that regulation among these shared proteins differed between the two strains (Wilcoxon matched-pairs signed-ranked test,  $p=0.87$ ).

Figure 4.2 Comparing two infection methods based on peptide ratios

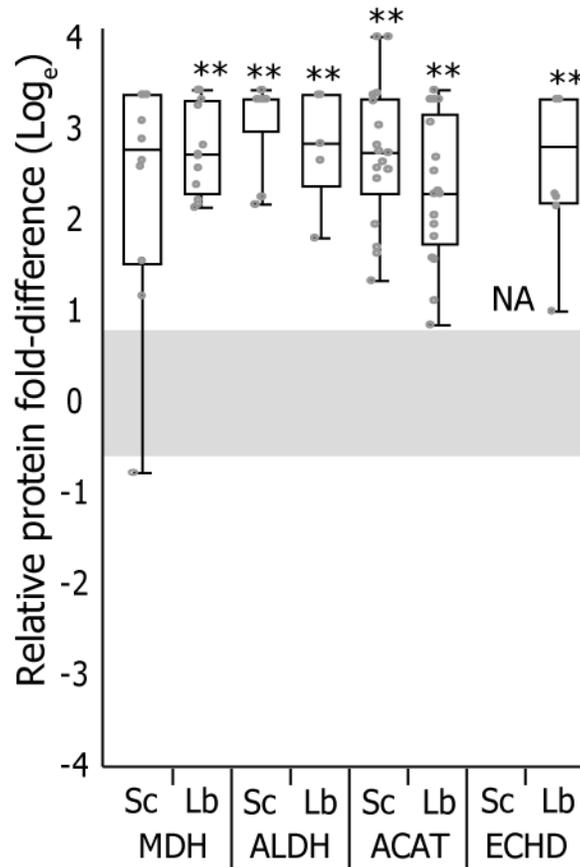


Proteins concomitantly quantified two different infection methods (PL-Lab and PL-Scale), relative to uninfected sample, are shown. Values in are in natural log. Linear regression is represented by the diagonal line: slope = 0.85, y-intercept = 0.041, R<sup>2</sup> = 0.76.

#### 4.4.2 Diseased honey bee larvae express higher levels of mitochondrial metabolic enzymes

A total of 33 proteins, out of 179 quantified, were regulated by a magnitude of at least 2-fold for either one or both inoculums ( $p < 0.05$  or  $p < 0.01$ , two-tailed, non-parametric Wilcoxon matched-pairs signed-ranked statistical test, full results listed in Supplementary Table 20 in Appendix A). Among the most up-regulated of all quantified proteins were several mitochondrial metabolic enzymes (Figure 4.3).

**Figure 4.3 American Foulbrood-induced changes in the levels of mitochondrial metabolic proteins**



Relative levels of protein expression levels in infected hemolymph, using one of two methods PL-Scale (Sc) or PL-Lab (Lb) are expressed in the natural log (y-axis), relative to uninfected bees (defined at 0). Each grey dot represents the level of a single peptide derived from a given protein, compiled from three biological samples. The horizontal bar represents the median level of protein regulation. Those with the median beyond 2-fold (outside of the shaded box) and meets statistical significance, as calculated by the two-tailed Wilcoxon matched-pairs signed-ranked test are marked by a single (\*,  $p < 0.05$ ) or double (\*\*,  $p < 0.01$ ) asterisk. NA = unquantifiable proteins. Abbreviations are defined as follows (name, [accession number]): ACAT (acetyl-CoA acyltransferase, [GI:48097100]); ALDH (aldehyde dehydrogenase, [GI:66530423]); ECHD (enoyl-Coa hydratase, [GI:110773271]); MDH (malate dehydrogenase, [GI:66513092]).

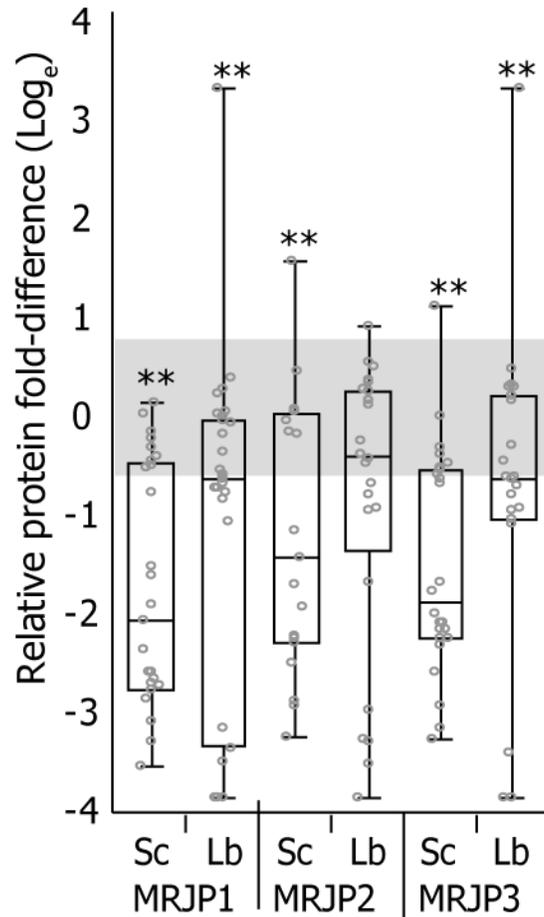
One malate dehydrogenase (MDH) [GI:66513092], for example, showed about a 14-fold increase by both infection methods ( $p < 0.01$ ). This bee MDH is 67% identical to human mitochondrial MDH2 [GI:12804929], implying its direct participation in the mitochondrial matrix and the TCA cycle, instead of the malate-aspartate shuttle that is carried out by human MDH1 [GI:66506786]. Further to this point, the levels of aspartate aminotransferase, another major enzyme of this shuttle, showed no change. An aldehyde dehydrogenase (ALDH) [GI:66530423], a homolog of the human mitochondrial

isoform [GI:118504] was up-regulated 25-fold ( $p < 0.01$ ) in PL-Lab-infected samples, and the same trend was observed in infection with PL-Scale, although it did not reach statistical significance ( $p < 0.1$ ). Acetyl-CoA acyltransferase [GI:48097100], which participates in beta-oxidation and the mevalonate pathway, was significantly ( $p < 0.01$ ) up-regulated at 14- and 9-fold in PL-Scale and PL-Lab, respectively.

#### **4.4.3 Infected larvae deplete their energy stores during infection**

Clearly the metabolic capacity of larvae is undergoing a massive change in response to infection (Figure 4.3), suggesting that concerted changes may also be occurring in their energy stores. Food proteins, which comprise a family of MRJPs, are consistently depleted (Figure 4.4), except in one case that was not significant at the  $p < 0.05$  level.

Figure 4.4 American Foulbrood-induced changes in the levels of food proteins

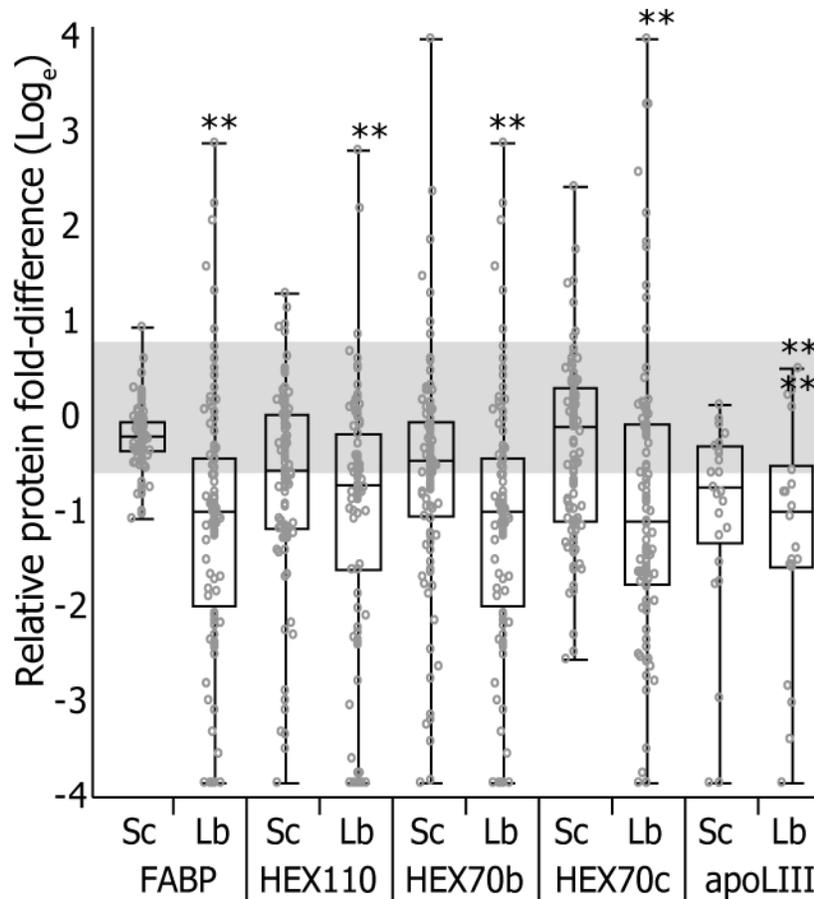


Relative levels of protein expression levels in infected hemolymph, using one of two methods PL-Scale (Sc) or PL-Lab (Lb) are expressed in the natural log (y-axis), relative to uninfected bees (defined at 0). Each grey dot represents the level of a single peptide derived from a given protein, compiled from three biological samples. The horizontal bar represents the median level of protein regulation. Those with the median beyond 2-fold (outside of the shaded box) and meets statistical significance, as calculated by the two-tailed Wilcoxon matched-pairs signed-ranked test are marked by a single (\*,  $p < 0.05$ ) or double (\*\*,  $p < 0.01$ ) asterisk. NA = unquantifiable proteins. Abbreviations are defined as follows (name, [accession number]): MRJP1 (major royal jelly protein 1, [GI:58585098]); MRJP2 (major royal jelly protein 2, [GI:58585108]); MRJP3 (major royal jelly protein 3, [GI:58585142]).

At the same time, larvae collected at this age (5 d post-hatching) should be accumulating enormous levels of hexamerin proteins in the hemolymph (Chan et al., 2008) as an amino acid source for later growth in the pupal stage. However, HEX110 [GI:110761029], HEX70b [GI:58585148] and HEX70c [GI:66549815] have a modest but significant ( $p < 0.01$ ) 2- to 3-fold decrease under PL-Lab infection conditions (Figure 4.5). Similar reductions were seen for two lipid carriers, retinoid- and fatty-acid binding

protein [GI:110758758] and apolipoprotein III [GI:66557660], while a putative neuropeptide Y (NPY) receptor [GI:110764421] that may regulate food intake was strongly up-regulated by the PL-Lab infection ( $p < 0.05$ )

**Figure 4.5 American Foulbrood-induced changes in the levels of energy storage proteins**

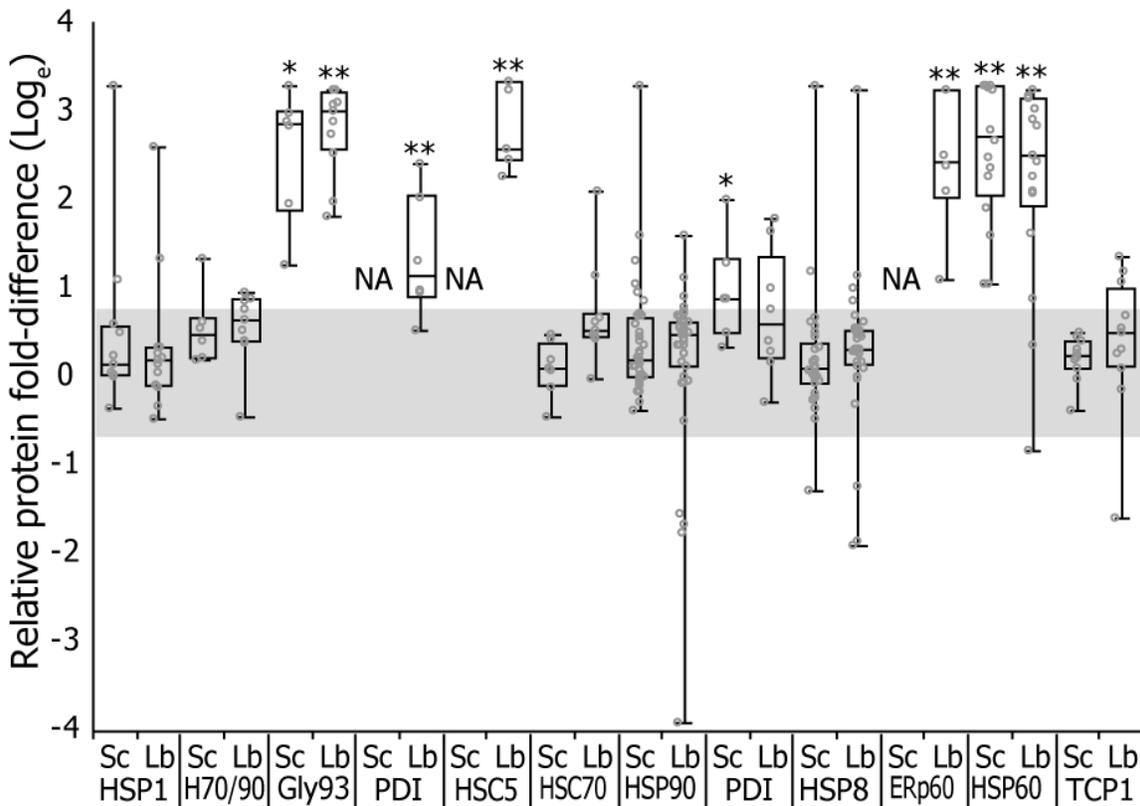


Relative levels of protein expression levels in infected hemolymph, using one of two methods PL-Scale (Sc) or PL-Lab (Lb) are expressed in the natural log (y-axis), relative to uninfected bees (defined at 0). Each grey dot represents the level of a single peptide derived from a given protein, compiled from three biological samples. The horizontal bar represents the median level of protein regulation. Those with the median beyond 2-fold (outside of the shaded box) and meets statistical significance, as calculated by the two-tailed Wilcoxon matched-pairs signed-ranked test are marked by a single (\*,  $p < 0.05$ ) or double (\*\*,  $p < 0.01$ ) asterisk. NA = unquantifiable proteins. Abbreviations are defined as follows (name, [accession number]): FABP (retinoid- and fatty acid binding protein, [GI:110758758]); HEX110 (hexamerin 110, was "larval serum protein 2", [GI:110761029]); HEX70b (hexamerin 70b, [GI:58585148]); HEX70c (hexamerin 70c, was "hexamerin 2 beta", [GI:66549815]); apoLIII (apolipoprotein III, was "hypothetical protein", [GI:66557660]).

#### 4.4.4 The protein-folding/quality control machinery is over-expressed in response to infection

Protein-folding chaperones and heat-shock proteins (HSPs) have been implicated in disease responses due to stress associated with tissue damage (Ranford et al., 2002), with evidence that they also have roles in signal transduction in immune pathways (Asea, 2003). Twenty-six molecular chaperones were detected in larval hemolymph, with many being up-regulated 3- to 20-fold in diseased larvae (Figure 4.6).

**Figure 4.6 American Foulbrood-induced changes in the levels of protein-folding chaperones**



Relative levels of protein expression levels in infected hemolymph, using one of two methods PL-Scale (Sc) or PL-Lab (Lb) are expressed in the natural log (y-axis), relative to uninfected bees (defined at 0). Each grey dot represents the level of a single peptide derived from a given protein, compiled from three biological samples. The horizontal bar represents the median level of protein regulation. Those with the median beyond 2-fold (outside of the shaded box) and meets statistical significance, as calculated by the two-tailed Wilcoxon matched-pairs signed-ranked test are marked by a single (\*,  $p < 0.05$ ) or double (\*\*,  $p < 0.01$ ) asterisk. NA = unquantifiable proteins. Abbreviations are defined as follows (name, [accession number]): ERp60 (a homlog of protein

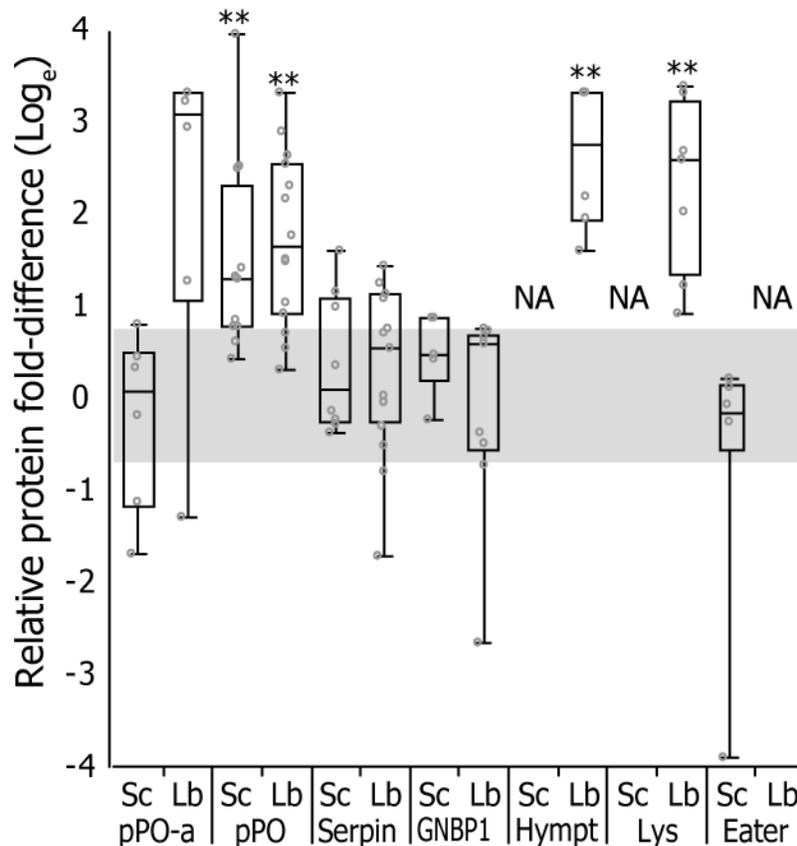
disulfide isomerase, [GI:66546657]); Gly93 (glycoprotein 93, a homolog of HSP90, [GI:110758921]); H70/90 (HSP70/90 organizing protein, [GI:110756123]); HSC5 (heat shock protein cognate 5, [GI:66501507]); HSC70Cb (heat shock cognate 70Cb, [GI:66505007]); HSP1 (heat shock protein 1, [GI:110749824]); HSP60 (heat shock protein 60, [GI:66547450]); HSP8 (heat shock protein 8, [GI:66537940]); HSP90 (heat shock protein 90, [GI:66512625]); PDI (disulfide isomerase, [GI:110768510]); PDI (disulfide isomerase, [GI:66531851]); TCP1 (a homolog of chaperonin, [GI:66560172]).

Among them are three proteins with multiple domains homologous to disulfide isomerases ([GI:110768510], [GI:66546657], [GI:66531851]), a 90 kDa heat shock protein HSP90 [GI:110758921], a 60 kDa heat shock protein HSP60 [GI:66547450] and a heat shock cognate 5 homolog [GI:66501507]. In human studies, heat shock proteins such as HSP60 have been repeatedly linked to macrophage activation (Coggins et al., 1991, Henderson et al., 2006). Hemocytes, being somewhat similar to macrophages in their phagocytic capacity, have been noted to undergo morphological changes during AFB infection, while at the same time populations of other hemolymphic cells increase (Zakaria, 2007) so these effects may be linked with the HSP up-regulation observed here.

#### **4.4.5 Lysozyme and hymenoptaecin levels increase with bacterial challenge**

We were able to identify four low molecular weight defense proteins: lysozyme [GI:66565246], hymenoptaecin [GI:58585174], apidaecin 22 [GI:58585226], and defensin [GI:58585176]. We observed a 13-fold increase of lysozyme in PL-Lab infections ( $p < 0.01$ ) and a 16-fold increase of hymenoptaecin (Figure 4.7) but there were too few peptides detected for apidaecin 22 and defensin to meet our criteria for quantitation (see Materials and Methods). Other immune factors that were identified but did not appear to be regulated by infection include a GGBP [GI:110755978], PGRP-SA [GI:110765019] and PGRP-SC2 [GI:66522804], suggesting that the response seen for lysozyme and hymenoptaecin is a specific response to *P. larvae* infection.

**Figure 4.7 American Foulbrood-induced changes in the levels of proteins with immune function**

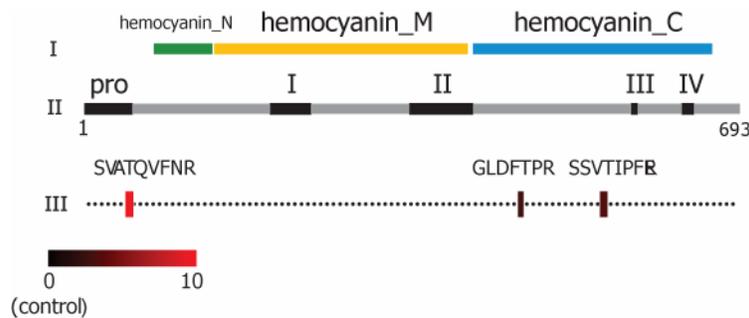


Relative levels of protein expression levels in infected hemolymph, using one of two methods PL-Scale (Sc) or PL-Lab (Lb) are expressed in the natural log (y-axis), relative to uninfected bees (defined at 0). Each grey dot represents the level of a single peptide derived from a given protein, compiled from three biological samples. The horizontal bar represents the median level of protein regulation. Those with the median beyond 2-fold (outside of the shaded box) and meets statistical significance, as calculated by the two-tailed Wilcoxon matched-pairs signed-ranked test are marked by a single (\*,  $p < 0.05$ ) or double (\*\*,  $p < 0.01$ ) asterisk. NA = unquantifiable proteins. Abbreviations are defined as follows (name, [accession number]): Eater (a homolog identified by [10], [GI:110763407]); GNBP1 (Gram-negative binding protein 1, [GI:110755978]); Hympt (hymenoptaecin, [GI:58585174]); Lys (lysozyme, [GI:66565246]); pPO (prophenoloxidase, [GI:58585196]); pPO-a (prophenoloxidase-activating factor, [GI:110758534]); Serpin (serine protease inhibitor 5, [GI:66566441]).

ProPO expression and proteolytic activation are enhanced during infection. The melanization cascade, which leads to the encapsulation of infectious agents, is one of the most important defensive mechanisms of insect innate immunity. One of the central steps in this mechanism is the cleavage of proPO to PO, the active form of the monooxygenase. The PO enzyme, which is activated by proteolytic action, catalyses a key step in the synthesis of melanin and plays a crucial role in melanotic encapsulation.

of invaders (Cerenius et al., 2004). We observed a 4-fold increase ( $p < 0.01$ ) in the expression of proPO in PL-Scale-infected larvae and a 5-fold increase ( $p < 0.01$ ) in PL-Lab-infected larvae (Figure 4.7). The tryptic peptide SVATQVFNR, whose C-terminus is the predicted propeptide cleavage site (Lourenco et al., 2005), was elevated by about 10-fold compared to tryptic peptides found in the remainder of the protein (Figure 4.8). The higher ratios for this propeptide versus the other peptides of the protein suggest that the increased PO response during infection is largely due to the proteolytic activation of an existing pool of PO in the hemolymph and only partially attributable to up-regulated expression.

**Figure 4.8 Mass spectrometry-based peptide analysis for prophenoloxidase**



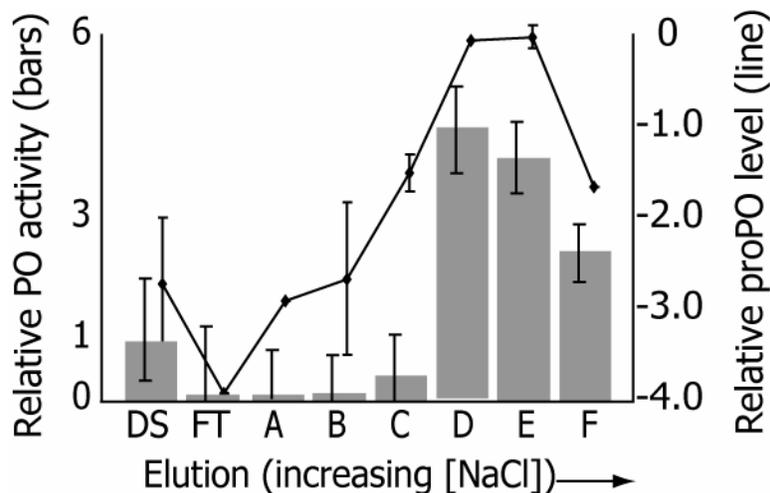
Protein domains (Marchler-Bauer et al., 2005) of proPO are shown in row I. Notable regions (Lourenco et al., 2005) and protein length are shown in row II. Row III describes the average quantity of three peptides in infected larval hemolymph, represented by a color scale to depict fold-differences relative to healthy controls (black). Peptides used in averages and their statistical significances using the two-tailed, paired t-test:  $n=3$  for SVATQVFNR ( $p < 0.05$ ),  $n=3$  for GLDFTPR ( $p < 0.1$ ),  $n=5$  for SSVTIPFER ( $p < 0.05$ ). Averages were generated by considering values from both infection methods (PL-Scale and PL-Lab) together.

#### 4.4.6 Phenoloxidase activity is not found in the first two days of larval development, but increases sharply afterwards

Although PO is well-known for its activity against pathogens, there is little indication so far that its expression level affects the outcome of infection by *P. larvae*. Recent data from our group suggests that proPO levels correlate positively with age (Chan et al., 2008), but we had been unable to establish a full profile of proPO levels during the entire course of larval development due to the low absolute levels of expression. A PO activity assay (Pye, 1974), used to test for the oxygenation of monophenols to diphenols and diphenols to quinones (Mason, 1965), should be more sensitive than mass spectrometry and so was employed here to detect PO activity in

developing, healthy larvae. PO activity was easily detected in crude hemolymph from fourth- to fifth-instar honey bee larvae but there are at least two gene products in the honey bee genome that could function in this assay based on domain comparisons (Marchler-Bauer et al., 2005). To determine which of the two possible proteins is responsible for the PO activity in hemolymph, hemolymph from healthy, fourth and fifth instar larvae was fractionated by strong anion exchange chromatography and the PO activity in each fraction was correlated (Figure 4.9) with the abundance of each protein (Foster et al., 2006).

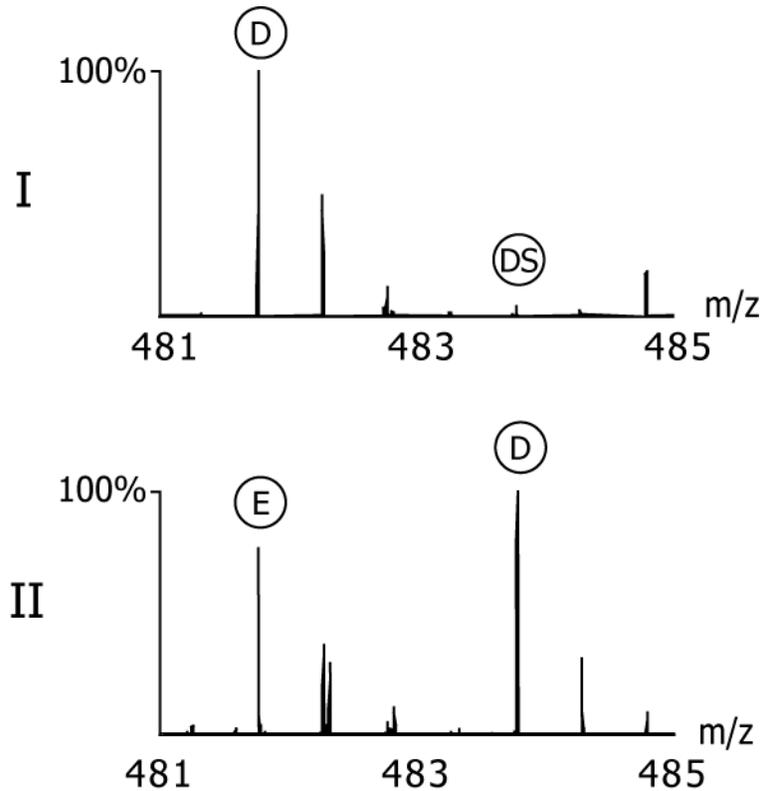
**Figure 4.9 Phenoloxidase activity in fractionated hemolymph**



Hemolymph from fourth- and fifth-instar larvae was fractionated by strong anion exchange using a step gradient of increasing salt (x-axis; A to F from 0.04M to 0.24M salt in 0.04M increments, plus DS=desalted hemolymph, and FT=flowthrough). Each fraction was normalized by protein concentration and was subjected to a PO assay (see Methods). Activity is represented by relative reaction rates (left axis, bars, with DS activity defined at 1). Using mass spectrometry, proPO levels were measured relative to the fraction containing the highest activity (Fraction D), shown on a natural log scale (right axis, N=3). An example is shown in Figure 4.10. Error bars = 2 standard deviations.

The measurement of relative proPO levels was accomplished using differential isotopic labeling of peptides in each fraction, selecting one fraction as the reference to compare against the others (Figure 4.10). The PO activity profile matched very closely to the levels of the gene product annotated as 'prophenoloxidase' [GI:58585196] across the chromatographic fractions and matched very poorly to HEX70b, which also has a putative PO catalytic domain.

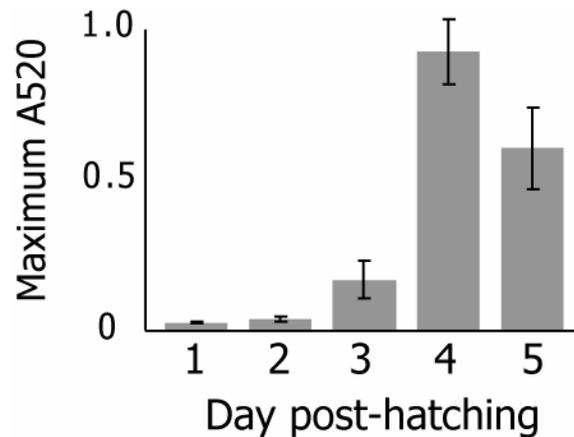
**Figure 4.10 Method for relative peptide quantitation by mass spectrometry: an example**



Two spectra of the differentially labeled (+28Da and +32Da) peptide FSDTIVPR are shown at a 1:1 mixture of peptides from (I) Fraction D and sample DS and (II) Fraction D and E.

Conceivably, older larvae can boost PO activity in response to infection, but could a lack of PO activity in the early larval stages explain the susceptibility of young larvae to *P. larvae* infection? Cell-free hemolymph was extracted from healthy larvae 1 to 5 days after eclosion and tested for PO activity and, indeed, there was no detectable activity in the first two days of development, with some activity detected in day three and substantial activity thereafter (Figure 4.11).

**Figure 4.11 Phenoloxidase activity profile during larval development**



The phenoloxidase (PO) activity assay (see Methods) was conducted on the samples, where activity is represented by the maximum A520 attained by the samples. All PO assay measurements were performed in triplicate. Error bars = 2 standard deviations.

## 4.5 Discussion

Here we have used a quantitative proteomics approach to compare the proteomes of healthy and *P. larvae*-infected *A. mellifera* larvae, leading to the discovery that the infected state is associated with an elevated expression of immunity proteins, chaperones, certain metabolic proteins with an accelerated consumption of energy stores. One particular immune factor, proPO, was particularly up-regulated in response to infection. Intriguingly, the activity level of this enzyme during development of larvae appears to correlate very tightly with susceptibility of the larvae to infection. Our data support a model where the host larva responds to infection not only by producing proteins that can fight the infection directly, but also by engaging its metabolic pathways and energy resources required to support the effort.

The observed depletion of energy stores in the form of MRJPs, hexamerins and lipid transporters suggest that the observed up-regulation of metabolic enzymes is, at least in part, tied to energy production. In larvae infected with *E. coli*, which is not a natural pathogen of bees, MRJP-1, MRJP-7/MRJP-2 were mildly lowered in expression compared to the mock-infected control (Randolt et al., 2008). The negative correlation between energy availability and infection survival has also been observed in other insects such as the butterfly *Pieris rapa* (Asgari et al., 2004) and the bumblebee *Bombus terrestris* (Moret et al., 2000). In further support of the increased energy demands of the infected state, the putative bee NPY receptor was also up-regulated in infected

larvae. While the ligand for this receptor in bees remains unknown, the receptor and its cognate ligand in mammals control feeding and appetite in mammals (Naslund et al., 2007).

The most obvious class of proteins expected to increase in response to infection are those involved in the innate immune response. Lysozyme's primary known function is to degrade the peptidoglycan shell of Gram-positive bacteria (Masschalck et al., 2003) and is therefore expected to have a significant role in inhibiting *P. larvae*. Interestingly, the C-type lysozyme [GI:110762174] that has been previously shown to be up-regulated upon infection (Evans et al., 2006) is not the lysozyme we have identified here, which is also known as the destabilase-lysozyme [GI:66565246]. Because these two forms are drastically different (e.g., the best-matched region is only 50 amino acids long and shares only 20% sequence identity), it is clearly not a case of the peptides identified by MS/MS being shared by both enzymes. The 13- to 16-fold up-regulation of destabilase-lysozyme suggests that it can be important in host defense, which is also supported by the observation that its homolog has antimicrobial activity in the medicinal leech *Hirudo medicinalis* (Zavalova et al., 2006). The AMPs comprise another humoral-based defense mechanism, killing Gram-negative and positive bacteria alike (Casteels et al., 1993, Randolt et al., 2008); many of them work by forming pores in the bacterial cell wall. Among those known in bees, we were only able to quantify hymenoptaecin in the hemolymph, and its dramatic up-regulation suggests that it plays a crucial role in defending against *P. larvae*, a conclusion that is supported by transcript and other protein-level data as well (Evans et al., 2006, Randolt et al., 2008). Conspicuously absent in our data, however, are defensin and abaecin, which have both been implicated in the larval response to *P. larvae* (Bachanova et al., 2002, Bilikova et al., 2001, Evans, 2004, Evans et al., 2004, Evans et al., 2006). Although we detected peptides from defensin, the signal:noise ratios in the MS1 spectra were not high enough to allow accurate quantitation; no abaecin peptides were detected. Our inability to detect these two AMPs with sufficient signal suggests that their concentration is likely much lower than hymenoptaecin, which is confirmed by a recent 1DGE study of larval hemolymph (Randolt et al., 2008).

The consistent up-regulation of proPO in both infection methods is in agreement with the well-characterized antimicrobial activity of this enzyme. The ability of larvae to

employ melanization as a defense mechanism has been questioned because the proPO levels are low compared to adults, to the point of being undetectable on a stained 1D gel (Randolt et al., 2008). In our own experiments, even with ion-exchange fractionation prior to LC-MS/MS analysis on an LTQ-OrbitrapXL, one of the most sensitive systems available, it was difficult to detect throughout most of larval development except for the oldest samples (5 days post-hatching) (Chan et al., 2008). However, we are clearly able to detect PO activity as early as 3 days after hatching, where the absence of activity in the earlier timepoints (days 1 and 2) precisely match the period of maximum susceptibility of the larvae to AFB (Shimanuki, 1997, Woodrow et al., 1942). Thus, our data argue that older larvae have significant levels of PO and that they are indeed capable of utilizing the PO pathway to fight infection.

## **4.6 Conclusions**

The larval stage of a honey bee represents a unique system for applying proteomics to probe host-pathogen interactions. Unlike most other systems, proteins in larvae not only play major roles in immune defense but also constitute one of their primary stores of energy. Studying such a response in most other systems with more conventional energy reserves (e.g., lipids) would necessitate a wide variety of tools in order to monitor energy usage, immune factor production and metabolic flux all at the same time. By monitoring all these aspects simultaneously, our data clearly demonstrate that host defense against bacterial challenge is a concerted response involving proteins that kill the microbes directly, as well as metabolic and cell/protein repair enzymes that indirectly support this defensive effort. By using proteomics techniques on this unique model organism where immunity and protein energy flux are tightly coupled, we have been able to build a more comprehensive picture of the insect innate immune response.

## 4.7 References

- Asea A. **Chaperokine-induced signal transduction pathways.** *Exerc Immunol Rev* 2003, 9:25-33.
- Asgari S, Schmidt O. **Isolation of an imaginal disc growth factor homologue from *Pieris rapae* and its expression following parasitization by *Cotesia rubecula*.** *J Insect Physiol* 2004, 50(8):687-694.
- Bachanova K, Kludiny J, Kopernicky J, Simuth J. **Identification of honeybee peptide active against *Paenibacillus* larvae larvae through bacterial growth-inhibition assay on polyacrylamide gel.** *Apidologie* 2002, 33:259-269.
- Bilikova K, Wu G, Simuth J. **Isolation of a peptide fraction from honeybee royal jelly as a potential antifoulbrood factor.** *Apidologie* 2001, 32:275-283.
- Brødsgaard CJ, Ritter W, Hansen H. **Response of in vitro reared honey bee larvae to various doses of *Paenibacillus* larvae larvae spores.** *Apidologie* 1998, 29:569-578.
- Casteels P, Ampe C, Jacobs F, Tempst P. **Functional and chemical characterization of Hymenoptaecin, an antibacterial polypeptide that is infection-inducible in the honeybee (*Apis mellifera*).** *J Biol Chem* 1993, 268(10):7044-7054.
- Cerenius L, Soderhall K. **The prophenoloxidase-activating system in invertebrates.** *Immunol Rev* 2004, 198:116-126.
- Chan QW, Howes CG, Foster LJ. **Quantitative comparison of caste differences in honeybee hemolymph.** *Mol Cell Proteomics* 2006, 5(12):2252-2262.
- Chan QW, Foster LJ. **Changes in protein expression during honey bee larval development.** *Genome Biol* 2008, 9(10):R156.
- Coggins CC, Gumbardo DJ. **The nursing process applied to staff development.** *J Nurs Staff Dev* 1991, 7(4):196-200.
- Davidson EW. **Ultrastructure of American foulbrood disease pathogenesis in larvae of the worker honey bee, *Apis mellifera*.** *J Invertebr Pathol* 1973, 21:53-61.
- Evans JD. **Transcriptional immune responses by honey bee larvae during invasion by the bacterial pathogen, *Paenibacillus* larvae.** *J Invertebr Pathol* 2004, 85(2):105-111.
- Evans JD, Lopez DL. **Bacterial probiotics induce an immune response in the honey bee (Hymenoptera: Apidae).** *J Econ Entomol* 2004, 97(3):752-756.
- Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, Kanost M, Thompson GJ, Zou Z, Hultmark D. **Immune pathways and defence mechanisms in honey bees *Apis mellifera*.** *Insect Mol Biol* 2006, 15(5):645-656.
- Foster LJ, de Hoog CL, Zhang Y, Zhang Y, Xie X, Mootha VK, Mann M. **A Mammalian Organelle Map by Protein Correlation Profiling.** *Cell* 2006, 125(1):187-199.
- Gregorc A, Bowen ID. **Histopathological and histochemical changes in honeybee larvae (*Apis mellifera* L.) after infection with *Bacillus* larvae, the causative agent of American foulbrood disease.** *Cell Biol Int* 1998, 22(2):137-144.

- Henderson B, Allan E, Coates AR. **Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection.** *Infect Immun* 2006, 74(7):3693-3706.
- Lourenco AP, Zufelato MS, Bitondi MM, Simoes ZL. **Molecular characterization of a cDNA encoding prophenoloxidase and its expression in *Apis mellifera*.** *Insect Biochem Mol Biol* 2005, 35(6):541-552.
- Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z *et al.* **CDD: a Conserved Domain Database for protein classification.** *Nucleic Acids Res* 2005, 33(Database issue):D192-196.
- Mason HS. **Oxidases.** *Annu Rev Biochem* 1965, 34:595-634.
- Masschalck B, Michiels CW. **Antimicrobial properties of lysozyme in relation to foodborne vegetative bacteria.** *Crit Rev Microbiol* 2003, 29(3):191-214.
- McCornack R. **Extended Tables of the Wilcoxon Matched Pair Signed Rank Statistic.** *Journal of the American Statistical Association* 1965, 60(311):864-871.
- Moret Y, Schmid-Hempel P. **Survival for immunity: the price of immune system activation for bumblebee workers.** *Science* 2000, 290(5494):1166-1168.
- Naslund E, Hellstrom PM. **Appetite signaling: from gut peptides and enteric nerves to brain.** *Physiol Behav* 2007, 92(1-2):256-262.
- Pye AE. **Microbial activation of prophenoloxidase from immune insect larvae.** *Nature* 1974, 251:610-613.
- Randolt K, Gimple O, Geissendorfer J, Reinders J, Prusko C, Mueller MJ, Albert S, Tautz J, Beier H. **Immune-related proteins induced in the hemolymph after aseptic and septic injury differ in honey bee worker larvae and adults.** *Arch Insect Biochem Physiol* 2008, 69(4):155-167.
- Ranford JC, Henderson B. **Chaperonins in disease: mechanisms, models, and treatments.** *Mol Pathol* 2002, 55(4):209-213.
- Shimanuki H. **Bacteria.** In: *Honey bee pests, predators, and diseases*. Edited by Morse RA, Flottum K. Medina: A.I. Root Co.; 1997: 35-54.
- Woodrow AW, Holst EC. **The mechanism of colony resistance to American foubrood.** *Journal of Economic Entomology* 1942, 35:892-895.
- Yue D, Nordhoff M, Wieler LH, Genersch E. **Fluorescence in situ hybridization (FISH) analysis of the interactions between honeybee larvae and *Paenibacillus* larvae, the causative agent of American foulbrood of honeybees (*Apis mellifera*).** *Environ Microbiol* 2008, 10(6):1612-1620.
- Zakaria ME. **The cellular immunity responses in the haemolymph of honey bee workers infected by american foulbrood disease (AFB).** *Journal of Applied Science Research* 2007, 3(1):56-63.
- Zavalova LL, Yudina TG, Artamonova, II, Baskova IP. **Antibacterial non-glycosidase activity of invertebrate destabilase-lysozyme and of its helical amphipathic peptides.** *Chemotherapy* 2006, 52(3):158-160.

## 5 CONCLUSION

Honey bee research has been largely overlooked as an important area of science. The past five years, during which the work in this thesis has been conducted, has been a time of immense change and excitement for bee research. The more recent trouble caused by the Colony Collapse Disorder has brought attention from the public and scientific community. Undeniably, the economic losses to beekeepers and agricultural products that rely on bee pollination have been great, but it has also motivated scientists to devote more research efforts to bees.

The resources to learn about this species, particularly regarding their molecular biology, have also been dramatically upgraded in the past five years. The official publication of the honey bee genome in 2006 (Consortium, 2006), and with the application of bioinformatic tools gave predicted gene and protein sequence libraries, which are inherently useful for comparing against other organisms, and learning how similar or different they may be from honey bees. Inferences by homology from well-studied species are often very helpful for understanding how various genes and proteins attribute to the anatomy and physiology of a less understood species such as bees. Algorithms such as BLAST or similar utilities are commonly employed. Usage of sequence data in a laboratory context however, frequently involves high-throughput methodologies such as microarrays for gene expression studies, and MS-based proteomics for protein expression studies. In the latter technology, instrumentation has become more widespread over the past five years; more importantly, instrument sensitivity, accuracy, and speed have been much improved. It has become routine to identify hundreds of proteins in one or two hours worth of time on a mass spectrometer. With the application of isotopic labels to paired samples, relative quantities of many proteins can be conveniently measured. This method is highly suitable for studying temporal and infection-induced changes in an organism, which are major goals of the research conducted in this thesis.

### 5.1 Addressing the project aims

The goal of this project was to improve our understanding of the molecular biology of honey bee larvae, particularly regarding the host immune response upon *P.*

*larvae* challenge; the differences between young and old larva that explains their highly age-dependent resistance against *P. larvae* infection.

In the three manuscripts included in this thesis, we mainly chose to work with hemolymph. This is an obvious place to start for insect research because it is well known that immune factors permeate to all the tissues in a body through this medium. Furthermore, as most of the other organs in a bee are rather small, hemolymph is relatively easy to obtain and in reasonably large amounts. This is an important consideration especially for Chapter 3, the profiling of larval proteome during development, since the youngest one day-old larvae were less than 1 mm long, making it virtually impossible to dissect any organ by conventional practices.

Our approach was to first determine whether there are measurable differences in the proteome among bees of different ages. For this we compared worker adults and larvae (Chapter 2). Even with gel electrophoresis of hemolymph we could see major disparities in their banding patterns, and beyond that, MS was able to reveal even finer and more differences. From the list of proteins we quantified, the ones with a predicted role in immunity were clearly more abundant in adults. While we had used older larvae that would have acquired resistance against *P. larvae* by that age, we reasoned that factors conferring protection should increase gradually with age. If we had tested younger larvae that are expected to have nearly no immunity and older larvae that do, we should be able to deduce what proteins are necessary for defense against bacterial invasion. Furthermore, we hope that the expression profile of one or more protective factors should closely mirror the observed age where *P. larvae* resistance is gained – approximately the third larval instar.

With this in mind, we collected specimens over the entire length of larval development for five consecutive days – both hemolymph and the solid tissue remains (Chapter 3). With MS we found the expression profiles of hundreds of proteins relative to their level in day three (the approximate midpoint of the larval stage). While it would have been ideal to track the quantities of immunity proteins found in Chapter 2, so that this paper can add to our knowledge of them, this was unfortunately not possible. Using practically the same method as in Chapter 2, we were only able to detect some of them, let alone quantify them. We saw only four immunity-related proteins with this effort, but only two – prophenoloxidase and the antimicrobial peptide apisimin – significantly

increased with age. With the age-associated increase of enzymes related with ATP synthesis, TCA cycle, and most dramatically the hexamerins, it seems clear that the larval phase is only mainly concerned with metabolism, presumably to power its incredible expansion in size over such a short time, and storage of energy in preparation for pupation. It seems reasonable to expect that, in face of bacterial challenge, energy for growth must be diverted towards immunity instead.

Given what we know from Chapter 3, we were able to give plausible explanations of our observations in Chapter 4. We compared the hemolymph proteins of healthy and *P. larvae* infected larvae, and saw that infection had induced significant changes in metabolic enzymes. Proteinaceous forms of energy stores such as hexamerins were clearly diminished, while TCA cycle proteins were upregulated. These samples were derived from five days post-eclosion larvae, a time in which they were expected to have a maximum level of hexamerins as energy storage. In the infected samples, we saw dramatic induction in the expression of the antimicrobial peptide hymenoptaecin, lysozyme, and prophenoloxidase. There was also a clear increase in some enzymes associated with protein folding, which is indicative of physiological stress. Based on the relatively few immunity-related proteins we observed in both Chapters 3 and 4, we can reasonably speculate that the immune defense of a healthy larvae – even an older one – maybe quite poor; yet older larvae have the energy reserves that makes it possible to mount a defense when the need arises, while the younger ones do not and they consequently succumb to infection.

## 5.2 Addressing the project hypothesis

We began the work in this thesis with the hypothesis that exposure of honey bee larva during the first 48 h after eclosion to *P. larvae* is the cause of American foulbrood due to the host's underdeveloped immune system; the corollary to this is that older larvae and adults have a sufficiently developed one. Implicit is the assumption that the strength of the immune system is a measurable quantity; represented primarily by proteins with defensive functions, which may be detected by MS-based techniques. In Chapter 2, where we compared the expression of proteins between adult and larval hemolymph, we demonstrated that differences were clearly discernable, especially regarding their repertoire of immune proteins. We followed this with a detailed study of

larval development, expecting to find one or more proteins whose expression level correlates with the gain of resistance against *P. larvae* (Chapter 3).

One of those we saw was prophenoloxidase: expressed at greater than 50-fold higher in worker adults compared to larvae of five days post-hatching, and about 24-fold higher in these larvae over those younger by two days. This protein was completely undetectable by MS in larvae younger than this, likely because the levels are too low. This gives reasonable grounds to suspect that the lack of phenoloxidase is associated with lethality in *P. larvae* challenge. We were also able to profile phenoloxidase activity over the course of larval development, and the resulting plot mirrors very closely to the gain of resistance against infection. The significance of this protein in host defense is further supported by the increase prophenoloxidase by 4- to 5-fold in infected larval hemolymph compared to healthy controls.

Based on our data we did not see any protein with which we could make the same argument as we did phenoloxidase. The expression profile of the antimicrobial peptide apisimin was somewhat similar, but the rise in its abundance on day four does not match as closely to *P. larvae* resistance as phenoloxidase activity, therefore this peptide may not be the critical factor necessary for survival. There is no knowledge of apisimin's activity against microorganisms, only that it is present in royal jelly and its cDNA is found in nurse bee heads (Bilikova et al., 2002).

Despite only finding one protein, phenoloxidase, with several lines of evidence suggesting its importance to host defense, there is good reason to believe that it can make a life-or-death difference to larvae. This enzyme is found even in healthy hemolymph, and acts within minutes to generate cytotoxic molecules. Nodulation of bacteria, which is the predominant insect defense, is often followed by melanization. The cell lysis that occurs in the generation of nodules quickly releases large amounts of phenoloxidase. Coupled with the metabolic resources that older larvae have, perhaps diverted from growth to the massive biosynthesis of phenoloxidase or other enzymes such as lysozyme, the host almost always survives *P. larvae* infection.

### **5.3 Other contributions to honey bee knowledge**

For the purposes of addressing the project aims and hypothesis, we focus on proteins of the immune system in the honey bee larvae. One benefit of shotgun

proteomics is to identify whatever proteins are contained in the sample, and based on our knowledge of their functions, would bring considerable insight regarding the sample's biological state.

As of yet there are very few bee proteins whose existence is backed by experimental evidence. Genome annotation efforts have predicted about 9000 proteins, using a combination of *ab initio* and homology methods (Elsik et al., 2007). With MS, we have identified thousands of peptides and matched them to protein sequences that lack prior evidence. With now a larger list of proteins that have been ascertained to exist, these can be used by informaticians as training sets in downstream gene prediction algorithms, which will improve the annotation. Furthermore, MS-sequenced peptides can also correct sequence mis-assignments, which often take place at intron-exon splice sites. We have taken mass spectra that have not been matched to a peptide from the predicted protein set and re-searched them against the six-frame translated genome itself. Based on similar work done on other organisms such as *C. elegans* (Merrihew et al., 2008), it is not surprising that the MS data generated here was able to correct splice site mis-annotations. There is also multi-peptide evidence for previously undiscovered proteins when the data was searched against the translated genome. Short proteins such as antimicrobial peptides can be missed by prediction software, since there is a higher risk of false-positives in small open reading frames (Basrai et al., 1997).

Peptide spectra can be used to help with selected reaction monitoring experiments. In this method, the targeted observation of a peptide fragment that is highly representative of a protein of interest is considered adequate to consider it identified. Tracking these so-called proteotypic peptides increases the detection sensitivity dramatically, since the mass spectrometer can focus on the monitoring of a specific fragment instead of the untargeted, discovery-based approach of shotgun proteomics. Therefore, the spectra, peptide and protein evidence generated from this work are very helpful resources for MS-based research for specific proteins and bee molecular biology in general. Plans are well under way to help ensure such contributions to the field are made publicly accessible (Section 5.4.3).

## **5.4 Future directions**

### **5.4.1 Phenoloxidase knockdown study**

A major significance of the work outlined in this thesis is to suggest the importance of phenoloxidase as a defensive protein against *P. larvae* infection. A more definitive experiment would be to perform a knockdown of this enzyme using RNAi in older larvae, and observe whether they will remain resistant against bacterial challenge. If this treatment is lethal to all or nearly all of the bees, it is a strong indication that phenoloxidase activity is highly critical to defense; if there is no apparent effect on survival, this enzyme is not as important as we thought; if there are intermediate levels of mortality, it suggests that phenoloxidase and some other proteins or mechanism is at work.

### **5.4.2 Treatments and prevention against AFB**

Ultimately, we hope these studies will pave way for new treatments or preventative measures against AFB. If phenoloxidase is truly important for host defense, one can conceive the design of a drug that can induce its expression, to be fed or sprayed on very young larvae that otherwise would not express it. Honey bee strains can be selected with a genetic predisposition to overexpress key this enzyme. Further to this, the breeding program can begin with bees with already high levels of this enzyme, or express at an earlier age than the average strain. This can be tested either by MS or by the phenoloxidase activity assay used in Chapter 4 of this thesis.

### **5.4.3 PeptideAtlas and inclusion of novel proteins**

One major value of having a database of peptide mass spectra, and knowing whether a given peptide is observed frequently or not is extremely useful in the MS-based selected reaction monitoring technique. Already under way is a collaboration with Eric W. Deutsch to create a honey bee PeptideAtlas: a publicly accessible database of peptide mass spectra, made from comprehensive re-processing of raw MS data to derive information such as proteotypic peptides, false discovery rate, frequency of protein or peptide observation across various samples – to name a few. Furthermore, the novel and corrected proteins due to the mis-annotation of start, end, and intro-exon splice sites found using our MS data will be included in PeptideAtlas. This will allow more

spectra to be matched to peptide hits, thus maximizing the information derived from each MS run.

#### **5.4.4 Comparison to transcriptomics data**

Although our focus is on honey bee immunity, numerous other proteins have also been concomitantly identified in our proteomic studies. Most of them have never been observed before, and new expression patterns have been seen across entire functional classes of enzymes. All of these discoveries will open doors to many unique and testable hypotheses for the scientific community to explore, particularly regarding honey bee larval development. It would also be very interesting to conduct similar experiments using transcriptomics as a readout, and compare it to the proteomics data we have obtained here.

### **5.5 Closing**

Our biochemical understanding of *A. mellifera* has lagged far behind other model organisms. As it has been shown in this thesis, MS-based proteomics is a powerful tool to provide rapid advancement of our knowledge of this most beneficial insect. American foulbrood is one of the oldest diseases of managed honey bees and with the current wave of problems resulting from Colony Collapse disorder on top of existing bee diseases, it is with hope that the knowledge we have gained in this thesis will serve a useful purpose in finding solutions to these problems.

## 5.6 References

- Basrai MA, Hieter P, Boeke JD. **Small open reading frames: beautiful needles in the haystack.** *Genome Res* 1997, 7(8):768-771.
- Bilikova K, Hanes J, Nordhoff E, Saenger W, Kludiny J, Simuth J. **Apisimin, a new serine-valine-rich peptide from honeybee (*Apis mellifera* L.) royal jelly: purification and molecular characterization.** *FEBS Lett* 2002, 528(1-3):125-129.
- The Honey Bee Genome Sequencing Consortium. **Insights into social insects from the genome of the honeybee *Apis mellifera*.** *Nature* 2006, 443(7114):931-949.
- Elsik CG, Mackey AJ, Reese JT, Milshina NV, Roos DS, Weinstock GM. **Creating a honey bee consensus gene set.** *Genome Biol* 2007, 8(1):R13.
- Merrihew GE, Davis C, Ewing B, Williams G, Kall L, Frewen BE, Noble WS, Green P, Thomas JH, MacCoss MJ. **Use of shotgun proteomics for the identification, confirmation, and correction of *C. elegans* gene annotations.** *Genome Res* 2008, 18(10):1660-1669.

## APPENDIX A

**Supplementary Table 1.** A list of 3657 peptides for the identification of hemolymph proteins. Spectra acquired from all honey bees were searched against the *A. mellifera* protein library using either Mascot (M) or X!Tandem (X). Peptide scores displayed are dependent on the search engine used. If the same peptide was identified by both programs, Mascot data takes precedence. All peptides shown satisfy the set criteria for significance (see Experimental Procedures of Chapter 2). Available online <http://www.mcponline.org/cgi/data/M600197-MCP200/DC1/2>.

**Supplementary Table 2.** Proteins considered identified based on criteria for significance (see Experimental Procedures of Chapter 2). Percent sequence coverage for each protein is shown for data combined from all honey bees (total, T), queen (Q), drone (D), worker adult (W) and larva (WL). Worker larvae were analyzed more extensively than the adult castes and so sequence coverage numbers are not directly comparable for larva versus adult. Available online <http://www.mcponline.org/cgi/data/M600197-MCP200/DC1/3>.

**Supplementary Table 3.** Peptides used for identification of the DWV polyprotein. Measured peptide mass, mass error, charge, and Mascot scores are shown. Peptide marked by an asterisk was crucial in differentiating between the identification of DWV polyprotein and the highly identical varroa destructor virus. Available online <http://www.mcponline.org/cgi/data/M600197-MCP200/DC1/4>.

**Supplementary Table 4.** Assignment of GO identification numbers to hemolymph proteins. Assignments were made by Blast2GO using criteria stated in Experimental Procedures of Chapter 2. Available online <http://www.mcponline.org/cgi/data/M600197-MCP200/DC1/5>.

**Supplementary Table 5.** List of hemolymph proteins under GO identification numbers at level 3 of a directed acyclic graph generated using terms under the Molecular

Function ontology. Available online <http://www.mcponline.org/cgi/data/M600197-MCP200/DC1/6>.

**Supplementary Table 6.** List of hemolymph proteins under GO identification numbers at level 3 of a directed acyclic graph generated using terms under the Biological Process ontology. Available online <http://www.mcponline.org/cgi/data/M600197-MCP200/DC1/7>.

**Supplementary Table 7.** Peptides found in direct genome search. Spectra not matching known honey bee proteins were searched against the Amel\_2.0 genome assembly using Mascot. Scores displayed were generated by Mascot. Available online <http://www.mcponline.org/cgi/data/M600197-MCP200/DC1/8>.

**Supplementary Table 8.** Relative quantitation of bee larval proteome. Available online <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760883/bin/gb-2008-9-10-r156-S1.xls> in tab "ST1."

**Supplementary Table 9.** Gene Ontology terms assigned to honey bee larval proteins. Available online <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760883/bin/gb-2008-9-10-r156-S1.xls> in tab "ST2."

**Supplementary Table 10.** Gene Ontology categorization of proteins by Molecular Function using directed acyclic graphs. Available online <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760883/bin/gb-2008-9-10-r156-S1.xls> in tab "ST3."

**Supplementary Table 11.** Gene Ontology "Molecular Function" vocabularies assigned to proteins on level 3 of a directed acyclic graph. Available online <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760883/bin/gb-2008-9-10-r156-S1.xls> in tab "ST4."

**Supplementary Table 12.** Manually assigned protein function and functional class. Available online <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760883/bin/gb-2008-9-10-r156-S1.xls> in tab "ST5."

**Supplementary Table 13.** Average slope values of proteins within manually assigned functional classes. Available online <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760883/bin/gb-2008-9-10-r156-S1.xls> in tab "ST6."

**Supplementary Table 14.** Enrichment analysis of hierarchical clustering of proteins profiled from the honey bee larval solid tissue. Available online <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760883/bin/gb-2008-9-10-r156-S1.xls> in tab "ST7."

**Supplementary Table 15.** Enrichment analysis of hierarchical clustering of proteins profiled from the honey bee larval hemolymph. Available online <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760883/bin/gb-2008-9-10-r156-S1.xls> in tab "ST8."

**Supplementary Table 16.** Enrichment analysis of hierarchical clustering of proteins profiled from the honey bee larval hemolymph. Available online <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760883/bin/gb-2008-9-10-r156-S1.xls> in tab "ST9."

**Supplementary Table 17.** Peptide sequence data including Mascot score, charge, and mass error. Available online <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760883/bin/gb-2008-9-10-r156-S1.xls> in tab "ST10."

**Supplementary Table 18.** List of proteins considered identified by mass spectrometry-based sequencing and the peptide sequences of each protein. Protein accession numbers preceded by "999" are proteins that have been falsely discovered by matches

reversed peptide sequences. Available online

<http://www.biomedcentral.com/content/supplementary/1471-2164-10-387-s1.xls>.

**Supplementary Table 19.** List of quantified peptides and their relative expression values. Available online <http://www.biomedcentral.com/content/supplementary/1471-2164-10-387-s2.xls>.

**Supplementary Table 20.** List of proteins, their median averaged values based on peptide relative expression. Available online <http://www.biomedcentral.com/content/supplementary/1471-2164-10-387-s3.xls>.

**Supplementary Table 21.** List of proteins discussed in the paper with direct mention of their known or putative function, and the evidence or resource for this information. Available online <http://www.biomedcentral.com/content/supplementary/1471-2164-10-387-s4.xls>.